# Fisher Discriminant Analysis with New Between-class Scatter Matrix for Audio Signal Classification

Yuechi Jiang
Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China
yuechi.jiang@connect.polyu.hk

Frank H. F. Leung
Electronic and Information Engineering
The Hong Kong Polytechnic University
Hong Kong, China
frank-h-f.leung@polyu.edu.hk

*Abstract*—**Fisher Discriminant Analysis (FDA) is a widely used technique for signal classification. Its application varies from face recognition to speaker recognition. FDA aims to project a given feature onto a projected space, where the features coming from the same class are moved closer, while those coming from different classes are moved farther. However, in the original formulation of FDA, the number of orthogonal projection directions is limited by the number of classes, which may hinder the effectiveness of FDA as a projection technique. In this paper, we propose to use new between-class scatter matrices to replace the original between-class scatter matrix, in order to increase the number of orthogonal projection directions. We call FDA with these new between-class scatter matrices the Modified FDA (MFDA). The effectiveness of MFDA and FDA as a projection technique is compared through doing two audio signal classification tasks. Both linear version and kernel version of MFDA and FDA are evaluated, and experimental results show that MFDA can outperform FDA in both classification tasks.**

***Keywords—Fisher discriminant analysis; modified Fisher discriminant analysis; new between-class scatter matrix; audio signal classification***

## I. INTRODUCTION

Fisher Discriminant Analysis (FDA) has been prevailing as a fundamental projection technique. Although the linear version of FDA (LDA) is quite popular, the kernel version of FDA (KDA) makes it more general [1][2]. As a dimensionality reduction technique, it can be used to extract low-dimension features from high-dimension features, such as Fisherfaces [3]-[5] and some modifications on Fisherfaces [6][7] for face recognition, and Fishervoices [8]-[11] for speaker recognition. A summary of different variants of FDA can be found in [12]. As a general projection technique, FDA can be used to project out the interference information embedded in the high-dimension features, such as the usage in i-vector to suppress channel variability [13][14]. It can also be used to improve the discrimination of the raw feature [15][16].

LDA aims at projecting a feature from the original feature space onto a projected feature space, such that in the projected feature space, the features in the same class are closer while those in different classes are farther. In the projection space, each projection direction (i.e. basis) is an eigenvector of $S_W^{-1}S_B$, where $S_W$ is the within-class scatter matrix and $S_B$ is the between-class scatter matrix [17]. However, the rank of $S_W^{-1}S_B$ is restricted by the number of classes, meaning that the number of orthogonal projection directions is restricted by the number of classes. This restriction limits the effectiveness of LDA. The same limitation also exists for KDA.

In this paper, we design two new between-class scatter matrices to be used in FDA, with which more orthogonal projection directions can be found. The FDA with new between-class scatter matrices is called Modified FDA (MFDA). The effectiveness of MFDA is then compared to FDA in doing two microphone classification tasks. Microphone classification is targeted at recognizing the recording microphone of a speech recording. In the experiments, Gaussian Supervector (GSV) is used as the feature, which has been applied to speaker recognition [14][18][19], recording device recognition [20][21] and acoustic scene identification [22]. Support Vector Machine (SVM) is employed as the classifier.

This paper is organized as follows. In Section II, the formulations of Linear FDA (LDA) and Linear MFDA (MLDA) are described. In Section III, the formulations of Kernel FDA (KDA) and Kernel MFDA (MKDA) are described. In Section IV, the microphone speech datasets are introduced. In Section V, experimental results and discussions are presented. A conclusion is drawn in Section VI.

## II. LDA AND MLDA

### A. Linear Fisher Discriminant Analysis (LDA)

Suppose we have $N$ training feature vectors $\{x_1, x_2 \dots x_N\}$, the objective of LDA is to find a projection matrix $W$ by solving the optimization problem defined in (1), where $S_W$ is the within-class scatter matrix and $S_B$ is the between-class scatter matrix [17] as given by (2) and (3). The relationship between $x_n$ and its projected version $y_n$ is $y_n=W^T x_n$. Let the $i$-th column of $W$ be denoted as $w_i$, which is a projection direction.

$$\max \quad J(W) = Trace\left\{\left(W^T S_W W\right)^{-1} W^T S_B W\right\} \tag{1}$$

In (2) and (3), $K$ is the number of classes, $N_k$ is the number of training feature vectors in class $k$, and $C_k$ denotes the set containing all the training feature vectors in class $k$.

The work described in this paper was substantially supported by a grant from The Hong Kong Polytechnic University (Project Account Code: RUG7).

$$S_W = \sum_{k=1}^{K} \sum_{x_n \in C_k} (x_n - m_k)(x_n - m_k)^T \quad (2)$$

$$S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T \quad (3)$$

where

$$m_k = \frac{1}{N_k} \sum_{x_n \in C_k} x_n \quad (4)$$

$$m = \frac{1}{N} \sum_{n=1}^{N} x_n = \frac{1}{\sum_{k=1}^{K} N_k} \sum_{k=1}^{K} N_k m_k \quad (5)$$

As shown in [17], $w_i$ is an eigenvector of $S_W^{-1} S_B$. Since the rank of $S_B$ is at most $K-1$, LDA can only find at most $K-1$ orthogonal projection directions [17].

### B. Modified Linear Fisher Discriminant Analysis (MLDA)

To increase the number of projection directions, we design two new between-class scatter matrices $S'_B$ and $S''_B$ to be used in MLDA, as defined in (6) and (7) respectively. MLDA with $S'_B$ being the between-class scatter matrix is called MLDA v1 (version 1), while MLDA with $S''_B$ is called MLDA v2.

$$\begin{aligned} S'_B &= \sum_{n=1}^{N} (x_n - m)(x_n - m)^T \\ &= \sum_{k=1}^{K} \sum_{x_n \in C_k} (x_n - m)(x_n - m)^T \end{aligned} \quad (6)$$

$$\begin{aligned} S''_B &= \sum_{n=1}^{N} \frac{1}{K-1} \sum_{k=1, x_n \notin C_k}^{K} (x_n - m_k)(x_n - m_k)^T \\ &= \sum_{k=1}^{K} \sum_{x_n \notin C_k} \frac{1}{K-1} (x_n - m_k)(x_n - m_k)^T \end{aligned} \quad (7)$$

As can be seen from (6) and (7), $S'_B$ and $S''_B$ can be used as a metric of the degree of separation of different classes. In addition, the rank of $S'_B$ or $S''_B$ is determined by the number of data, while that of $S_B$ is determined by the number of classes. This means that MLDA may find more projection directions than LDA. Similar to LDA, the solution of MLDA v1 is the collection of the eigenvectors of $S_W^{-1} S'_B$, and the solution of MLDA v2 is the collection of the eigenvectors of $S_W^{-1} S''_B$.

## III. KDA AND MKDA

### A. Kernel Fisher Discriminant Analysis (KDA)

As shown in [1] and [16], $w_i$ should be a linear combination of the training vectors, namely $w_i = \sum_{n=1}^{N} x_n (v_i)_n$, where $v_i$ is an $N\times1$ vector containing the coefficients in the linear combination, and $(v_i)_n$ is its $n$-th element. Furthermore, if we define a new matrix $V$ whose $i$-th column is $v_i$, then we can express $W^T m_k$, $W^T m$ and $W^T x_n$ in terms of $V$, as given by (8), where $l_k$, $l$ and $q_n$ are $N\times1$ vectors, whose $j$-th elements are given by (9), and $k(x_j, x_n)$ is a kernel function [16].

$$W^T m_k = V^T l_k, \quad W^T m = V^T l, \quad W^T x_n = V^T q_n \quad (8)$$

where

$$(l_k)_j = \frac{\sum_{x_n \in C_k} k(x_j, x_n)}{N_k}, \quad (l)_j = \frac{\sum_{n=1}^{N} k(x_j, x_n)}{N}, \quad (q_n)_j = k(x_j, x_n) \quad (9)$$

Then we can express $W^T S_W W$ and $W^T S_B W$ as in (10) and (11), where $U_W$ involved in (10) and $U_B$ involved in (11) are defined in (12) and (13) respectively.

$$\begin{aligned} W^T S_W W &= \sum_{k=1}^{K} \sum_{x_n \in C_k} W^T (x_n - m_k)(x_n - m_k)^T W \\ &= \sum_{k=1}^{K} \sum_{x_n \in C_k} V^T (q_n - l_k)(q_n - l_k)^T V = V^T U_W V \end{aligned} \quad (10)$$

$$\begin{aligned} W^T S_B W &= \sum_{k=1}^{K} N_k W^T (m_k - m)(m_k - m)^T W \\ &= \sum_{k=1}^{K} N_k V^T (l_k - l)(l_k - l)^T V = V^T U_B V \end{aligned} \quad (11)$$

where

$$U_W = \sum_{k=1}^{K} \sum_{x_n \in C_k} (q_n - l_k)(q_n - l_k)^T \quad (12)$$

$$U_B = \sum_{k=1}^{K} N_k (l_k - l)(l_k - l)^T \quad (13)$$

Adopting the expressions in (10) and (11), (1) becomes (14), which is KDA. Similar to LDA, the solution of KDA is the collection of the eigenvectors of $U_W^{-1} U_B$ [16].

$$\max \quad J(V) = Trace\left\{ \left(V^T U_W V\right)^{-1} V^T U_B V \right\} \quad (14)$$

### B. Modified Kernel Fisher Discriminant Analysis (MKDA)

Similar to KDA, $W^T S'_B W$ (for MKDA v1) and $W^T S''_B W$ (for MKDA v2) can also be expressed in terms of $V$, which is given by (15) and (16) respectively, where $U'_B$ and $U''_B$ involved in the equations are given by (17) and (18) respectively. Then the solution of MKDA v1 is the collection of the eigenvectors of $U_W^{-1} U'_B$, while that of MKDA v2 is the collection of the eigenvectors of $U_W^{-1} U''_B$.

$$\begin{aligned} W^T S'_B W &= \sum_{k=1}^{K} \sum_{x_n \in C_k} W^T (x_n - m)(x_n - m)^T W \\ &= \sum_{k=1}^{K} \sum_{x_n \in C_k} V^T (q_n - l)(q_n - l)^T V = V^T U'_B V \end{aligned} \quad (15)$$

$$\begin{aligned} W^T S''_B W &= \sum_{k=1}^{K} \sum_{x_n \notin C_k} \frac{1}{K-1} W^T (x_n - m_k)(x_n - m_k)^T W \\ &= \sum_{k=1}^{K} \sum_{x_n \notin C_k} \frac{1}{K-1} V^T (q_n - l_k)(q_n - l_k)^T V = V^T U''_B V \end{aligned} \quad (16)$$

where

$$U'_B = \sum_{k=1}^{K} \sum_{x_n \in C_k} (q_n - l)(q_n - l)^T \quad (17)$$

$$U''_B = \sum_{k=1}^{K} \sum_{x_n \notin C_k} \frac{1}{K-1} (q_n - l_k)(q_n - l_k)^T \quad (18)$$

TABLE I. AHUMADA-25 MICROPHONE SPEECH DATASET

| Notation | Microphone Model | Number of Speech Samples | |
|---|---|---|---|
| | | *Training* | *Testing* |
| M1 | AKG C410B Head Mounted | 240 | 260 |
| M2 | AKH D80S Desktop | 240 | 260 |
| M3 | SONY ECM 66B Lapel | 240 | 260 |
| M4 | TARGET Lapel | 240 | 260 |
| UBM | All the models | 599 | |

TABLE II. GAUDI-25 MICROPHONE SPEECH DATASET

| Notation | Microphone Model | Number of Speech Samples | |
|---|---|---|---|
| | | *Training* | *Testing* |
| M1 | AKG C410 Desktop | 240 | 260 |
| M2 | AKG Tripower Desktop | 240 | 260 |
| M3 | AKH D80S Desktop | 240 | 260 |
| M4 | SONY ECM 66B Lapel | 240 | 260 |
| M5 | TARGET CPT3GX Desktop | 240 | 240 |
| UBM | All the models | 744 | |

## IV. SPEECH DATASETS AND EXPERIMENTAL SETTINGS

In this paper, we utilize two microphone speech datasets to compare the performance of FDA (including LDA and KDA) and MFDA (including MLDA v1 and v2, MKDA v1 and v2). One dataset is Ahumada-25, comprising 4 different microphones for classification; the other is Gaudi-25, comprising 5 different microphones for classification. Both datasets come from AHUMADA speech corpus [23]. Each dataset is divided into a training set, a testing set, and a set for Universal Background Model (UBM), as shown in Table I and Table II. For Ahumada-25, totally 960 microphone speech samples are used for training, 1040 microphone speech samples are used for testing, and 599 microphone speech samples are used for UBM. For Gaudi-25, totally 1200 microphone speech samples are used for training, 1280 microphone speech samples are used for testing, and 744 microphone speech samples are used for UBM. The feature representing a speech sample is Gaussian Supervector (GSV), which is calculated by adapting a 32-mixture UBM with a relevance factor 5 [15][19][20]. The classifier is linear Support Vector Machine (SVM) [24].

On using FDA and MFDA, to prevent singularity, a regularization term $\varepsilon I$ is added to $S_W$ and $U_W$, where $\varepsilon$ is a non-negative value and $I$ is an identity matrix. Adding a regularization term is a simple yet effective way to prevent singularity [25]. Therefore, we find the eigenvectors of $(S_W+\varepsilon I)^{-1}S_B$ (for LDA), $(S_W+\varepsilon I)^{-1}S'_B$ (for MLDA v1), $(S_W+\varepsilon I)^{-1}S''_B$ (for MLDA v2), $(U_W+\varepsilon I)^{-1}U_B$ (for KDA), $(U_W+\varepsilon I)^{-1}U'_B$ (for MKDA v1), $(U_W+\varepsilon I)^{-1}U''_B$ (for MKDA v2). On using KDA or MKDA, linear kernel $k(a,b)=a^T b$ and Gaussian kernel $k(a,b)=e^{-(a-b)^T(a-b)/h}$ are employed.

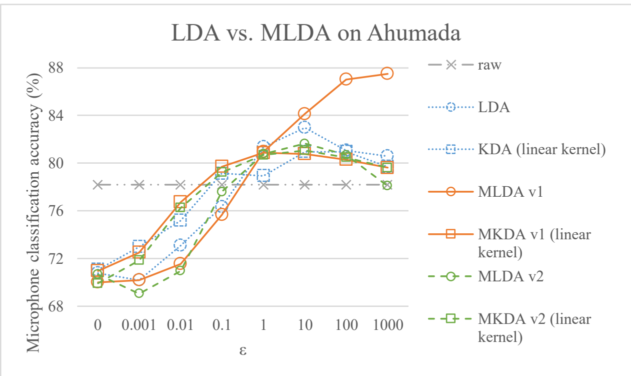


Fig. 1. Classification results on Ahumada-25 using LDA, linear kernel KDA, MLDA, and linear kernel MKDA.

## V. EXPERIMENTAL RESULTS AND DISCUSSIONS

### *A. Classification Results and Discussions on Ahumada-25*

Microphone classification results on Ahumada-25 are shown in Figs. 1 and 2. Fig. 1 shows the performances of linear projection techniques (i.e. LDA, MLDA, linear kernel KDA and MKDA), while Fig. 2 shows the performances of nonlinear projection techniques (i.e. Gaussian kernel KDA and MKDA). The polyline labelled with "raw" is the result without applying any projection technique. An interesting observation is that, the regularization parameter $\varepsilon$ plays an important role; when $\varepsilon$ is small, the linear projection techniques may not take effect (Fig. 1). With a large value of $\varepsilon$, MLDA v1 can outperform other linear projection techniques, which demonstrates the effectiveness of the new scatter matrix involved in MLDA v1, while MLDA v2 merely gives similar performance to LDA. From Fig. 2, performance improvement is also gained on applying the nonlinear projection techniques, with a good choice of the kernel parameter $h$. Particularly, both MKDA v1 and v2 can outperform KDA.

### *B. Classification Results and Discussions on Gaudi-25*

Microphone classification results on Gaudi-25 are shown in Figs. 3 and 4. Fig. 3 shows the performances on using linear projection techniques, while Fig. 4 shows the performances on using nonlinear projection techniques. It can be observed that, the effectiveness of the projection techniques highly depends on the regularization parameter $\varepsilon$, and only when $\varepsilon$ is properly chosen, can the projection techniques take effect. In general, MFDA is less dependent on $\varepsilon$ (i.e. more stable) and can

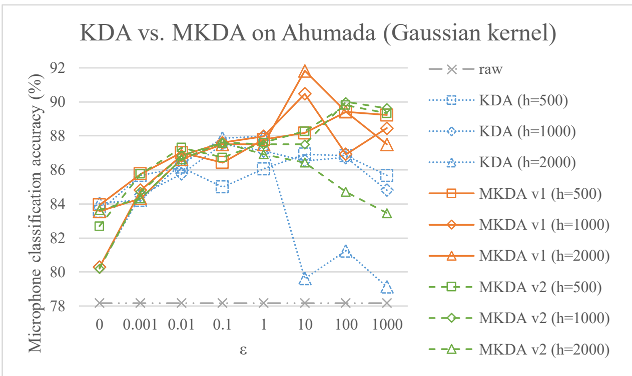


Fig. 2. Classification results on Ahumada-25 using Gaussian kernel KDA and Gaussian kernel MKDA.

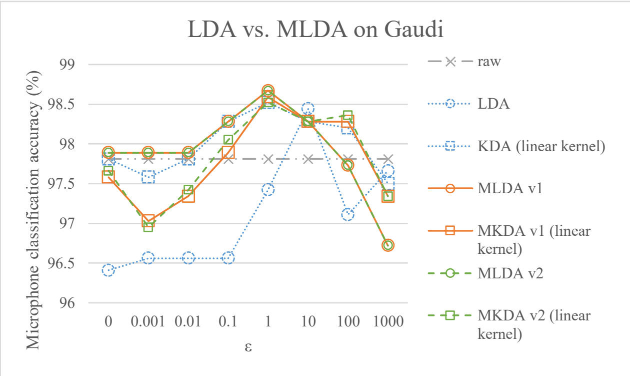


Fig. 3. Classification results on Gaudi-25 using LDA, linear kernel KDA, MLDA, and linear kernel MKDA.

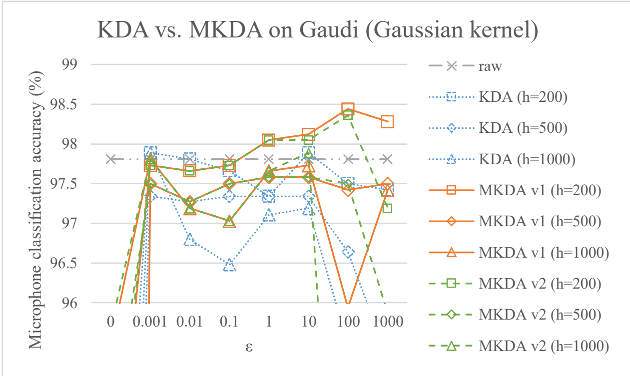


Fig. 4. Classification results on Gaudi-25 using Gaussian kernel KDA and Gaussian kernel MKDA.

perform better, than FDA. Comparing Fig. 3 and Fig. 4, the linear projection techniques and nonlinear projection techniques give similar improvement over the raw feature; however, the nonlinear projection techniques are also highly affected by the kernel parameter $h$.

### C. *Dimensionality Reduction Ability of LDA and MLDA*

In this part, the ability of LDA and MLDA as a dimensionality reduction technique is compared. Full dimensionality requires including all the column vectors in the projection matrix $W$, while the reduced dimensionality only involves a few of the column vectors in $W$. By sorting the column vectors in $W$ according to the significance (as each column vector is an eigenvector, the larger the eigenvalue, the more significant the column vector will be), only a few of the significant projection directions may be good enough.

Microphone classification results with different dimensionalities of the projected features are shown in Figs. 5 and 6, where the dimensionality varies from 768 (i.e. all the column vectors are included) to 76 (i.e. only the first 76 significant column vectors are included). It can be observed that the performance of MLDA v1 is quite stable with respect to different dimensionalities, and even when the dimensionality of the projected feature is small (e.g. 76), the classification accuracy is still high. On Ahumada-25 dataset (Fig. 5), the performance of LDA and MLDA v2 tends to drop with the decrease of the dimensionality of the projected features. On Gaudi-25 dataset (Fig. 6), the performance of LDA tends to improve even with the decrease of the dimensionality of the projected features, while the performance of MLDA v1 and v2 are relatively stable. If the raw feature is noisy, probably only the first several significant projection directions are effective while the rest projection directions are not only useless but also interfering, so reducing the dimensionality (i.e. only keeping the most significant projection directions) may improve the quality of the projected feature.

## VI. Conclusion

In this paper, we propose Modified Fisher Discriminant Analysis (MFDA), which uses different between-class scatter matrices from what the traditional Fisher Discriminant Analysis (FDA) does. Compared to FDA, MFDA can find more orthogonal projection directions. The performance of utilizing MFDA and FDA as a projection technique, is compared in doing two audio signal classification tasks. Experimental results show that MFDA gains higher classification accuracy than FDA, which demonstrates its effectiveness. Another finding is that, the performance of both MFDA and FDA is highly dependent on the regularization parameter. Moreover, we also compare the effectiveness of using the linear version of MFDA (i.e. MLDA) and the linear version of FDA (i.e. LDA) as a dimensionality reduction technique. Experimental results demonstrate that MLDA can be more stable than LDA when the dimensionality of the projected feature varies.

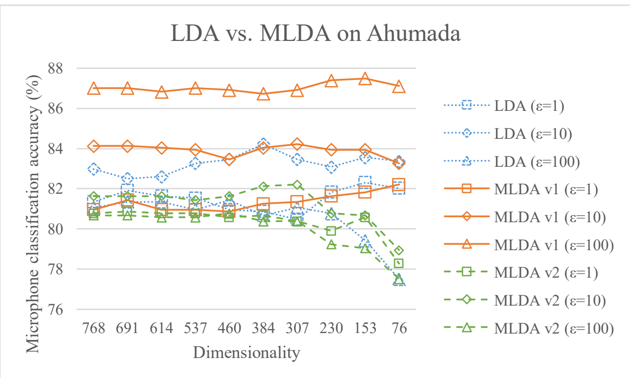


Fig. 5. Classification results on Ahumada-25 using projected features with different dimensionalities.

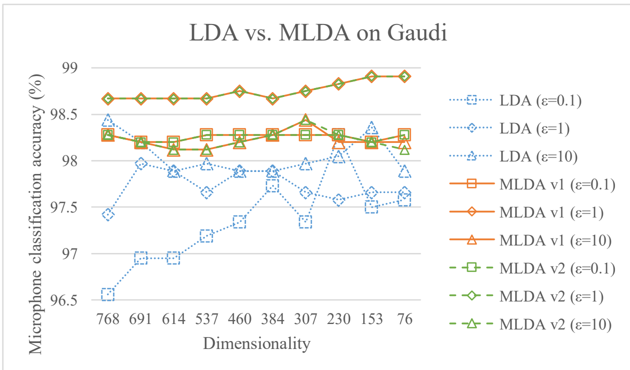


Fig. 6. Classification results on Gaudi-25 using projected features with different dimensionalities.

## References


[1] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller, "Fisher discriminant analysis with kernels," in *Proc. IEEE Neural Networks for Signal Processing Workshop*, 1999, pp. 41-48.

[2] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Computation*, vol. 12, no. 10, pp. 2385-2404, 2000.

[3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711-720, 1997.

[4] M. H. Yang, "Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods," in *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 2002.

[5] J. Yang, A. F. Frangi, J. Y. Yang, D. Zhang, and Z. Jin, "KPCA plus LDA: a complete kernel fisher discriminant framework for feature extraction and recognition," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 2, pp. 230-244, 2005.

[6] X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *Proc. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2004.

[7] S. Yan *et al.*, "Multilinear discriminant analysis for face recognition," *IEEE Trans. on Image Processing*, vol. 16, no. 1, pp. 212-220, 2007.

[8] S. M. Chu, H. Tang, and T. S. Huang, "Fishervoice and semi-supervised speaker clustering," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2009, pp. 4089-4092.

[9] Z. Li, W. Jiang, and H. Meng, "Fishervoice: a discriminant subspace framework for speaker recognition," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 4522-4525.

[10] C. Zhang and T. F. Zheng, "A fishervoice based feature fusion method for short utterance speaker recognition," in *Proc. IEEE China Summit & International Conference on Signal and Information Processing (ChinaSIP)*, 2013, pp. 165-169.

[11] N. Li, W. Jiang, H. Meng, and Z. Li, "Clustering similar acoustic classes in the fishervoice framework," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2013, pp. 7726-7730.

[12] Y. Jiang and F. H. F. Leung, "Generalized Fisher discriminant analysis as a dimensionality reduction technique," in *Proc. Int. Conf. on Pattern Recognition (ICPR)*, 2018.

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

[14] J. H. L. Hansen and T. Hasan, "Speaker recognition by machines and humans: a tutorial review," *IEEE Signal Processing Magazine*, vol. 32, no. 6, pp. 74-99, 2015.

[15] Y. Jiang and F. H. F. Leung, "Using regularized Fisher discriminant analysis to improve the performance of Gaussian supervector in session and device identification," in *Proc. IEEE Int. Joint Conf. on Neural Networks (IJCNN)*, 2017, pp. 705-712.

[16] Y. Jiang and F. H. F. Leung, "Using double regularization to improve the effectiveness and robustness of Fisher discriminant analysis as a projection technique," in *Proc. IEEE Int. Joint Conf. on Neural Networks (IJCNN)*, 2018, pp. 3275-3281.

[17] C. M. Bishop, "Linear models for classification," in *Pattern Recognition and Machine Learning*, Springer, 2006, ch. 4, pp. 179-224.

[18] Y. Jiang and F. H. F. Leung, "The scalable version of probabilistic linear discriminant analysis and its potential as a classifier for audio signal classification," in *Proc. IEEE Int. Joint Conf. on Neural Networks (IJCNN)*, 2018, pp. 72-78.

[19] W. M. Campbell, D. E. Sturim, and D. A. Reynolds, "Support vector machines using GMM supervectors for speaker verification," *IEEE Signal Processing Letters*, vol. 13, no. 5, pp. 308-311, 2006.

[20] D. G. Romero and C. Y. E. Wilson, "Automatic acquisition device identification from speech recordings," in *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010, pp. 1806-1809.

[21] C. L. Kotropoulis, "Source phone identification using sketches of features," *IET Biometrics*, vol. 3, no. 2, pp. 75-83, 2014.

[22] Y. Jiang and F. H. F. Leung, "Comparison of supervector and majority voting in acoustic scene identification," in *Proc. IEEE Int. Conf. on Digital Signal Processing (DSP)*, 2018.

[23] J. O. Garcia, J. G. Rodriguez, and V. M. Aguiar, "AHUMADA: a large speech corpus in Spanish for speaker characterization and identification," *Speech Communication*, vol. 31, no. 2, pp. 255-264, 2000.

[24] C. C. Chang and C. J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-27, 2011.

[25] T. V. Bandos, L. Bruzzone, and G. C. Valls, "Classification of hyperspectral images with regularized linear discriminant analysis," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 862-873, 2009.