Detection of Vocal Fold Paralysis and Edema using Linear Discriminant Classifiers

Euthymius Ziogas and Constantine Kotropoulos

Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki 54124, Greece
thimiouc@otenet.gr, costas@zeus.csd.auth.gr

Abstract. In this paper, a two-class pattern recognition problem is studied, namely the automatic detection of speech disorders such as vocal fold paralysis and edema by processing the speech signal recorded from patients affected by the aforementioned pathologies as well as speakers unaffected by these pathologies. The data used were extracted from the Massachusetts Eye and Ear Infirmary database of disordered speech. The linear prediction coefficients are used as input to the pattern recognition problem. Two techniques are developed. The first technique is an optimal linear classifier design, while the second one is based on the dual-space linear discriminant analysis. Two experiments were conducted in order to assess the performance of the techniques developed namely the detection of vocal fold paralysis for male speakers and the detection of vocal fold edema for female speakers. Receiver operating characteristic curves are presented. Long-term mean feature vectors are proven very efficient in detecting the voice disorders yielding a probability of detection that may approach 100% for a probability of false alarm equal to 9.52%.

1 Introduction

Speech processing has proved to be an excellent tool for voice disorder detection. Among the most interesting recent works are those concerned with Parkinson's Disease (PD), multiple sclerosis (MS) and other diseases which belong to a class of neuro-degenerative diseases that affect patients speech, motor, and cognitive capabilities [1, 2]. Such studies are based on the special characteristics of speech for persons who exhibit disorders on voice and/ or speech. They aim at either evaluating the performance of special treatments (i.e. LSVT [2, 3]) or developing accessibility in communication services for all persons [4]. Thus, it would possibly be a matter of great significance to develop systems able to classify the incoming voice samples into normal or pathological ones before other procedures are further applied.

In this paper, we are concerned with vocal fold paralysis and vocal fold edema, which are both associated with communication deficits that affect the perceptual characteristics of pitch, loudness, quality, intonation, voice-voiceless contrast etc, having similar symptoms with PD and other neuro-degenerative

diseases [5]. In either case, a two-class pattern recognition problem is essentially studied.

Closely related previous works are the detection of vocal fold cancer [6], where a Hidden Markov Model (HMM)-based classifier was employed and the binary classification between normal subjects and subjects suffering from different pathologies in [7], where Mel frequency cepstral coefficients and pitch were used as features for classification that was performed by the linear discriminant classifier, the nearest mean classifier, and classifiers based on Gaussian mixture models or HMMs. Three parameters namely the number of discrimination, the level of clustering, and the average clustering were assessed for disease discrimination based on acoustic features in [8]. The performance of Fisher's linear classifier, the K-nearest neighbor classifier, and the nearest mean one for vocal fold paralysis and vocal fold edema was assessed in [9]. An attempt is presented to identify pathological disorders of the larynx such as vocal fold paralysis using wavelet analysis and multilayer neural networks in [10]. The detection of certain voice pathologies from the cepstral content of the mucosal wave that is reconstructed by inverse filtering based on findings from the behavior of a 2 m vocal cord model is discussed in [11].

In this paper, two techniques based on linear classifiers are developed. The first one is a sample-based optimal linear classifier design, while the second one is based on the dual-space linear discriminant analysis. The work presented in this paper extends previously reported results in [9]. We are not interested in the detection of pathological speech as in [7], but in the assessment of the discriminatory capability of the aforementioned classifiers for detecting vocal fold paralysis in male speakers and the detection of vocal fold edema in female speakers. The pattern recognition experiments were conducted by employing either frame-based 14th order linear prediction coefficients or their long-term mean vectors for each speaker. Leave-one-out estimates of the probability of false alarm and the probability of detection are derived and receiver operating characteristic (ROC) curves are demonstrated.

The outline of the paper is as follows. Section 2 describes the design of sample-based linear parametric classifiers. The design of dual space linear discriminant classifiers is discussed in Section 3. The data-set used is presented in Section 4 along with the feature extraction. Experimental results are reported in Section 5 and conclusions are drawn in Section 6.

2 Sample-Based Linear Parametric Classifiers

We focus on a two-class pattern recognition problem. Let X denote a sample (i.e. a feature vector). In this paper, linear parametric classifiers are studied regardless of the pattern distributions and hence the decision rule is of the form

$$h(X) = V^T X + v_0 \begin{array}{c} \Omega_1 \\ < \\ > \\ \Omega_2 \end{array} 0, \tag{1}$$

where V is the classifier coefficient vector, v_0 is the threshold, and Ω_i, $i = 1, 2$ denote the two classes. The optimal linear classifier is of the form [12]:

$$V = [s\Sigma_1 + (1-s)\Sigma_2]^{-1}(M_2 - M_1), \tag{2}$$

where Σ_i is the covariance matrix of the samples that belong to class Ω_i and M_i is the corresponding mean vector. The optimal linear parametric classifier can be designed using the iterative Algorithm 1.

Algorithm 1. Linear parametric classifier design

Step 1: Divide the available samples into two groups namely the *design sample set* and the *test sample set*.

Step 2: Using the design samples, compute the sample mean $\widehat{M}_i$ and the sample covariance matrix $\widehat{\Sigma}_i$, $i = 1, 2$.

Step 3: Change s from 0 to 1.

Step 4: Calculate V for a given s by $V = [s\widehat{\Sigma}_1 + (1-s)\widehat{\Sigma}_2]^{-1}(\widehat{M}_2 - \widehat{M}_1)$.

Step 5: Using the coefficient vector V obtained in Step 4, compute $y_j^{(i)} = V^T X_j^{(i)}$, for $i = 1, 2$ and $j = 1, 2, \ldots, N$, where $X_j^{(i)}$ is the jth test sample in the class Ω_i.

Step 6: The scalar values $y_j^{(1)}$ and $y_j^{(2)}$ that do not satisfy $y_j^{(1)} < -v_0$ and $y_j^{(2)} > -v_0$ are counted as classification errors. Changing v_0 from $-\infty$ to $+\infty$ find v_0 that yields the smallest classification error.

Step 7: Record the classification error determined in Step 6 and go to Step 3.

Algorithm 1 makes no assumption concerning the distributions of the feature vectors X. It is known as *holdout method* and produces a *pessimistic bias* in estimating the classification error. If Step 1 is omitted and the classifier is designed using all the available samples and tested on the same samples in Step 5, then the so called *resubstitution method* results. The latter method produces an *optimistic bias* in estimating the classification errors. As the number of samples increases towards ∞, both the bias of the holdout method and that of the resubstitution method are reduced to zero. As far as the parameters are concerned, we can get better estimates by using a larger number of samples. However, in most cases, the number of the available samples is fixed.

3 Dual Space Linear Discriminant Analysis

A Dual Space Linear Discriminant Analysis algorithm was proposed for face recognition in [13]. In contrast to the linear parametric classifier described in Section 2, this algorithm is not restricted to a two-class problem.

Let the training set contain L classes and each class Ω_i, $i = 1, 2, \ldots, L$ have n_i samples. Then the within class scatter matrix S_w and the between class scatter matrix S_b are defined as

$$S_w = \sum_{i=1}^{L} \sum_{X_k \in \Omega_i} (X_k - M_i)(X_k - M_i)^T \tag{3}$$

$$S_b = \sum_{i=1}^{L} n_i (M_i - M)(M_i - M)^T \tag{4}$$

where M is the gross-mean of the whole training set and M_i, $i = 1, 2, \ldots, L$ are the class centers for Ω_i, $i = 1, 2, \ldots, L$. The Dual Space Linear Discriminant Analysis is summarized in Algorithm 2.

Algorithm 2. Dual space linear discriminant classifier design
At the *design (training) stage*:

Step 1: Compute S_w and S_b using the design set.

Step 2: Apply principal component analysis (PCA) to S_w and compute the principal subspace F defined by the K eigenvectors $V = [\Phi_1|\Phi_2|...|\Phi_k]$ and its complementary subspace $\overline{F}$. Estimate the average eigenvalue ρ in $\overline{F}$.

Step 3: All class centers are projected onto F and are normalized by the K eigenvalues. Then S_b is transformed to

$$K_b^P = \Lambda^{-\frac{1}{2}} V^T S_b V \Lambda^{-\frac{1}{2}}, \tag{5}$$

where $\Lambda = \text{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_K\}$ is the diagonal matrix of the K largest eigenvalues that are associated with F. Apply PCA to K_b^P and compute the l_P eigenvectors Ψ_P of K_b^P with the largest eigenvalues. The l_P discriminative eigenvectors in F are defined as

$$W_P = V \Lambda^{-\frac{1}{2}} \Psi_P. \tag{6}$$

Step 4: Project all the class centers to $\overline{F}$ and compute the reconstruction difference as

$$A_r = (I - VV^T)A, \tag{7}$$

where $A = [M_1|M_2|...|M_L]$ is a matrix whose columns are the class centers. A_r is the projection of A onto $\overline{F}$. In $\overline{F}$, S_b is transformed to

$$K_b^C = (I - VV^T) S_b (I - VV^T). \tag{8}$$

Compute the l_C eigenvectors of K_b^C with the largest eigenvalues Ψ_C. The l_C discriminative eigenvectors in $\overline{F}$ are defined as

$$W_C = (I - VV^T)\Psi_C. \tag{9}$$

At the *test stage*:

Step 1: All class centers M_j, $j = 1, 2, \ldots, L$ as well as the test samples X_t are projected to the discriminant vectors in F and $\overline{F}$ yielding

$$a_j^P = W_P^T M_j \tag{10}$$

$$a_j^C = W_C^T M_j \tag{11}$$

$$a_t^P = W_P^T X_t \tag{12}$$

$$a_t^C = W_C^T X_t. \tag{13}$$

Step 2: The test sample X_t is assigned to the class

$$j^* = \arg\min_{j=1}^{L} \left\{ \|a_j^P - a_t^P\|^2 + \frac{1}{\rho} \|a_j^C - a_t^C\|^2 \right\}. \tag{14}$$

4 Datasets and Feature Extraction

Due to the inherent differences of the speech production system for each gender, it makes sense to deal with disordered speech detection for male and female speakers separately. In the first experiment that concerns vocal fold paralysis detection, the dataset contains recordings from 21 males aged 26 to 60 years who were medically diagnosed as normals and 21 males aged 20 to 75 years who were medically diagnosed with vocal fold paralysis. In the second experiment that concerns vocal fold edema detection, 21 females aged 22 to 52 years who were medically diagnosed as normals and 21 females aged 18 to 57 years who were medically diagnosed with vocal fold edema served as subjects. The subjects might suffer from other diseases too, such as hyperfunction, ventricular compression, atrophy, teflon granuloma, etc. All subjects were assessed among other patients and normals at the MEEI [14] in different periods between 1992 and 1994. Two different kinds of recordings were made in each session: in the first recording the patients were called to articulate the sustained vowel "Ah" (/a/) and in the second one to read the "Rainbow Passage". The former recordings are those employed in the present work. The recordings made at a sampling rate of 25 KHz in the pathological case, while at a rate of 50 KHz in the normal case. In the latter case, the sampling rate was reduced to 25 KHz by down-sampling. The aforementioned datasets are the same used in [9]. However, in this work more frames are considered per speaker utterance.

As in [9], 14 linear prediction coefficients were extracted for each speech frame [15]. The speech frames have a duration of 20 ms. Neighboring frames do not possess any overlap. Both the rectangular and the Hamming window are used to extract the speech frames. In the first experiment, the sample set consists of 4236 14-dimensional feature vectors (3171 samples from normal speech and another 1065 samples from disordered speech) for male speakers. In the second experiment, the sample set consists of 4199 14-dimensional feature vectors (3096 samples from normal speech and another 1103 samples from disordered speech vectors) for female speakers.

Besides the frame-based feature vectors, the 14-dimensional mean feature vectors for each speaker utterance are also calculated. By doing so, a dataset of 21 long-term feature vectors from males diagnosed as normal and another 21 long-term feature vectors from males diagnosed with vocal fold paralysis is created in the first experiment. Similarly, a dataset of 21 long-term feature vectors from females diagnosed as normal and another 21 long-term feature vectors from females diagnosed with vocal fold edema is collected in the second experiment.

5 Experimental Results

The assessment of the classifiers studied in the paper was done by estimating the probability of false alarm and the probability of detection using the just described feature vectors and the leave-one-out (LOO) method. The probability

of detection P_d is defined as

$$P_d = \frac{\#\text{ correctly classified pathological samples}}{\#\text{ pathological samples}} \tag{15}$$

and the probability of false alarm P_f is given by

$$P_f = \frac{\#\text{ normal samples misclassified as pathological ones}}{\#\text{ normal samples}}. \tag{16}$$

where # stands for number. There is no difficulty in the application of the LOO concept for long-term feature vectors. However, for frame-based feature vectors, the LOO method that excludes just one feature vector associated to speaker $\mathcal{S}$ leaves another $N_{\mathcal{S}}-1$ feature vectors of this speaker in the design set, where $N_{\mathcal{S}}$ is the number of feature vectors extracted from speaker $\mathcal{S}$ utterance. To guarantee that the test set is comprised of totally unseen feature vectors (i.e. samples), we apply the LOO method with respect to speakers and not the frame-based samples. Then the test set is comprised by feature vectors of the same speaker and a unique decision can be taken by assigning the test speaker to the class where the majority of the test feature vectors is classified to.

5.1 Sample-Based Linear Parametric Classifier

It is worth noting that for the linear parametric classifier the aforementioned probabilities of false alarm and detection are threshold-dependent. Accordingly, a ROC curve can be derived by plotting the probability of detection versus the probability of false alarm treating the threshold as an implicit parameter.

Vocal Fold Paralysis in Men

Frame-based feature vectors. Using the rectangular window and increments $\Delta s = 0.01$ Algorithm 1 yields the minimum total classification error 14.2857% for $s = 0.19$. The aforementioned classification error corresponds to a misclassification of 1 out of 21 normal utterances and 5 out of 21 disordered speaker utterances. The ROC curve is depicted in Figure 1a. Working in the same way with the Hamming window, Algorithm 1 yields the minimum total classification error for $s = 0.4$. However, in this case considerably more errors are committed in recognizing the normal patterns. Figure 1b depicts the corresponding ROC curve. By constraining the probability of false alarm at 10 %, the linear parametric classifier yields a probability of detection slightly higher than that achieved by the Fisher linear discriminant classifier [9].

Long-term feature vectors. If we design the classifier using the parameters derived by the LOO method on the frame-based feature vectors in order to classify the mean feature vectors per speaker, we obtain the ROC curve of Figure 1c, when the rectangular window is used. We see that the two classes are now considerably separable and we achieve a perfect classification for $P_f \approx 10\%$ that corresponds to 2 speakers. When the Hamming window is employed, the ROC curve plotted in Figure 1d results.

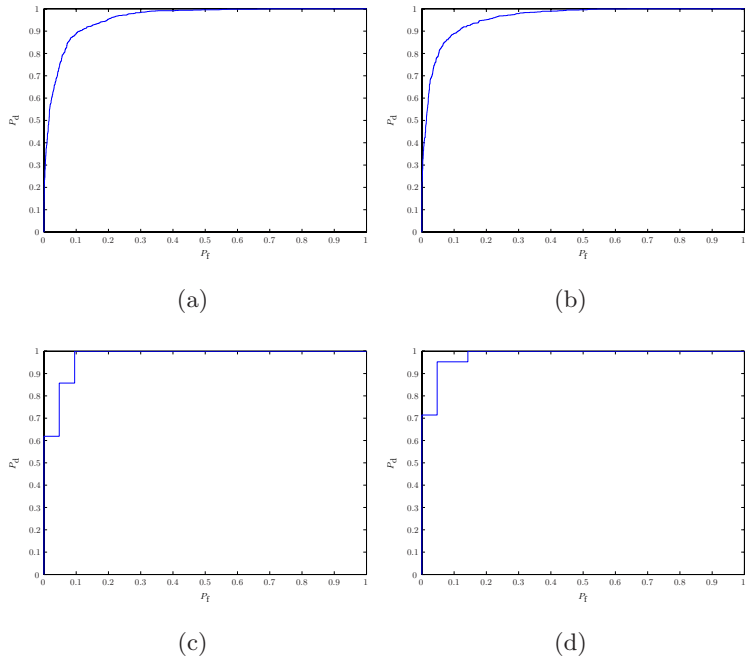

Fig. 1. Receiver operating characteristic curves of a linear parametric classifier designed to detect vocal fold paralysis in men using: (a) frame-based features and the rectangular window; (b) frame-based features and the Hamming window; (c) long-term feature vectors and the rectangular window; (d) long-term feature vectors and the Hamming window.

Vocal Fold Edema in Women

Frame-based feature vectors. Algorithm 1 yields the smallest total classification error of 9.5238% for $s = 0.92$ that corresponds to misclassification of 4 disordered speech utterances. Figure 2a depicts the ROC curve when the rectangular window is used. By comparing the ROC curves plotted in Figures 1a and 2a we notice that the classifier detects more efficiently vocal fold edema in women than vocal fold paralysis in men. A much better performance is obtained when the Hamming window replaces the rectangular one. The minimum classification error is only 7.1429% for $s = 0.84$, corresponding to misclassification of one normal and two disordered speech utterances. Indeed, the ROC curve of Figure 2b indicates a more accurate performance than that of Figure 2a. By constraining the probability of false alarm at 10 %, the linear parametric classifier yields a probability of detection 20% higher than that achieved by the Fisher linear discriminant classifier [9].

Long-term feature vectors. By using the rectangular window we can achieve a $P_d = 100\%$ for a misclassification of only one normal utterance, as can be seen in Figure 2c. The corresponding ROC is plotted in Figure 2d, when the Hamming window is used with the mean feature vectors.

From the ROC curves of Figure 2c and 2d, we notice that no false alarm can by obtained at the expense of only one misclassified disordered speech utterance. By allowing for 2 misclassifications of the normal utterances, we can obtain a perfect detection of vocal fold edema.

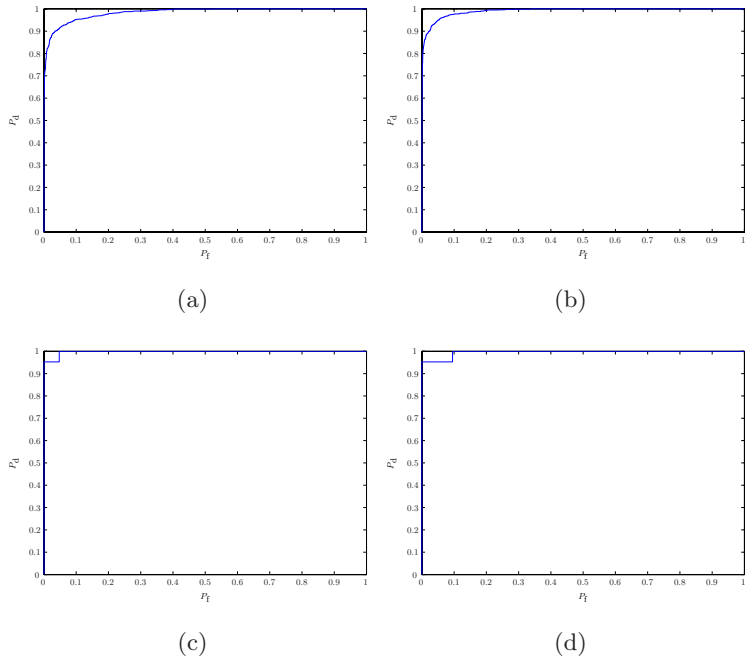

Fig. 2. Receiver operating characteristic curves of a linear parametric classifier designed to detect vocal fold edema in women using: (a) frame-based features and the rectangular window; (b) frame-based features and the Hamming window; (c) long-term feature vectors and the rectangular window; (d) long-term feature vectors and the Hamming window.

Tables 1 and 2 summarize the performance of the parametric classifier when frame-based features and long-term features are used, respectively.

Table 1. Performance of the parametric classifier for frame-based features (E_N stands for normal speech errors - i.e. false alarms and E_D stands for disordered speech errors - i.e. miss-detections).

Pathology	Window	E_N	P_f	E_D	P_d
Paralysis	Rectangular	1	4.761905%	5	76.1905%
Paralysis	Hamming	1	4.761905%	5	76.1905%
Edema	Rectangular	0	0%	4	80.952381%
Edema	Hamming	1	4.761905%	2	90.47619%

Table 2. Performance of the parametric classifier for long-term features (E_N stands for normal speech errors - i.e. false alarms and E_D stands for disordered speech errors - i.e. miss-detections).

Pathology	Window	E_N	P_f	E_D	P_d
Paralysis	Rectangular	0	0%	8	61.90476%
Paralysis	Rectangular	2	9.52381%	0	100%
Paralysis	Hamming	0	0%	6	71.4286%
Paralysis	Hamming	3	14.2857%	0	100%
Edema	Rectangular	0	0%	1	95.2381%
Edema	Rectangular	1	4.761905%	0	100%
Edema	Hamming	0	0%	1	95.2381%
Edema	Hamming	2	9.52381%	0	100%

5.2 Dual Space Linear Discriminant Classifier

Before applying the dual space linear discriminant classifier (Algorithm 2) to either frame-based or long-term feature vectors, we must note the following:

- We are interested in a two class problem, hence $L = 2$.
- Considering the ratio of the largest to the smallest eigenvalue of S_w in either case, we found that it was of the order of 10^3 or larger. For this reason, we shall consider as null subspaces the ones defined by eigenvectors associated with eigenvalues that are ten times larger than the smallest eigenvalue at most. S_w is a full rank matrix in any case. Thus, we obtain $K = 12$ and hence the dimension of the null subspace of S_w is equal to 2.
- In our case, S_b is a rank 1 matrix. Therefore, $l_P, l_C > 1$ does not make any sense and we choose that $l_P = l_C = 1$.
- The probabilities of detection and false alarm are not threshold-dependent. Accordingly, the classifier operates at a single point and no ROC curve is obtained.

Having clarified the above, we applied the dual space linear discriminant classifier to the detection of vocal fold paralysis in men and vocal fold edema in women. The results are summarized in Tables 3 and 4.

From the cross-examination of either Tables 1 and 3 or Tables 2 and 4, we conclude that the parametric classifier is more accurate than the dual space

Table 3. Dual space linear discriminant classifier applied to frame-based feature vectors (E_N stands for normal speech errors - i.e. false alarms and E_D stands for disordered speech errors - i.e. miss-detections).

Pathology	Window	E_N	P_f	E_D	P_d
Paralysis	Rectangular	1	4.761905%	7	66.6667%
Paralysis	Hamming	2	9.52381%	8	61.90476%
Edema	Rectangular	1	4.761905%	7	66.6667%
Edema	Hamming	5	23.80952%	4	80.952381%

Table 4. Dual space linear discriminant classifier applied to long-term feature vectors ((E_N stands for normal speech errors - i.e. false alarms and E_D stands for disordered speech errors - i.e. miss-detections).

Pathology	Window	E_N	P_f	E_D	P_d
Paralysis	Rectangular	2	9.52381%	7	66.6667%
Paralysis	Hamming	2	9.52381%	8	61.90476%
Edema	Rectangular	1	4.761905%	4	80.952381%
Edema	Hamming	4	19.04762%	4	80.952381%

linear discriminant classifier. For vocal fold paralysis, the use of the rectangular window yields better results than the use of the Hamming window. The opposite is true for vocal fold edema.

6 Conclusions

Two linear classifiers, namely the sample-based linear classifier and the dual space linear discriminant classifier have been designed for vocal fold paralysis detection in men and vocal fold edema detection in women. The experimental results indicate that the sample-based linear classifier achieves better results than the dual space linear discriminant classifier.

Acknowledgments

This work has been supported by the FP6 European Union Network of Excellence MUSCLE "Multimedia Understanding through Semantics, Computation and Learning" (FP6-507752).

References

1. F. Quek, M. Harper, Y. Haciahmetoglou, L. Chen, and L. O. Ramig, "Speech pauses and gestural holds in Parkinson 's Disease," in *Proc. 2002 Int. Conf. Spoken Language Processing*, 2002, pp. 2485–2488.
2. L. Will, L. O. Ramig, and J. L. Spielman, "Application of Lee Silverman Voice Treatment (LSVT) to individuals with multiple sclerosis, ataxic dysarthria, and stroke," in *Proc. 2002 Int. Conf. Spoken Language Processing*, 2002, pp. 2497–2500.

3. J. L. Spielman, L. O. Ramig, and J. C. Borod, "Oro-facial changes in Parkinson's Disease following intensive voice therapy (LSVT)," in *Proc. 2002 Int. Conf. Spoken Language Processing*, 2002, pp. 2489–2492.
4. V. Parsa and D. G. Jamieson, "Interactions between speech coders and disordered speech," *Speech Communication*, vol. 40, no. 7, pp. 365–385, 2003.
5. `www.emedicine.com/ent/byname/vocal-fold-paralysis-unilateral.htm`.
6. L. Gavidia-Ceballos and J. H. L. Hansen, "Direct speech feature estimation using an iterative EM algorithm for vocal fold pathology detection," *IEEE Trans. Biomedical Engineering*, vol. 43, pp. 373–383, 1996.
7. A. A. Dibazar, S. Narayanan, and T. W. Berger, "Feature analysis for automatic detection of pathological speech," in *Proc. Engineering Medicine and Biology Symposium 02*, 2002, vol. 1, pp. 182–183.
8. M. O. Rosa, J. C. Pereira, and M. Grellet, "Adaptive estimation of residue signal for voice pathology diagnosis," *IEEE Trans. Biomedical Engineering*, vol. 47, pp. 96–104, 2000.
9. M. Marinaki, C. Kotropoulos, I. Pitas, and N. Maglaveras, "Automatic detection of vocal fold paralysis and edema," in *Proc. 2004 Int. Conf. Spoken Language Processing*, 2004.
10. J. Nayak and P. S. Bhat, "Identification of voice disorders using speech samples," in *Proc. IEEE TenCon2003*, 2003, number 395.
11. P. Gómez, J. I. Godino, F. Rodríguez, F. Díaz, V. Nieto, A. Álvarez, and V. Rodellar, "Evidence of vocal cord pathology from the mucosal wave cepstral contents," in *Proc. 2004 IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, 2004, vol. 5, pp. 437–440.
12. K. Fukunaga, *Introduction in Statistical Pattern Recognition*, Academic Press, San Diego CA, 2nd edition, 1990.
13. X. Tang and W. Wang, "Dual space linear discriminant analysis for face recognition," in *Proc. 2004 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, 2004, pp. 1064–1068.
14. Voice and Speech Laboratory, Massachusetts Eye and Ear Infirmary, Boston MA, *Voice Disorders Database*, 1.03 edition, 1994, Kay Elemetrics Corp.
15. J. R. Deller, J. G. Proakis, and J. H. L. Hansen, *Discrete Time Processing of Speech Signals*, MacMillan Publishing Company, N. Y., 1993.