Asian Journal of Computer Science and Technology
ISSN: 2249 - 0701 Vol. 3 No. 1, 2014, pp. 20 - 23

A Pathological Voices Assessment Using Classification

T. Arikrishnan and C.P. Darani

Department of Computer Science and Engineering, Anna University (BIT Campus), Trichy, Tamil Nadu, India

E-mail: ari0244@gmail.com

(Received on 15 January 2014 and accepted on 28 March 2014)

Abstract - **The diagnosing of pathological voice may be a tedious topic and it receives abundant attention. There are several diseases that adversely have an effect on our human speech (voice).The doctor will use only the equipments for detection of pathological voice. However, it is invasive and needs a skilled analysis of diverse human speech signal parameters. Automatic voice analysis for pathological speech has its own blessings, like 1) its quantitative and non-invasive nature. 2) permitting the identification and observance of vocal system diseases . Within the pathological voice classification techniques gathered by the voice of a patient, the goal is to discriminate whether the given voice is normal or pathological. From the speech Mel-Frequency Cepstral Coefficients (MFCC) has been extracted from the voice information and classified into two categories. However, the accuracy of the earlier classification methodology may need additional improvement. In my project work, Support Vector Machine (SVM) classifier is used for pathological voice classification with non-invasive nature to diagnose and analyze the voice of the patient.**

Keywords: **Pathological voices, SVM**

I. Introduction

Dysphonia or pathological voice refers to speech problems resulting from damage to or malformation of the speech organs. Dysphonia is more common in people who use their voice professionally, for example, teachers, lawyers, salespeople, actors, and singers, and it dramatically effects these professional group's lives both financially and psychosocially. The presence of pathologies in the vocal folds affects the normal vibration pattern of the glottis and cause changes in voice quality. In the past 20 years, a significant attention has been paid to the science of voice pathology diagnostic and monitoring. The purpose of this work is to help patients with pathological problems for monitoring their progress over the course of voice therapy. Currently, patients are required to routinely visit a specialist to follow up their progress. Moreover, the traditional ways to diagnose voice pathology are subjective ,invasive methods such as the direct inspection of the vocal folds and the observation of the vocal folds by endoscopic instruments and depending on the experience of the specialist, different evaluations can be resulted. These techniques are expensive, risky, time consuming, discomfort to the patients and require costly resources, such as special light sources, endoscopic instruments and specialized video-camera equipment. In order to circumvent the above problems, non-invasive methods have been developed to help the ENT clinicians and speech therapists for early detection of vocal fold pathology and can improve the accuracy of the assessments.

Clinical measurement of pathological vocal function typically involves processing of acoustic, aerodynamic and stroboscopic data. Among these, acoustic measurements of voice are particularly appealing due to the simplicity and non-invasiveness of the measurement procedure. Acoustic analysis-based techniques are an effective tool for the objective support to vocal and voice disease screening and especially in their early detection and diagnosis.

Classically, a large amount of long-term parameters have been introduced to measure the quality and "degree of normality" of voice records, such as the pitch (fo), jitter, shimmer, amplitude perturbation quotient (APQ), pitch perturbation quotient (PPQ), harmonics to noise ratio (HNR), normalized noise energy (NNE), voice turbulence index (VTI), soft phonation index (SPI), frequency amplitude tremor (FATR), glottal to noise excitation ratio (GNE), and many others. But up to now rigorous studies about their application to large populations are lacking in order to fix the values related with normality for

every control group regarding sex and age. Yumoto [2] introduced the Harmonics-to-Noise Ratio (HNR) parameter which qualifies the amount of glottal noise in the vowel waveform, and showed that it can be an effective parameter for predicting pathological diseases. L.Eskenazi [3] tried to find the correlates of percent jitter parameter with the people's perceptual ratings of breathing. Klingholz [4] discussed the influence of the SNR ratio to the automatic detection of pathological voices [6]-[8] has also make some research of LP-residue, CPP. K.Umapathy [7] found octave of Time frequency transformation to deal with the problem of continuous voice disorder speech. Some research was based on traditional parameters from speech recognition. Godino-Llorente [9] used a multilayer perceptron (MLP) on mel frequency cepstral coefficients (MFCC) to achieve a classification rate of 96%. In this paper, SVM have been used to implement these experiments, and SVM had superior performance than GMM method.

II. Methodology

The system requires the digitized speech signal with their quality level. The short time acoustic characteristics are denoted by Mel Frequency Cepstral Coefficients. In order to discriminate the categories of speech signal based on quality, MFCC features are extracted to characterize the speech signal content. These feature are applied to the classifiers(SVM) for the classification of voices into normal voice and pathology voice.

III. Speech Data And Parametrization

The speech signal is recorded from 20 people (10 from normal peoples and 10 from pathological patients),at a rate of 8000 samples per second using sound recorder. Each person is asked to speak a same small sentence in Tamil.

Feature extraction of speech is one of the most important issues in the field of speech technology. There are two dominant acoustic measurements of speech signal. The first is the parametric modelling approach, developed to match closely the resonant structure of the human vocal tract that produces the corresponding speech sound. It is mainly derived from Linear Predictive analysis, such as LPC and LPC-based cepstrum (LPCC) [5]. The second is the nonparametric modelling method, that is basically originated from the human auditory perception system. FFT-based me1 frequency cepstral coefficients are used for this purpose and it is shown in Fig 1. The term M*el* refers to a kind of measurement related to perceived frequency. The mapping between the real frequency scale (Hz) and the perceived frequency scales *(Mels)* is approximately linear below 1 KHz and logarithmic at higher frequency [5]. The bandwidth of the critical band varies according to the perceived frequency. It is about linear up to 1 KHz and increases logarithmically above 1KHz. The suggested formula that models their relationship is described as follows:

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right)$$

The advantages are that, those parameters are capable of being immune to noise and it is easy to warp frequency into a non-uniform scale, such as Mel scale. An overview of the method is provided.

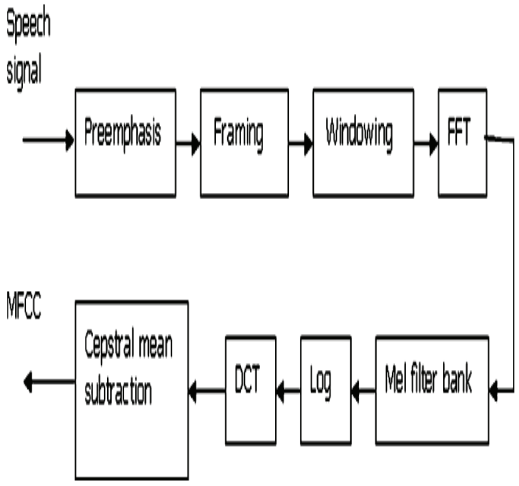

Fig.1Extraction of MFCC from speech signal

Overview of MFCC Algorithm

We assume that *y[n]* denotes the input speech signal. The complete calculation process of the coefficients can be described in the next four steps as follows:

Step 1: Transform the input speech signal from time domain to frequency domain by applying short-time Fast Fourier Transform (FFT) method

$$Y(\Omega) = \sum_{n=0}^{F-1} y[n] \cdot w[n] \cdot e^{-j2\pi \cdot n \cdot \frac{m}{F}}$$

where $m = 0,1,2 \ldots F\text{-}1$; F is frame size, which is generally equal to the power of 2; w $[n]$is the Hamming window function, which is based on the fact that the signal can be regarded as stationary and uninfluenced by the others within a short period of time, i.e. the frame size.

Step 2: Find the energy spectrum of each frame.

$$X(\Omega) = |Y(\Omega)|^2$$

Step 3: Calculate the energy in each Mel window.

$$S_k = \sum_{j=0}^{\frac{k}{2}-1} W_k(j) \cdot X(j)$$

where $1 \le k \le M$; M is the number of the Mel windows in Mel scale, which generally ranges from 20 to 24. $W_k(j)$: the triangular weighted function is associated with the k^{th} *M*el window in Mel scale.

Step 4: Proceeding with logarithm and cosine transforms, we can figure out the Mel Frequency Cepstral Coefficients:

$$mc_m = \sum_{k=1}^{M} \log(S_k) \cos\left[n \cdot (k - 0.5) \frac{\pi}{M} \right]$$

where $1 \le n \le L$; L is the desired order of MFCC.

IV. Experiments

In this paper, and SVM models are used to classify the given input signals.

SVM Classifier

Support vector machine (SVM) is based on the principle of structural risk minimization (SRM). Like RBFNN, support vector machines can be used for pattern classification and nonlinear regression. SVM constructs a linear model to estimate the decision function using non-linear class boundaries based on support vectors. If the data are linearly separated, SVM trains linear machines for an optimal hyperplane that separates the data without error and into the maximum distance between the hyperplane and the closest training points. The training points that are closest to the optimal separating hyperplane are called support vectors. SVM maps the input patterns into a higher dimensional feature space through some linear mapping chosen a priori. A linear decision surface is then constructed in this high dimensional feature space.

SVM Principle

Support vector machine (SVM) ban be used for classifying the obtained data (Burges, 1998). SVM are a set of related supervised learning methods used for classification and regression. They belong to a family of generalized linear classifiers. Let us denote a feature vector (termed as pattern) by x-(x_1, x_2,...., x_n) and its class label by y such that y = {+l, -l}. Therefore, consider the problem of separating the set of n-training patterns belonging to two classes,

$$(x_i, y_i), x_i \in R^n, \text{ y} = \{+1, -1\}, \ i = \ 1, 2, ..., n$$

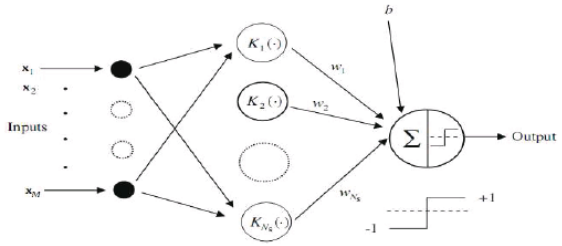

Fig. 2 Architecture of the SVM *(N_s* is the number of support vectors)

SVM Kernels

SVM generally applies to linear boundaries. In the case where a linear boundary is inappropriate SVM can map the input vector into a high dimensional feature space. By choosing a linear mapping, the SVM constructs an optimal separating hyperplane in this higher dimensional space.The function K is defined as the kernel function for generating the inner products to construct machines with different types of linear decision surfaces in the input space.

$$K(x, x_i) = \Phi(x) . \Phi(x_i)$$

The kernel function may be any of the symmetric functions that satisfy the Mercer's conditions (Courant and Hilbert, 1953).

An example for SVM kernel function $\Phi(x)$ maps 2-dimensional input space (x_1, x_2) to higher 3-dimensional feature space $x_1^2, x_2^2, \sqrt{2x_1x_2}$ as shown in SVM was originally developed for two class classification problems. The N class classification problem can be solved using N SVMs.

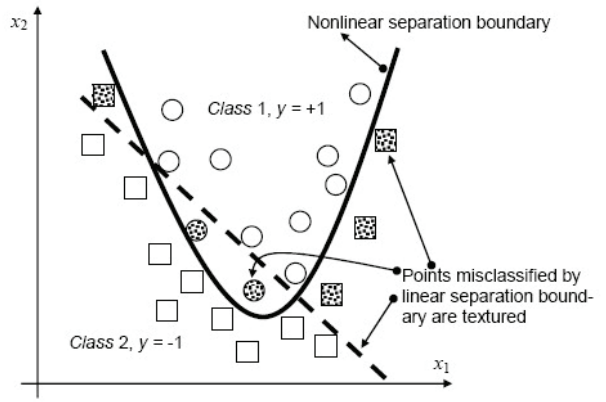

Fig.3 An example for SVM kernel function $(as) maps two dimensional input space to higher three dimensional feature space, (a) Nonlinear problem, (b) Linear problem.

V. Discussion and Conclusion

This classification between normal and pathological voice from mfcc parameter are extracted from voice samples and applied to svm classifier.

Since the total number of data was small,we tried to train and test the SVM model by splitting total data sets into two parts.Two thirds of the data were used for training and the remaining one third of data was used for test. In each stage,the Support Vector Machine were trained and tested separately.

References

[1] R.A.Prosek, A. A. Montgemery, B. E. Walden, "An evaluation of residue features as correlates of voice disorders," *J. Commun. Disorders,* Vol. 20, pp. 105-117, 1987.

[2] E.Yumoto,"Harmonics-to-noise ratio as an Index of the degree of horseness," *J.Acoust.Soc.Am.,* Vol 71, pp.1544-1549,1989.

[3] L.Eskenazi, D.G.Chinders, and D.M.Hicks,"Acoutics correlates of vocal quality", J.Speech Hear.Res., Vol. 33,pp. 298-306, 1990.

[4] F.Klingholz,"Acoustic recognition of voice disorders :A comparative study, running speech versus sustained vowels," *J. Acoust. Soc. Am.,* Vol.88, pp. 2218-2224, 1990.

[5] L. Rabiner and B. H. Huang, *Fundamentals of speech recognition*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[6] Y.Qi and R.E. Hillman, "Temporal and spectral estimations of harmonics to- noise ratio in human voice signals," *J.Acoust. Soc. Am.,* Vol. 102, No. 1, pp. 537-543,1997.

[7] K.Umapathy, S.Krishnan, V.Parsa, and D.Jamieson, "Time-frequency modelling and classification of pathological voices," in Proc. IEEE Engineering in Medicine and Biology Society (EMBS) 2002 Conference, Houston,TX,Oct.2002, pp. 116-117.

[8] Y.D.Heman-Ackah et al., "Cepstral peak prominence: A more reliable measure of dysphonia," *Ann Otol., Rhinol., Laryngol.,* Vol. 112, No. 4, pp. 324-329, Apr. 2003.

[9] Godino-Llorente and P.Gomez-Vilda, "Automatic detection of voice impairment by means of short-term cepstral parameters and neuralnetwork based detectors," IEEE Trans. Biomed. Eng., Vol.51, No.2, pp.380-384,Feb. 2004.

[10] J. Nayak, P.S. Bhat, R. Acharya, U.V. Athal, "Classification and analysis of speech abnormatities", *ITBM – RBM,* 26, pp. 319-327.2005.