Disease Detection using Analysis of Voice Parameters

Sonu and R.K. Sharma

Department of Electronics and Communication
National Institute of Technology Kurukshetra, Haryana, India
E-mail: sonu_27jan@yahoo.co.in, mail2drrks@gmail.com

Abstract
This paper investigates the adaptation of automatic speech recognition to disease detection by analyzing the voice parameters. The analysis of the voice allows the identification of the diseases which affect the vocal apparatus and currently is carried out from an expert doctor through methods based on the auditory analysis. This paper presents a novel method to keep track of patient's pathology: Easy to use, fast, non invasive for the patient and affordable for the clinician. This method uses parametric method (jitter, shimmer, harmonic to noise etc…) to evaluate the pathological voice. The method for this task also relies on Mel Frequency Cepstral Coefficient (MFCC) as feature extraction and Dynamic Time Warping (DTW) as feature Matching. The aim of the study is to evaluate the voice quality in patients with mild-to-acute asthma by parametric method and non parametric method. Comparative analysis is also done between parametric and non-parametric methods.

Keywords: MFCC, DTW, Dysphonia, Jitter, shimmer, HNR, Acoustic parameters

Introduction

Asthma as a chronic inflammatory disorder of the airways associated with increased airway hyper-responsiveness, recurrent episodes of wheezing, breathlessness, chest tightness, and coughing, particularly at night/early morning.Airway inflammation caused by allergies and asthma can hurt the sound quality of the voice. The vocal cords cover the larynx, the top part of the trachea. These mucus-covered muscular bands are the vibrating "strings" that produce voice sound, which is then filtered and shaped by the resonating cavities of the throat, nose, and mouth. Inflammation along the passageways from the nose down to the larynx can impair vocal quality. Bronchial asthma, labored breathing and wheezing, and allergies can also cause sore throat and inflammation around the vocal cords. Swollen, inflamed cords don't vibrate efficiently and can make the voice sound hoarse or scratchy [2]. Nearly half of the patients complain about permanent voice disorders. Any modification of above system may cause a qualitative and/or quantitative alteration of the voice, defined as dysphonia. Dysphonia can be due to both organic factors (organic dysphonia) and other factors (dysfunctional dysphonia)[7]. Spectral "noise" is strictly linked to air flow turbulences in the vocal tract, mainly due to irregular vocal folds vibration and/or closure, causing dysphonia. Such symptom requires a set of endoscopic analysis (by using videolaryngoscope, VLS) for accurate analysis [3]. However, clinical experience has pointed out that dysphonia is often underestimated by patients and, sometimes, even by family doctors. But early detection of dysphonia is of basic importance for pathology recovering. Several methods for assessing speech pathologies have been introduced. In general, objective speech quality measures are usually evaluated in the spectral, time or cepstral domains. In the spectral analysis methods, researchers have tried to keep track of the spectral variations of signal such as amplitude, bandwidth and frequency of formants including sub-band processing methods. In time domain, method based on temporal measurements of signal and their statistics, such as average pitch variation, jitter, shimmer, etc. to distinguish between normal and pathological speech is used. Moreover, Speech processing based on cepstral analysis has proved to be an excellent tool for voice disorder detection. In this paper, we investigate both time domain methods and the adaptation of Automatic Speaker Recognition for dysphonic voice assessment. Mel-frequency cepstral coefficients (MFCC) have traditionally been used in speaker identification applications. In this paper we have used MFCC for the feature extraction from the speech signals provided in the database and dynamic time warping (DTW) is used for feature matching in order to discriminate non-asthmatic persons from the asthmatic's patients. This paper is organized as follows: the speech and subject database and methods for feature extraction and feature analysis are described in section II. The results are presented in Section III and conclusion in Section IV.

Materials and Methods

Speech and subject database:

For this analysis, the speech record database consisted of a sustained phonation of the vowel /a/.The asthmatic group consisted of 21 patients with asthma's disease, 13 males (aged between 26 and 82 years, mean of 51.923 years) and eight females (aged between 46 and 62 years, mean of 53.4 years) and duration of the disease from 1 month to 30 years, with an average of 9.5 years. The control group was composed of 21 individuals non-asthma, four males (aged from 17 to 45 years, mean of 31 years) and twelve females (aged from 20 to 72 years, mean of 45.9 years) and five children (aged from 6 to 10, mean of 8 years).Acoustic assessment was performed by analysis of vowels phonated in isolation and in a constant linguistic test starting with the following words :(" hum sab ek

hein"). Acoustic analysis was performed with PRAAT Software programme. The following parameters were analyzed: F0, F1, F2, F3 Formants frequency levels, degree of voice break in isolated vowels, constant, fundamental frequency, Fo (Hz), Jitter (frequency perturbation – local, %), Shimmer (amplitude perturbation –local, %), Harmonic to noise ratio (HNR – dB), Intensity(dB).

Acoustic Parameters:

Jitter (local): This is the average absolute difference between consecutive periods, divided by the average period. MDVP calls this parameter *Jitt,* and gives 1.040% as a threshold for pathology.

Shimmer (local): This is the average absolute difference between the amplitudes of consecutive periods, divided by the average amplitude. MDVP calls this parameter *Shim*, and gives 3.810% as a threshold for pathology.

Harmonics-to-Noise Ratio (HNR) : A Harmonicity object represents the degree of acoustic periodicity, also called Harmonics-to-Noise Ratio (HNR). Harmonicity is expressed in dB. Harmonicity can be used as a measure for voice quality. For instance, a healthy speaker can produce a sustained a or i with a harmonicity of around 20 dB, and an u at around 40 dB; Hoarse speakers will have an a with a harmonicity much lower than 20 dB.

Degree of Voice Breaks DVB /%/ [15] - the ratio of the total length of areas representing voice breaks to the time of the complete voiced sample; and number of voice breaks NVB. The criteria for voice break area can be a missing impulse for the current period or an extreme irregularity of the pitch period.

Formant Frequency Measures: Frequency component amplified by resonator (vocal tract) and acoustic properties that distinguish speech sounds. Typically measured by LPC (Linear Predictive Coding)or spectrographic analysis. F_1 related to tongue height; F_2 related to tongue advancement. F_2 transition: change in frequency value of formant over time; reflects change in position of articulators.

Mel Frequency Cepstral Coefficient :

The extraction of the best parametric representation of acoustic signals is an important task to produce a better recognition performance. The efficiency of this phase is important for the next phase since it affects its behaviour. MFCC is based on human hearing perceptions which cannot perceive frequencies over 1Khz. In other words, in MFCC is based on known variation of the human ear's critical bandwidth with frequency [8-10]. MFCC has two types of filter which are spaced linearly at low frequency below 1000 Hz and logarithmic spacing above 1000Hz. A subjective pitch is present on Mel Frequency Scale to capture important characteristic of phonetic in speech.

The speech input is recorded at a sampling rate of 16000Hz.This sampling frequency is chosen to minimize the effects of aliasing in the analog-to-digital conversion process.

The MFCC processor consists of seven computational steps is shown in the figure1. Each step has its function and mathematical approaches as discussed briefly in the following:

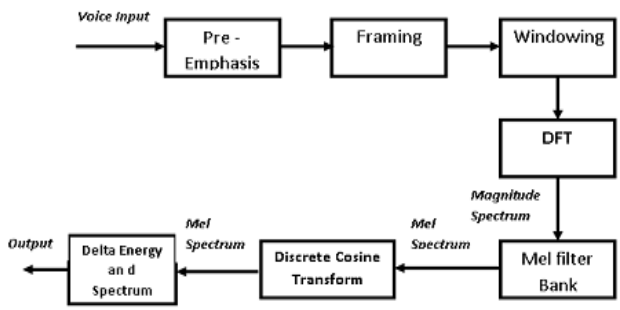

Figure 1: Block diagram of MFCC Processor.

Step 1: Pre–emphasis

This step processes the passing of signal through a filter which emphasizes higher frequencies. This process will increase the energy of signal at higher frequency.[8]

Step 2: Framing

The process of segmenting the speech samples obtained from analog to digital conversion (ADC) into a small frame with the length within the range of 20 to 40 msec. The voice signal is divided into frames of N samples. Adjacent frames are being separated by M (M<N). Typical values used are M = 100 and N= 512.

Step 3: Hamming windowing

Hamming window is used as window shape by considering the next block in feature extraction processing chain and integrates all the closest frequency lines.

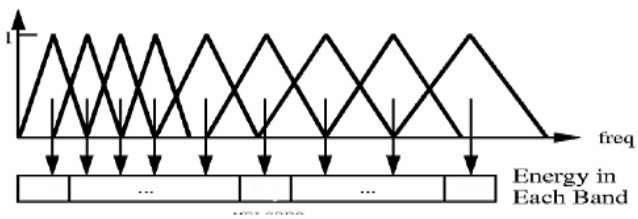

Figure 2: Mel scale filter bank from (young et al 1997).

This figure shows a set of triangular filters that are used to compute a weighted sum of filter spectral components so that the output of process approximates to a Mel scale. Each filter's magnitude frequency response is triangular in shape and equal to unity at the centre frequency and decrease linearly to zero at centre frequency of two adjacent filters [7, 8]. Then, each filter output is the sum of its filtered spectral components.

Step 4: Discrete Cosine Transform

This is the process to convert the log Mel spectrum into time domain using Discrete Cosine Transform (DCT).The result of the conversion is called Mel Frequency Cepstrum Coefficient.

Step 5: Delta Energy and Delta Spectrum

The voice signal and the frames changes, such as the slope of formant at its transitions. Therefore, there is a need to add feature related to the change in cepstral features over time. 13 delta velocity features (12 cepstral features plus energy), and 39 features double delta or acceleration feature are added.

Dynamic time warping

DTW algorithm is based on Dynamic Programming techniques as describes in [11]. This algorithm is for measuring similarity between two time series which may vary in time or speed. This technique also used to find the optimal alignment between two times series if one time series may be "warped" non-linearly by stretching or shrinking it along its time axis. This warping between two time series can then be used to find corresponding regions between the two time series or to determine the similarity between the two time series. Figure 4 shows the example of how one times series is 'warped' to another [12].

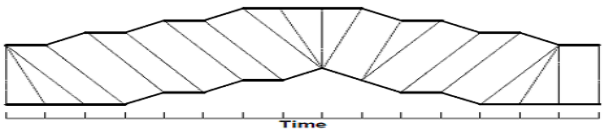

Figure 3: A Warping between two time series[12].

Suppose we have two time series *Q* and *C*, of length *n* and *m* respectively, where:

$$Q = q1, q2, \ldots, qi, \ldots, qn \ldots. \quad (1)$$

$$C = c1, c2, \ldots, cj, \ldots, cm\ldots.. \quad (2)$$

To align two sequences using DTW, an *n*-by-*m* matrix where the (*i*th, jth) element of the matrix contains the distance *d* (*q*i, *cj*) between the two points *q*i and *c*j is constructed. Then, the absolute distance between the values of two sequences is calculated using the Euclidean distance computation:

$$d\,(qi, cj) = (qi - cj)^2\ldots. \quad (3)$$

Each matrix element (*i*, *j*) corresponds to the alignment between the points *q*i and *c*j. Then, accumulated distance is measured by:

$$D(i, j) = \min[D(i-1, j-1), D(i-1, j), D(i, j-1)] + d(i, j)$$

Using dynamic programming techniques, the search for the minimum distance path can be done in polynomial time P(t), using equation below:

$$P(t)=o(N^2V) \quad (4)$$

where, N is the length of the sequence, and V is the number of templates to be considered[8].

Methodology

1. Database of 21 asthmatic patients (13 Male, 8female) undergoing treatment in the military hospital and 21 non-asthmatic person, is collected.
2. Vowels uttered by each person are extracted from the sentence with sampling frequency 16KHz, Mono, 8bit PCM.
3. Programming is done in Matlab to calculate Mel Frequency Cepstral Coefficient (MFCC) as feature extraction and dynamic time warping as feature matching.
4. Acoustic parameters such as fundamental frequency, jitter, shimmer, harmonic to noise ratio, formant frequency, intensity are extracted using PRAAT software.
5. Voice features of asthmatic patient and non-asthmatic persons are compared.

Algorithm for Proposed Architecture

An analysis of acoustic feature of asthmatic patient and an attempt to relate the variation in the voice characteristics of asthmatic. Recognition experiments is done by database of asthmatic patients recorded from the Military Hospital. Block Diagram illustrated in figure-4 describes the speech processing step to diagnose asthmatic patients

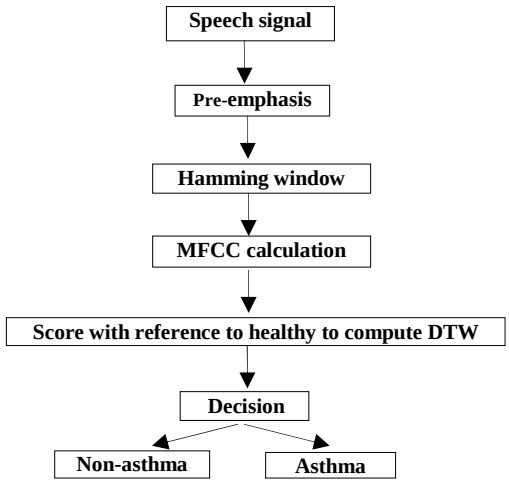

Figure 4: Algorithm for DTW score calculation.

Results

We recorded 21 phonation uttered by the asthmatic patients and 21 phonation by non-asthmatic group. For the acoustic analysis all 21 phonation of vowels is considered and for programming part 16 asthmatic and 16 non asthmatics voices are taken. Table 1 and Table 2 shows the results obtained by acoustic analysis using Praat software. Figure-5 shows the DTW Scores calculated and plotted.

Table 1: Acoustic Parameters of Asthmatic Group

SNo.	Parameters	Asthmatic group				
		min.	max	mean	overall result/success	
1	Pitch(Hz)	97.44	213	161	slighty higher	
2	Std. Devi(Hz)	4.5	53.67	24.8	higher	
3	UVF	15.09	70.11	29	mix value	
4	DVB(%)	5.14	62.75	31.1	higher	
5	Intensity(db)	67.34	90.51	80.4	lower	
6	Shim(db) local	4.912	17.94	8.44	<=3.8	85.71%
7	Jitter(db) local	0.83	4.27	1.73	<=1.04	95.23%
8	H/N ratio(db)	8.26	19.3	14.4	< 20	
9	Formant1	329	846	604	higher	
10	Formant 2	1697	2250	1972	lower	
11	Formant3	2503	3276	2967	mix value	

(CALCULATIONS BY PRAAT)

Table 2: Acoustic Parameter of Non-Asthmatic Group.

SNo.	Parameters	Non-asthmatic group				
		min.	max	mean	Overall result /success	
1	Pitch(Hz)	106	296	197	lower	
2	Std. Devi(Hz)	0.68	10.3	5.54	lower	
3	UVF	11.7	54.2	30.6	mix value	
4	DVB(%)	6.7	55.5	26.4	lower	
5	Intensity(db)	70.8	89.2	86.9	higher	
6	Shim(db)local	1.72	6.76	3.7	>3.8	71.28%
7	Jitter(db)local	0.12	1.53	0.62	>1.04	95.23%
8	H/N ratio(db)	12.5	21.8	17.6	nearly 20	
9	Formant1	363	857	458	lower	
10	Formant 2	1692	2771	2309	higher	
11	Formant3	1992	3369	2966	mix value	

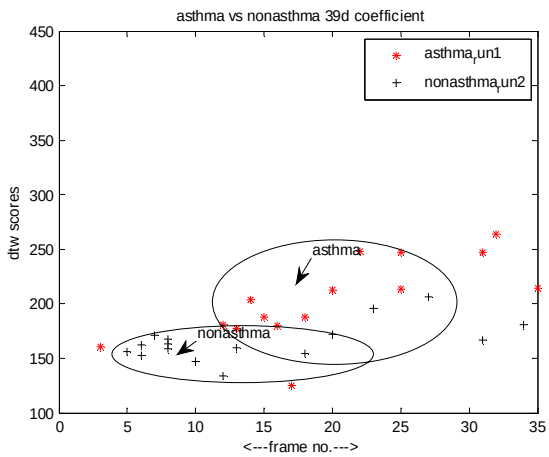

Figure 5: Comparison of asthma and non-asthma group (Vowel 'a' extracted from continuous speech).

Table 3: Rate of Classification From Both Groups

Method	Asthma	Nonasthma	Rate of classification
Acoustics analysis (using praat)	21	21	85%
MFCC/DTW	16	16	62.5%

Conclusion

Acoustic analysis of voice signal is showing better outcome though it is time consuming process. The application of cepstral analysis for the clinical evaluation of voice function has been qualitatively reviewed on asthmatic patients. The mathematical transformations involved in the analysis have been described as well as the suitability of the analysis for this application is described.

References

[1] [My paper]S Hackenberg, T Hacki, R Hagen, N H Kleinsasser, "Voice Disorders in Asthma" Laryngorhinootologie. 2010 Aug460-4. Epub 2010.

[2] Lee L, Chamberlain LG, Loudon RG, Stemple JC, " Speech segment durations produce by healthy and asthmatic subjects", J Speech Hear Disord. 1988 May;53(2):186-93.

[3] L Lee, R G Loudon, B H Jacobson, R Stuebing, " [My paper]Department of Communication Sciences and Disorders, University of Cincinnati, Ohio 45221.

[4] Md. Rashidul Hasan, Mustafa Jamil, Md. Golam Rabbani Md. Saifur Rahman, "Speaker Identification Using Mel Frequency Cepstral Coefficients", 3rd Proceedings of International Conference on Electrical & Computer Engineering, ICECE 2004, 28-30 December 2004, Dhaka, Bangladesh, pp 565- 568

[5] S. B. Davis, "Acoustic characteristics of normal and pathological voices, " Speech and Language: Advances in Basic Research and Practice", vol. 1, pp. 271–335, 1979.

[6] P.Yu, M.Ouaknine, J.Revis, and A.Giovanni, "Objective Voice Analysis for Dysphonic Patients: A Multiparametric Protocol Including Acoustic and Aerodynamic Measurements", Journal of Voice Vol. 15, No. 4, pp.529–542, 2001

[7] B. Boyanov, S.Hadjitodorov: "Acoustic analysis of pathological voices: a voice analysis system for screening of laryngeal diseases", Proc. IEEE Engineering in Medical and Biology, (1997), vol. 16, no. 4, 74-82.

[8] Lindasalwa muda, Mumtaj Begam, I. Elamvazuthi, "Voice Recognition Algorithms using MFCC & DTW Techniques", Journal of Computing, volume2, issue3, march 2010.

[9] Jamel Price, Ali Eydgahi, "Design of Matlab based Automatic Speaker Recognition Systems", 9th International Conference on Engineering Education, July2006.

[10] Rosalyn J. Moran*, Richard B. Reilly, "Telephony-Based Voice Pathology Assessment Using Automated

Speech Analysis'. ieee transactions on biomedical engineering, vol. 53, no. 3, march 2006.
[11] By Jackie Ehrlich, "Asthma, Allergies, Articulation: a Speech Therapy Perspective"
[12] Dirk Michaelis, "Selection and combination of acoustic features for the description of pathologic voices", 1998,
[13] Tripti kapoor, Dr. R.K.Sharma, " Parkinson's disease Diagnosis using Mel Frequency Cepstral Coefficients and Vector Quantization", International Journal of Computer Applications (0975 – 8887) Volume 14– No.3, January 2011
[14] Batler, "Inhaled Beclomethasone Dipropionate Improves Acoustics Measure of Voice in Asthmatic Patient", Journal of Speech and Hearing Research, Volume 39, 126-134, February 1996
[15] Lingyun Gu, John G. Harris, " Disordered Speech Assessment Using Automatic Methods Based on Quantitative Measures", Journal on Applied Signal Processing, 2005
[16] "Subjective and objective evaluation of voice quality in patients with asthma", Journal of Voice, 2007
[17] L. Rabiner, and B.H. Juang, "Fundamentals of Speech Recognition". Prentice Hall, 1993.
[18] Risa J Robinson, Richard L Doolittle, John N Diflorio, "Use of asthmatic pulmonary function test data to predict lung deposition", [My paper]J Aerosol Med. 2007 ;20 (2):141-62
[19] Kosztyła-Hojna Bb, " Objective analysis of voice quality in asthma patients", Int. Rev. Allergol. Clin. Immunol., 2009; Vol. 15, No. 1-2
[20] Praat software www.praat.org.
[21] Han Su Kim, Jin Wook Moon, " A Short-Term Investigation of Dysphonia in Asthmatic Patients Using Inhaled Budesonide, '' Journal of Voice, Vol. 25, No. 1, pp. 88-93.
[22] http://www.mathsworks.com