Classification of Pathological Voice Signal with Severe Noise Component

Tao Li* · Cheolwoo Jo*

ABSTRACT

In this paper we tried to classify the pathological voice signal with severe noise component based on two different parameters, the spectral slope and the ratio of energies in the harmonic and noise components (HNR). The spectral slope is obtained by using a curve fitting method and the HNR is computed in cepstrum quefrency domain. Speech data from normal peoples and patients are collected, diagnosed and divided into three different classes (normal, relatively less noisy and severely noisy data). The mean values and the standard deviations of the spectral slope and the HNR are computed and compared with in the three kinds of data to characterize and classify the severely noisy pathological voice signals from others.

Keywords: classification, noise, spectral slope, HNR, cepstrum

I. INTRODUCTION

These days there are many attempts to analyze and classify the pathological and normal voice by the original parameters (Jitter, Shimmer, NHR, SPI, etc.). The major purpose of such researches is to get some good standards and methods to classify and diagnose the patients who have diseases on their vocal folds. [1][2][3][4][5][6]

Although most of the parameters require pitch information, some severely noisy voice from patient makes it difficult to compute pitch values because of noise. Recent researches on discriminating patient's voice are performed using relatively clear voice, of which parameters can be computed without error. Also some of the commercial softwares cannot analyze the characteristics of voice. So those data have been excludedfrom the analysis. But to classify the patient's voice correctly without intervention of human experts, these kinds of voice have to be classified and analyzed as well as clear voice.

In this paper, we tried to classify severely noisy voice from the input based on the fact that their noise components are dominant compared with normal or relatively less noisy voice.

Two different parameters are used to analyze voice and the validities are examined.

* SASPL, School of Mechatronics, Changwon National University, Korea

II. DATA COLLECTION

To collect original voice data, the collection system is installed in a room of the ENT department of hospital.

The recording process is performed semi-automatically with the intervention of operator to control the quality and procedure. Also the voice materials from the different male speakers are collected using DAT (Digital Audio Tape). [7] The sampling rate is 50 KHz and the resolution is 16 bits. The collection is

conducted in a soundproof room of hospital. All the subjects are asked to pronounce /a/. Total voice data include 10 normal cases, 10 relatively less noisy cases and 3 severely noisy cases after removing invalid data from the raw data sets. The vocal diseases in the relatively less noisy cases include Vocal Polyp, Vocal Cord Palsy, Vocal Nodule, Vocal Cyst, Vocal Edema, Laryngitis and Glottic Cancers. Those in the severely noisy cases are all Glottic Cancers.

III. ANALYSIS METHOD

In this section, we discuss the two different analysis methods used in our study to classify and characterize the pathological voice signals with severe noise component from others. They are: spectral slope and harmonic-to-noise ratio (HNR). [8]Also the mean values and the standard deviations of the two different parameters are computed.

A. *Spectral Slope*

The method to get the spectral slope is a kind of curve fitting actually. In this analysis, we first use linear predictive coefficients (LPC) method (52nd-order) to analyze the segment (2048 points) of the original voice signal directly. After that, we can get the vocal tract function and the log magnitude spectrum of the all-pole LP filter. Finally the polynomial coefficients of a 1st-order linear function are found by using the curve fitting method. The linear function fits the discrete spectral data obtained in the log magnitude spectrum above. Then it is the slope of the beeline that is the spectral slope. Through the different spectral slopes of three kinds of voice signals, we can study the voice source characteristics, especially the severely noisy pathological voice.

B. *Harmonic-to-Noise Ratio*

HNR is a direct measure of the noise component, defined (in decibels) by

$$HNR = 10 \times \log_{10}\left(\frac{E_p}{E_{ap}}\right) \tag{1}$$

where E_p and E_p are the energies in the harmonic and the noise components respectively. The total energy is defined as the sum of the squared amplitudes for all samples. [8]

In order to get the HNR, the harmonic and the noise components are identified using the cepstrum of the signal in the analysis segment (2048 points). [9][10] First the data in the segment are multiplied with the Hamming window and a DFT is computed. Then we consider the log magnitude spectrum of the DFT. And the IDFT of this log magnitude spectrum gives the real cepstrum in the quefrency domain.

The quefrency domain of the cepstrum can be splitapproximately into three distinct regions. The low quefrency region in the range 0 - 0.0015 s (75 points) is mainly due to the vocal tract system characteristics. The rest of the quefrency region can be attributed to the excitation part. And in the region corresponding to the excitation, the range (15 points) whose center is the highest cepstral peak at the pitch period can be attributed to the harmonic part. The range is easily found with eyes in the cepstrum of the normal and partial relatively less noisy voices. For example, the range in the normal voice cepstrum is around 0.01 s almost all. But it is very difficult to identify the ranges in the rest relatively less noisy voices and especially severely noisy voices because the noise components are increased remarkably. Therefore we use computer program to identify the highest cepstral peak and get the corresponding range. Obviously the remaining excitation portion of the quefrency region can be attributed to the noise part. The choices for the widths for the three parts in the quefrency domain are only approximate. Then we compute the HNR in the quefrency domain according to the equation (1).

C. Mean Value and Standard deviation

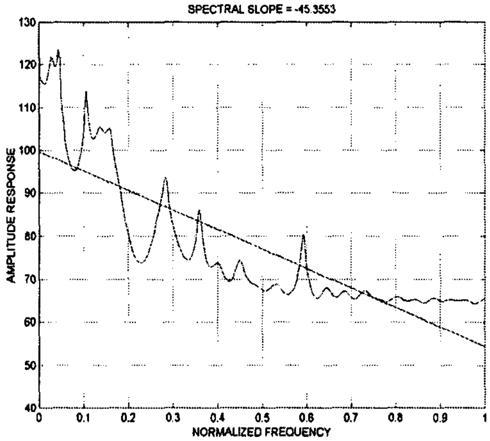
After obtaining the spectral slope and the HNR of all voice signals, we compute the mean values and the standard deviations of the two different parameters respectively. Then the computation results are compared with in the three kinds of data to classify and characterize the severely noisy pathological voice signals from others.

IV. EXPERIMENT AND RESULTS

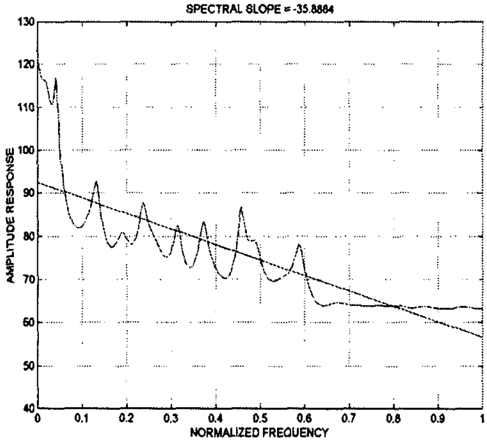
In order to accurately compare, characterize and classify the severely noisy pathological voice signals from others, the log magnitude spectrum and the fitting beeline are plotted and the spectral slope is computed for each of the three classes of voice signals respectively in the experiments. Also the same operations are done for the cepstrum and the HNR. In the plot and computation results using the spectral slope analysis method, we find there is the similar characteristic for any of the same class voice signals. So is there using the HNR analysis method. Thereby we choose and analyze the voices whose spectral slope and HNR characteristics represent the difference most among the three classes of voice signals. Here a normal voice, two relatively less noisy voices and a severely noisy voice are analyzed. Because the relatively less noisy voice is the transitional signal between the normal voice and the severely noisy voice, we choose the two relatively less noisy voices to observe the detailed change and difference of the characteristic in the two kinds of analysis methods. The same voice signal is analyzed by the spectral slope method and the HNR method respectively.

Fig.1 shows the different log magnitude spectra, fitting beelines and spectral slopes of the normal voice, the two relatively less noisy voices and the severely noisy voice. The corresponding fitting beeline and spectral slope are in the spectra and on the top of spectra respectively.

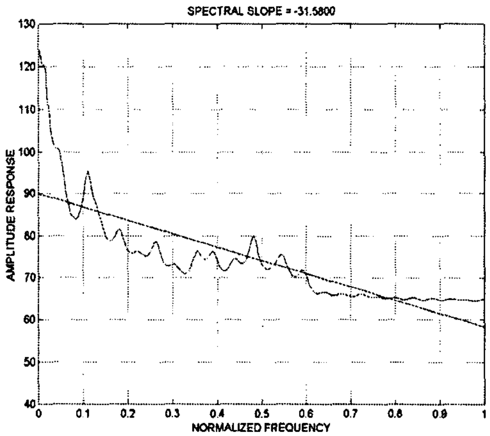
Fig.2 shows the comparisons of the mean values and the standard deviations of the spectral slopes from all of the normal, the relatively less noisy and the severely noisy voices (23 signals). The graph shows the relative change of the spectral slopes among the three classes of voice signals.



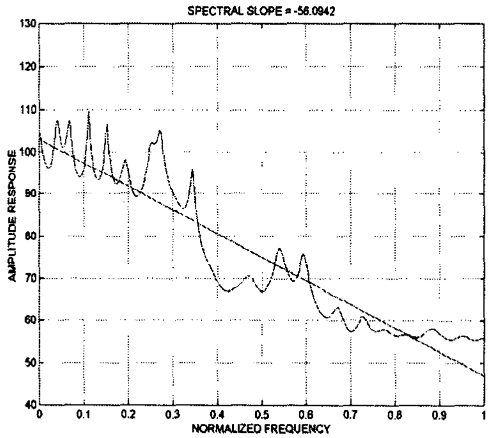
(a) Plot of the normal voice



(b) Plot of the relatively less noisy voice 1



(c) Plot of the relatively less noisy voice 2



(d) Plot of the severely noisy voice

Fig.1. Comparisons of the spectral slope analysis

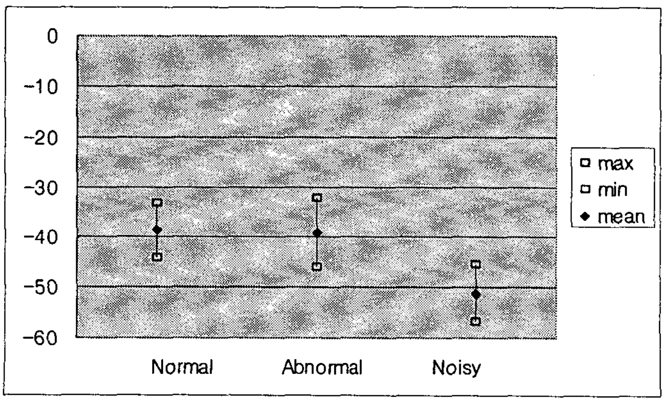
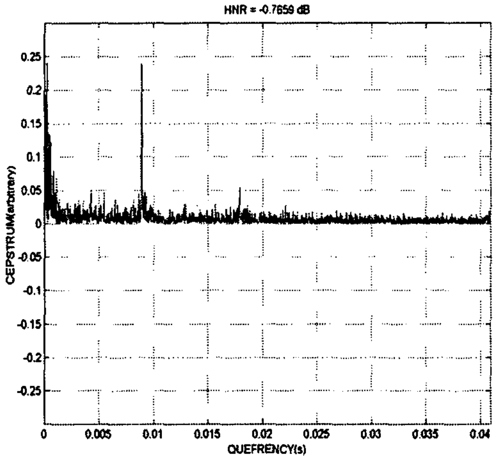


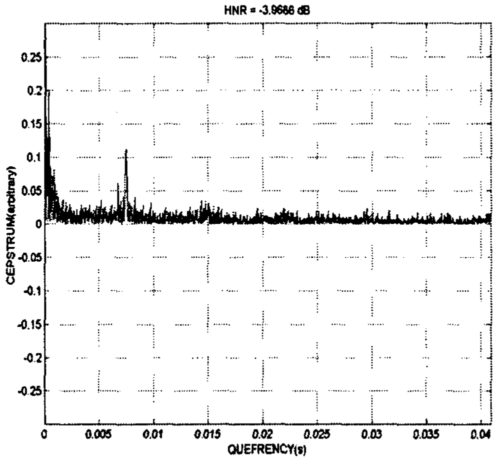
Fig.2. Comparisons of the mean values and the standard deviations of the spectral slope

Fig.3 shows the different cepstrum and HNR of the normal voice, the two relatively less noisy voices and the severely noisy voice. The corresponding HNR is on the top of cepstra.

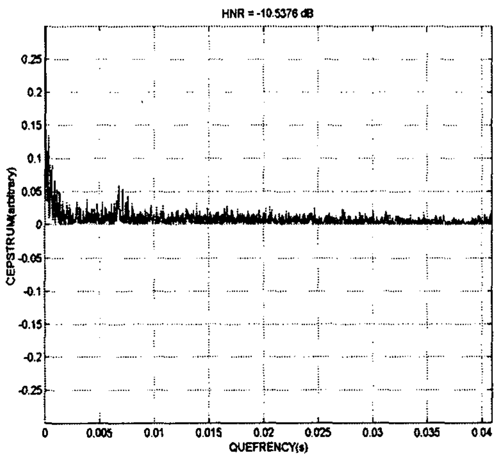
Fig.4 shows the comparisons of the mean values and the standard deviations of the HNR's from all of the normal, the relatively less noisy and the severely noisy voices (23 signals). The graph shows the relative change of the HNR's among the three classes of voice signals.



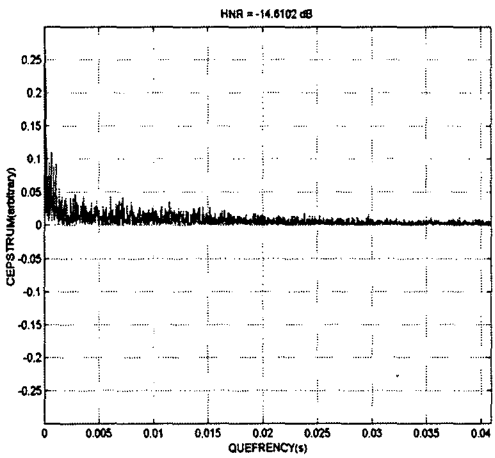
(a) Plot of the normal voice



(b) Plot of the relatively less noisy voice 1



(c) Plot of the relatively less noisy voice 2



(d) Plot of the severely noisy voice

Fig.3. Comparisons of the HNR analysis

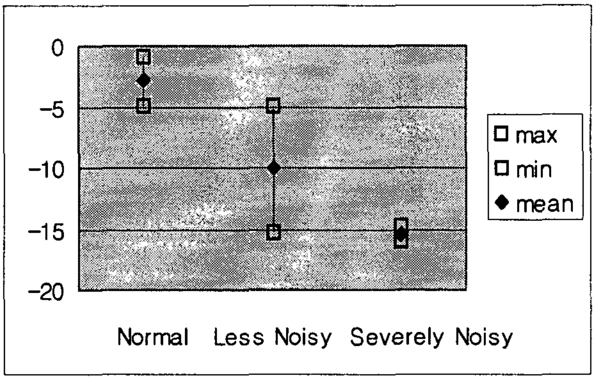


Fig.4. Comparisons of the mean values and the standard deviations of the HNR

V. Discussion

From the experimental results using the spectral slope analysis method, we can observe the significant different distribution characteristic of the spectral slopes between the severely noisy voice signals and other two kinds of remaining voice signals as shown in figure 2. The spectral slope of the severely noisy voice is smaller than that of the two kinds of remaining voices. The mean of the spectral slope of the severely noisy voice is around -50, but that of the two kinds of remaining voices around -40. Especially, the spectral slope of the severely noisy voice is -56.0942, but that of the normal voice -45.3553 and those of the two relatively less noisy voices -35.8884 and -31.5800 as shown in figure 1. So the results look good.

Also from the experimental results using the HNR analysis method, we can observe the significant different shape of the cepstrum and the significant different distribution characteristic of the HNR among the three classes of voice signals as shown in figure 3 and figure 4. The highest cepstral peak range (15 points) is very clear and outstanding in the cepstrum of the normal voice, and those of the two relatively less noisyvoices are changed to be non-outstanding gradually but clear still, but that of the severely noisy voice is not clear and outstanding especially as shown in figure 3. And the HNR of the severely noisy voice is much smaller than that of the normal voice and the relatively less noisyvoice as shown in figure 4. The mean of the HNR of the severely noisy voice is around -15 dB, but that of the normal voice around -3 dB and the relatively less noisy voice around -10 dB. With the noise component increased, the HNR of the voice is smaller and smaller. Especially, the HNR of the severely noisy voice is -14.8102 dB, but that of a normal voice -0.7659 dB and those of two relatively less noisy voices -3.9686 dB and -10.5376 dB as shown in figure 3. So the results look good, too.

VI. CONCLUSION

In this paper we collected the voice materials using DAT and tried to classify the pathological voice signal with severe noise component based on two kinds of parameters, the spectral slope and the HNR.

From the experiments we can obtain the good classification achievements using the two analysis methods. And we can conclude that the spectral slope and the HNR can be both considered as the good and useful parameters for the classification of the pathological voice signal with severe noise component from others.

But the total amount of voice data is still not enough (especially the severely noisy data) to generalize the performance and characteristic, more data collection is required. Also there have to be more researches in looking for other better methods and parameters to classify the severely noisy voice signal and improve the classification veracity.

ACKNOWLEDGEMENT

This study was supported by a grant of the Korea Health 21 R&D Project, Ministry of Health & Welfare, Republic of Korea. (No. 02-PJ1-PG10-31401-0005)

REFERENCES

[1] Godino-Llorente. Santiago Aguilera-Navarro*. Carlos Hernandez-Espinosa**. Merced-es Fernandez-Redondo**. Pedro Gomez-vilda+. 1999. "On The Selection of Meanin gful Speech Parameters Used by a Pathologic/Non pathologic Voice Register Classifier". *Eurospeech '99*, Volume 1, Page 563-566.

[2] Jo, Cheolwoo., Kim, Daehyun, Wang, Soogeon. 1999. "Classification of Pathological Voice into Normal/Benign/Malignant State." *Eurospeech'99*, Budapest, Hungary, pp.399-402.

[3] Kim, Daehyun., Jo, Cheolwoo. 1998. "A study on the classification of pathological voice." *15th Workshop on Speech Communication and Signal Processing*, pp.388-391.

[4] Kim, Kwangin., Jo, Cheolwoo., Kim, Daehyun., Wang, Soogeon., Jeon, Gyerock., Ahn, Sihun., Kim, Kiryon., Kim, Yongju. 2000. "Collection, Analysis and Classification of Pathological Voice using ARS and Neuralnetwork." *13rd Korean Conference on Signal Processing*, pp.955-958.

[5] Jo, Cheolwoo., Kim, Daehyun., Kim, Kwangin., Wang, Soogeon., Jeon, Gyerock. Ahn, Sihun., Kim, Kiryon., Kim Yongju. 2000. "Implementation of Analysis Tools for Pathological Voice." *17th Workshop on Speech Communication and Signal Processing*, pp.211-214.

[6] Jo, Cheolwoo., Kim, Kwangin., Kim, Daehyun., Wang, Soogeon. 2001. "Screening of

Pathological Voice from ARS using Neural Networks." *international workshop on Maveba*. Firenze. Italy.
[7] Jo, Cheolwoo., Kim, Kwangin., Kim, Daehyun., Wang, Soogeon., Jeon, Gyerok. 2001. "Comparisons of Acoustical Characteristics between ARS and DAT voice" *submitted to Eurospeech 2001*. Denmark.
[8] C. d'Alessandro. V. Darsinos. B. Yegnanarayana. 1998. "Effectiveness of a periodic and aperiodic decomposition method for analysis of voice sources." *IEEETransactions on Speech and Audio processing*, Vol. 6, no 1, pp. 1223.
[9] Yegnanarayana B. D'Alessandro C. Darsinos V. 1998. "An iterative algorithm for decomposition of speech signals into periodic and aperiodic components." *IEEE Transactions on Speech and Audio processing*, Vol. 6, no 1.
[10] Krom. G. 1993. "A cepstrum-based technique for determining a harmonicsto -noise ratio in speech signals." *J. Speech Hearing Res.*, vol. 36, pp.254266.

Received: November 14, 2003.
Accepted: December 15, 2003.

▲ Lee Tao
Changwon National University, Sarim 9-dong, Changwon, Kyungnam, Korea
SASPL, School of Mechatronics, Changwon National University
Tel: +82-55-279-7559 Fax: +82-55-262-5064
E-mail: litao_com@hotmail.com

▲ Jo Cheol-woo
Changwon National University, Sarim 9-dong, Changwon, Kyungnam, Korea
SASPL, School of Mechatronics, Changwon National University
Tel: +82-55-279-7552 Fax: +82-55-262-5064
E-mail: cwjo@sarim.changwon.ac.kr