ANALYSIS OF DISORDERED SPEECH SIGNAL USING WAVELET TRANSFORM

Cheol-Woo Jo, Dae-Hyun Kim

Department of Control and Instrumentation Eng., Changwon National University

Changwon, Kyeongnam 641-773 Korea

Phone: +82-551-279-7552, Fax: +82-551-262-5064, Email: cwjo@sarim.changwon.ac.kr

ABSTRACT

In this paper a method to analyze pathological speech signal using wavelet transform is suggested. Pathological speech signal is commercially available pathological voice database and analyzed by the suggested method. Normal speech signal is also from the same database and analyzed as well. Then the results are compared to find the differences between normal and pathological speech. Three level wavelet transform is used. Normalized energy ratios between the levels and normalized peak-to-peak values are used as parameters. As a result, it was possible to distinguish between normal and pathological speech signal.

1. INTRODUCTION

These days the importance of human health is increasing more and more. And there is needs for distinguishing between normal and pathological speech signals. Generally medical doctors are using special apparatus, similar to endoscope, to diagnose the mal-functioning vocal cords. Using such apparatus, doctors can see vocal cords. But there are needs to diagnose vocal cords by only speech signal. And there have been many researches about such applications.[1][2] In this paper we suggest a method using wavelet transform which is a kind of orthogonal transform.

2. CHARACTERISTICS OF PATHOLOGICAL SPEECH

There are many diseases which can be occurred on vocal cord. Some kind of diseases do not affect the quality of speech signal. But also there are some which do affect. Human listeners, doctors or even plain peoples, can distinguish the differences between normal and abnormal speech. Generally some conceptual measures are used for such diagnosis, for example husky, creaky etc. As a numerical parameters, which can measure the variations of vocal sources, jitter and shimmer are most widely used ones. But these two parameters require pitch information to be computed. And extracting pitch is sometimes difficult. Especially for pathological voices, pitch on voice is often irregular. So this affects the jitter and shimmer causing wrong decision. Generally pathological speech has more higher frequency noise and more irregular pitch.

In this paper, to simplify the experiment we limited the kinds of deseases to two, i.e. vocal fold edema and polypoid degeneration (Reinke's), which is thought to be related with cancer or symptoms are close to cancer. That is because our main aim is am early finding of cancer and prevent them.

3. WAVELET TRANSFORM

Wavelet transform if a kind of time-frequency orthogonal transform. This transform has much flexibility in its time and frequency resolution. In case of DWT(Dyadic Wavelet Transform), it shows narrower time resolution at lower frequencies while it shows wider time resolution at higher frequencies. And an arbitrary division of the time-frequency space is possible. Expression (1) is the definition of WT.

$$WT(a,b) = \int f(t)\Psi_{a,b}(t)dt \tag{1}$$

$$\Psi_{a,b}(t) = \frac{1}{\sqrt{a}} \cdot h\left(\frac{t-b}{a}\right)$$

For our experiment we use dyadic wavelet transform. In dyadic wavelet transform frequency band is divided into the number of levels by the factor of two. Figure1 shows the structure of wavelet transform we used.

Wavelet transform is used because of its multi-resolution capability. It is actually a multi-bandpass filter which has different resolution according to frequency band.

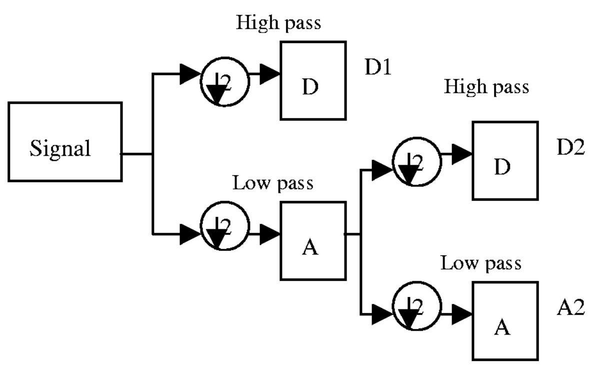

Figure 1: Dyadic Wavelet Transform.

4. PROPOSAL FOR NEW PARAMETER

There are already many parameters which can specify the characteristics of the vocal cords. Jitter and shimmer is the most widely used one. And there are more parameters , which is currently used to diagnose pathological speech. But almost all of those parameters require exact pitch information as a preprocessing. The problem is that some of the pathological voice does not show definite pitch. Sometimes it is very hard to detect pitch even by the speech experts. Accordingly much of the errors in such parameters are caused by the wrong pitch detection. In such cases detecting pitch is almost impossible if human expert cannot recognize the pitch. So to reduce such problem some of the parameters such as NHR is computed without computing pitch information. One of the reason that we use wavelet transform is computing a parameter which does not require pitch information. Another reason is to use the characteristics of the wavelet transform, i.e. multi-resolution characteristics.

To find an effective parameter which can be used to distinguish between normal and pathological speech, at first we obtained output signals from the three scales of wavelet transform. The observed periodicity in lower band scale from wavelet transform is more dominant than in higher-band scales in general. Also it is observed that the periodicity of speech is consistent in normal speech meanwhile it is decreasing rapidly in pathological speech. We choose two parameters from the result of wavelet transform to measure the consistency of periodicity across the scales.

The first parameter is defined as follows and it represents the consistency of normalized energy across the scales.

$$E_N(i) = \frac{E_i}{E_T} (i = 1,2,3) \quad (2)$$

This parameter shows how well the signal's energy is spread across the scale. In general normal speech shows more consistency than in pathological speech. The second parameter is the consistency of energy-normalized peak-to-peak amplitude across the scales.

$$EV_{pp}(i) = \frac{V_{pp}(i)}{E(i)} (i = 1,2,3) \quad (3)$$

This parameter is the ratio of maximum peak-to-peak amplitude versus scale energy. It shows how much bigger the periodic components is in the scale. From the two parameters energy consistency slope(ECS) and peak consistency slope(PCS) is measured by obtaining linear regression coefficients as a final parameter. The suggested parameters does not require the tricky pitch extractions for measurement.

Figure2 shows the clock diagram of the signal processing.

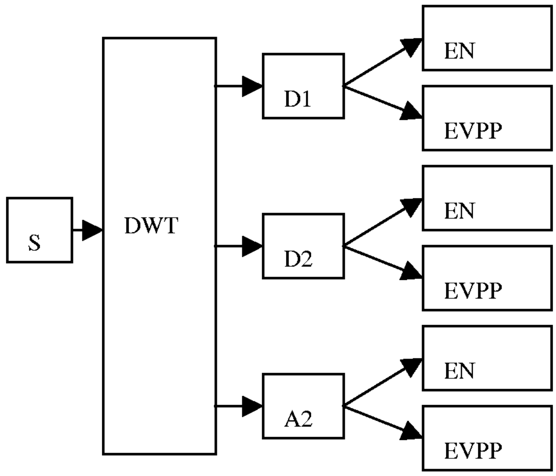

Figure 2: Signal Processing Block Diagram

5. EXPERIMENT AND DISCUSSIONS

To test the validity of the parameters the disordered voice database from KAY elemetrics Ltd is used.[3] We computed ECS, PCS and jitter, shimmer and NHR(noise to harmonic ratio) are used as a reference parameter. Jitter, shimmer and NHR is pre-computed by MDVP(Multi-dimensional voice program: voice analysis program) and included in the database.

Figure3 shows the variations of ECS across the scales. Figure4 shows the variations of PCS across the scales.

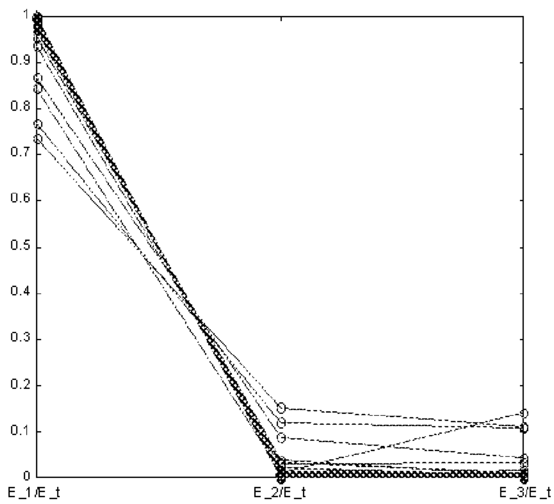

Figure 3: Variations of ECS parameter (+: normal voice, o: pathological speech)

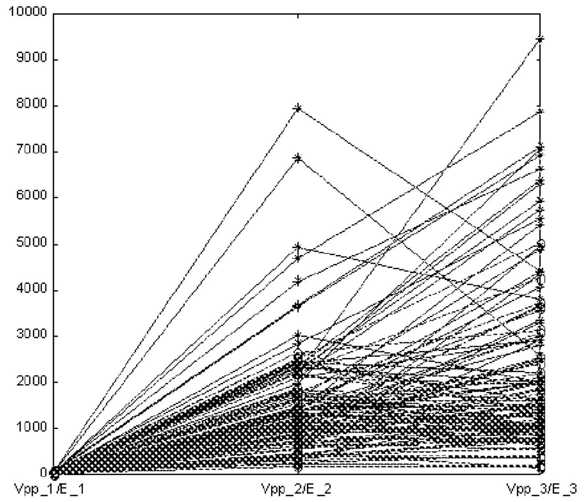

Figure 4: Variations of PCS parameter (+: normal voice, o: pathological speech)

Figure5 shows the ECS and the PCS versus each conventional parameters. From the figure3 it is shown that generally normal voice have more energy than the pathological voice. From the figure3 normalized peak-to-peak amplitude is bigger in higher frequency scales for normal voice meanwhile it is smaller in higher frequency scales for pathological voice.

Table1 shows mean and standard deviations of proposed parameters and conventional parameters.

Table 1: Mean and Standard Deviations of Parameters

Parameters		Mean	S.D.
Jitter	Normal	0.6157	0.4377
	Patient	1.9645	2.4118
Shimmer	Normal	2.2055	0.9242
	Patient	6.5540	4.1706
NHR	Normal	0.1158	0.0124
	Patient	0.1945	0.1241
ECS	Normal	-0.4985	0.0010
	Patient	-0.4850	0.0356
PCS	Normal	1668.4	1089.4
	Patient	674.5	470.0

From table1 we can notice that in PCS and ECS, the differences of mean and standard deviation between normal and pathological speech is similar to the conventional parameters. And it is verified that proposed parameters can be used to diagnose pathological speech without explicit pitch extraction.

6. CONCLUSION

In this paper we proposed two parameters which can discriminate pathological voices from unknown speech. Wavelet transform is used as a basic parameter extractor. And ECS and PCS parameters are suggested. From the analyzed result and comparisons with conventional parameters, it is verified that suggested parameters can be effectively used to diagnose the pathological speech. So wavelet transform can be applied to analysis the pathological speech signal. There is no need of pitch extraction with this parameter.

But still a single parameter is not sufficient for diagnosis. In further researches it is required developing a method to combine current parameters. It is verified from our experiment that combination of two parameters can better distinguish between normal and pathological speech. Next step following this research is the decision of threshold values for each parameters.

Ackknowledgement

This paper is a partial result from the research funded by Korea Research Foundation. The authors wish to acknowledge the financial support of the Korea Research Foundation made in the program year of 1997.

7. REFERENCES

1. F.Plante, H.Kessler, B.Cheetham, J.Earis, “Speech Monitoring of Infective Laryngitis”, Proceedings of ICSLP’96, pp.749-752, Philadelpia, 1996
2. M.N.Vieira, F.R.McInnes, M.A.Jack, “Robust F0 and Jitter Estimation in Pathological Voices”, Proceedings of ICSLP’96, pp.745-748
3. “Disordered Voice Database”, Version1.03, Kay Elemetrics Corp, 1994

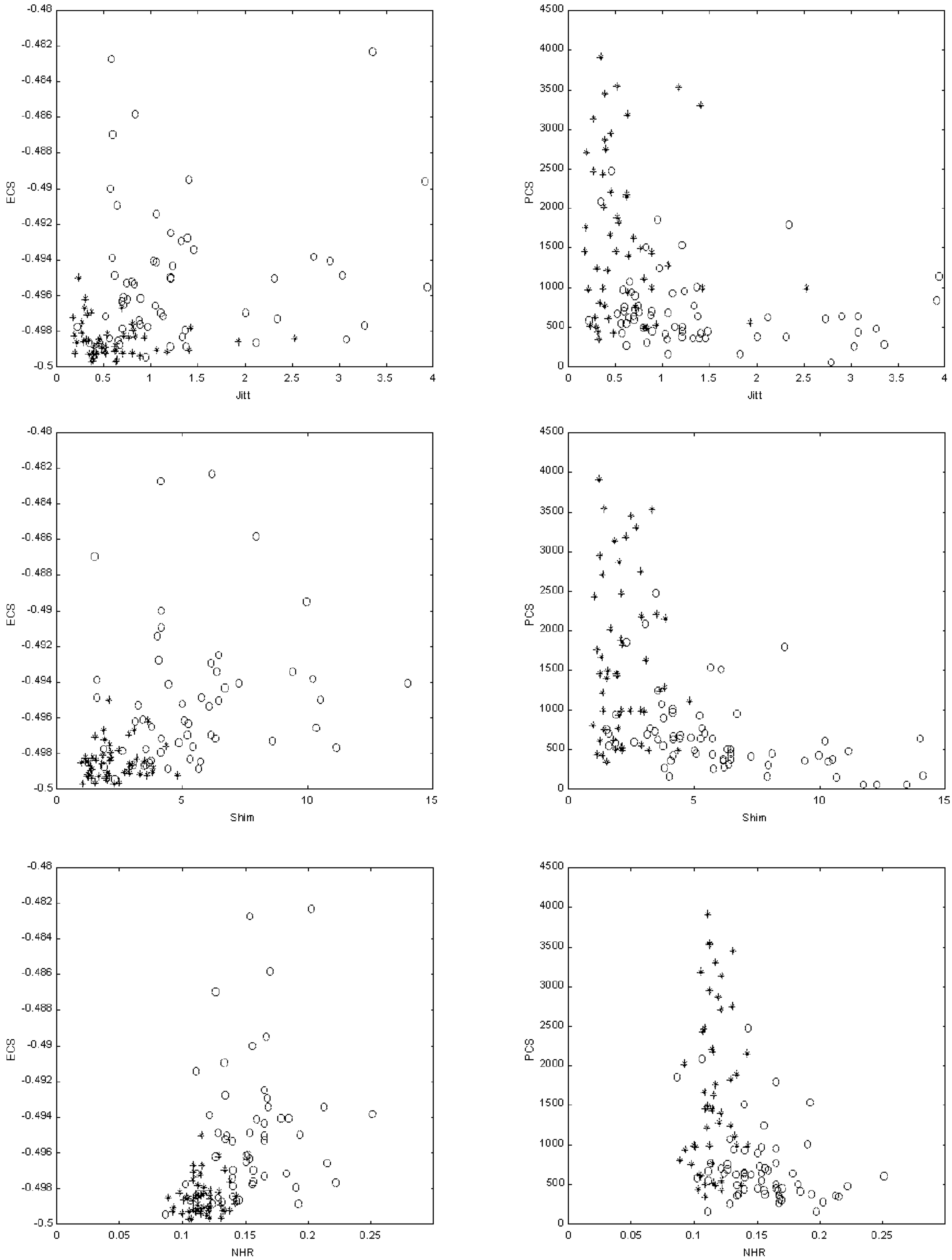

Figure 5: ECS, PCS vs Jitter, Shimmer, NHR (+:Normal Speech, o: Pathological Speech)