Diagnosis of Pathological Speech Signals Using Wavelet Transform*

Cheol-Woo Jo · Dae-Hyun Kim**

ABSTRACT

In this paper a method to diagnose pathological voices using wavelet transform is sug gested. Pathological voices are collected from hospital and analyzed by the suggested method. Normal voices are collected separately and analyzed. Then the results are compared to find the differences in their characteristics. 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 voices.

Keywords : wavelet, diagnosis, pathological speech

1. Introduction

These days the importance of human health is increasing more and more. And there is need for distinguishing between normal and pathological speech signals. Generally medical doctors are using special apparatus to diagnose the mal-functioning vocal cords. Using such apparatus, doctors can see vocal cords and disgnose the patients. But there is also the need to diagnose vocal cords by only speech signals. And there have been many trials for such applications. In this paper we suggest a method using wavelet transform which is a kind of orthogonal transform.

* This paper is a partial result from the research funded by Korea Research Foundation.
** Dept. of Control and Instrumentation Eng., Changwon National Univ.

2. Diagnosis of Pathological Voices

There are many diseases which can occur on the vocal cord. Some kinds of diseases do not affect the quality of speech signals, but there are also some which do. Human listeners, doctors or even plain people, can distinguish the differences between normal and abnormal speech. Generally some conceptual measures are used for such diagnosis, for example husky, creaky etc. As numerical parameters, which can measure the variations of vocal sources, jitter and shimmer are widely used. But these two parameters that require pitch information to be computed. Extracting pitch is sometimes difficult, especially for pathological voices, pitch on voice is often irregular. So this affects the jitter and shimmer values causing wrong decision. [1][2]

In this paper we confined to the limited kind of voice diseases only to simplify the experiment.

3. Wavelet Transform

Wavelet transform is a kind of time-frequency orthogonal transform. This has much flexibility in its time and frequency resolution. In case of DWT (Dyadic Wavelet Transform), it shows a 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 (Wavelet Transform).

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

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

Here WT is wavelet coefficient, $\Psi_{a,b}$ is a wavelet which is dilated by a, translated by b and h(t) is a mother wavelet. ere WT is wavelet coefficient.

Figure 1 shows the structure of dyadic wavelet transform used here.

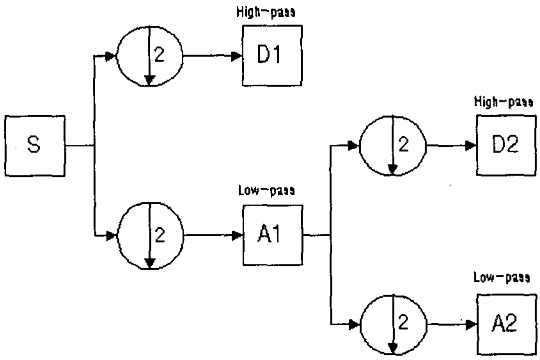


Figure 1. Dyadic Wavelet Transform

Input signal S is down sampled by two and divided into higher band and lower band. The resulting lower band output is again down sampled and filtered.

4. Signal Processing Method

To find an effective parameter which can be used to distinguish between normal voice and pathological voice, at first we obtained output signals from the three scales of wavelet transform. The observed periodicity in the 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 decreases rapidly in pathological speech. We chose two parameters from the result of wavelet transform to measure the consistency of periodicity across the scales.

The first parameter is the consistency of normalized energy across the scales.

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

This parameter shows how well the signal's energy is spread across the scales. 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.

$$EV_{pp}(i) = \frac{V_{pp}(i)}{E_N(i)}, \qquad (i=1,2,3) \tag{3}$$

This parameter is the ratio of maximum peak-to-peak amplitude versus scale energy. It shows how much bigger the periodic components are in the scale. From the two parameters energy consistency slope (ECS) and peak consistency slope (PCS) are measured by obtaining linear regression coefficients as a final parameter. ECS parameter is defined by the slope of a line which connects EN (1), EN (2) and EN (3). PCS is defined by the slope of a line which connects EVpp (1), EVpp (2) and EVpp (3). The suggested parameters do not require the tricky pitch extraction procedures.

Figure 2 shows the block diagram of the signal processing.

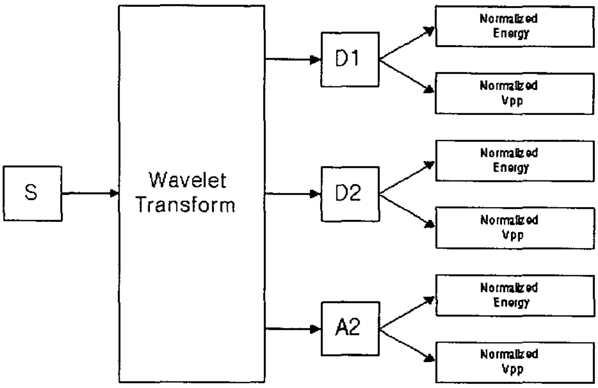


Figure 2. Signal Processing Block Diagram

5. Experiments, Results and Discussions

To test the parameters the disordered voice database from Kay elemetrics is used. In the database, speech samples are sampled by 16bit resolution and by 25 KHz sampling rate. All the voice is sustained vowel /a/s. Jitter, shimmer and NHR (noise to harmonic ratio) is precomputed and stored in the database. We computed PCS and ECS for 53 normal voices and 68 pathological voices.

Figure 3 shows the variations of ECS across the scales. Figure 4 shows the variations of PCS across the scales.

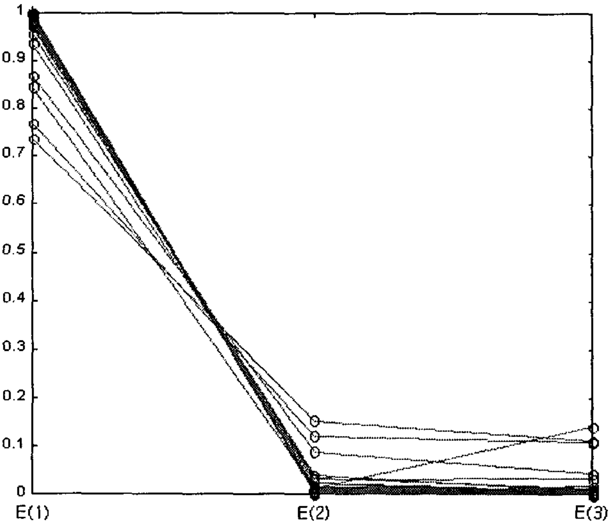


Figure 3. Variations of ECS
(* : normal voice, others : pathological voice)

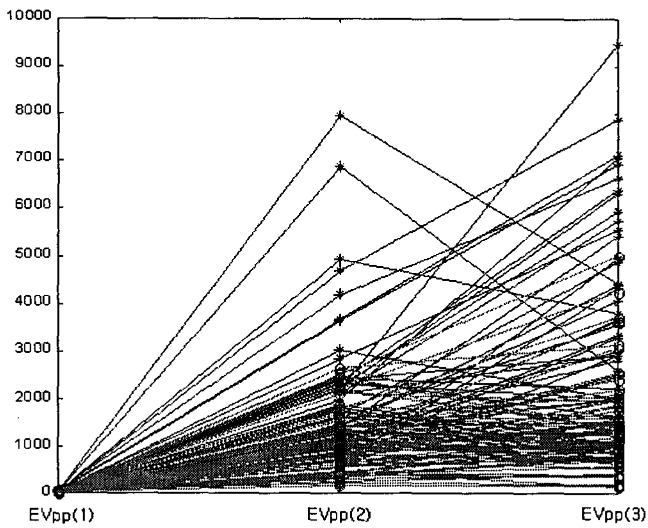


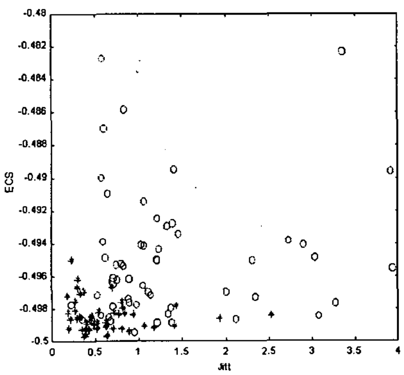
Figure 4. Variations of PCS
(* : normal voice, others : pathological voice)

From the figure 3 it is shown that generally normal voices have more energy than pathological voices. From the figure 4 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.

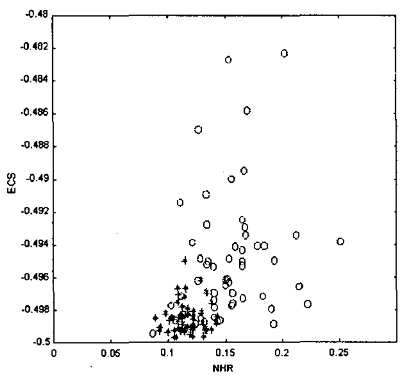
Table 1 shows mean and standard deviations of conventional three parameters. Figure 5 shows the ECS and the PCS versus each conventional parameter. Also we can see the discrimination of each voice by each parameter and by the combination of the two parameters. By the values from table1 and figure 5, we can observe that any single parameter cannot provide a good discrimination. For example in case of jitter, mean is 0.6157 for normal voice and 1.9645 for pathological voice. But the standard deviation of pathological voice is 2.4118 and its range spreads to normal voice. And discrimination of pathological voice from normal voice is not possible with any single parameters. But we can observe the better discrimination between the normal and the pathological voice by combining two parameters together in figure 5.

Table1. Mean and standard deviations of parameters

Parameter		Mean	Standard Deviation
Jitter	Normal	0.6157	0.4377
	Pathological	1.9645	2.4118
Shimmer	Normal	2.2055	0.9242
	Pathological	6.5540	4.1706
NHR	Normal	0.1158	0.0124
	Pathological	0.1945	0.1241



(a)



(b)

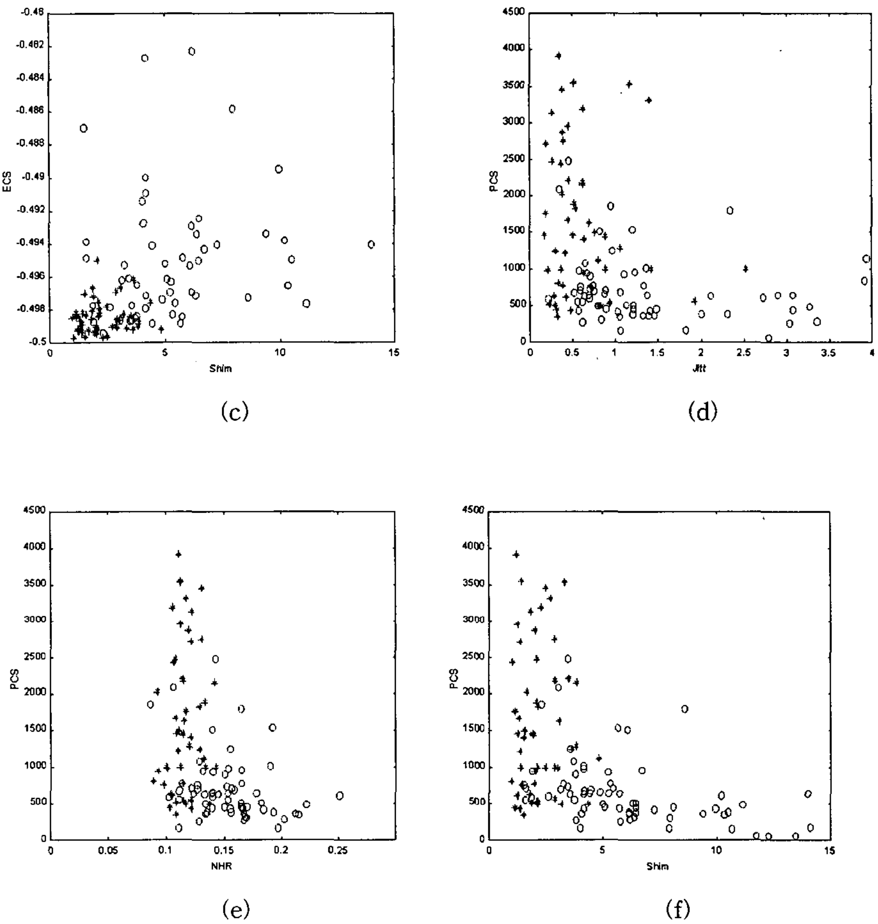


Figure 5. ECS and PCS versus jitter, NHR, Shimmer

6. Conclusions

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. Our aim is to create a new parameter which do not require an explicit pitch extraction.

Also it is verified that suggested parameters can be effectively used to diagnose the pathological voices with the combination of the conventional parameters.

The next step following this research is decision of threshold values for each parameters.

REFERENCES

[1] F. Plante, H. Kessler, B. Cheetham, J. Earis, 1996, Speech Monitoring of Infective Laryngitis, *Proceedings of ICSLP96*, 749-752, Philadelpia.
[2] M. N. Vieira, F. R. Mclnnes, M. A. Jack, 1996, Robust F0 and Jitter Estimation in Pathological Voices, *Proceedings of ICSLP96*, 745-748
[3] Disordered Voice Database, *Version 1.03, Kay Elemetrics Corp.*, 1994.

접수일자 : '98. 9. 30.
게재결정 : '98. 11. 1.

▲ 조철우
경남 창원시 사림동 9번지
창원대학교 공과대학 제어계측공학과 (우: 641-773)
Tel : (0551) 279-7552
e-mail : cwjo@sarim.changwon.ac.kr

▲ 김대현
경남 창원시 사림동 9번지
창원대학교 공과대학 제어계측공학과 (우 : 641-773)
Tel : (0551) 279-7559