Source Analysis of Pathological Voice

Cheolwoo Jo, *Member, IAENG*

***Abstract*— In this paper, we discuss the effect of voice source parameters according to the analysis of the LF model parameters from the database of pathological voice. Pathological voice consists of voice signals from the patients who have disease on their vocal folds. Vocal folds disease can affect the quality of the sound which is produced from there. KDSD v1.0 was used for the analysis[2][3]. Indirect analysis method using linear prediction was applied for the analysis. The statistical distributions of the parameters for the vowel part of the database are obtained.**

***Index Terms*—Analysis, Discrimination, Pathological Voice, Source.**

I. INTRODUCTION

There have been researches to identify the disorders of the vocal folds from the analysis of voice. Disorders on the vocal folds, such as polyp, edema, palalysis etc, cause the change of the oscillation type and accordingly the change of the voice signal. General diagnosis of the vocal folds disorder can be done visually using endoscope etc. But sometimes it is required to do the non-invasive diagnosis by the non-expert or by remote diagnosis. But some kind of vocal fold disease do not affect the oscillatory characteristics of the vocal folds. So we limit the diseases to the kinds which cause the oscillatory characteristics of the vocal folds.

The changes from the vocal folds diseases include the irregular vibration, increase of the noise components, increase of the amplitude perturbation, change of the structure of the harmonic components etc. Conventional parameters for measuring these phenomena are Jitter, Shimmer, NHR etc. The comparison of those parameter values from normal voice to those of the disordered voice can enable the diagnosis. Methods of the current discrimation include various pattern classification methods such as artificial neural network, GMM, HMM etc.[1] The major approach to such research is to find out the parameters which can discriminate the disorder and to collect the database of the disordered voice.

In this paper, voice source model parameters are used to find out the possibility of diagnosing disorders. Voice source model is the model which parameterized the shape of the vocal fold wave. The basic assumption is that if the vibratory action is affected by the disorder, it will cause the change of the source parameters. As the source model, Fujisaki-Ljungqvist(FL) model[2] is used to observe 7 parameters and 8 additional parameters are also used.

Cheolwoo Jo is with the Changwon National University, Changwon, Gyeongnam, 641-773 Korea (corresponding author to provide phone: +82-55-213-3662; fax:+82-55-262-5064; e-mail: cwjo@changwon.ac.kr).

This work was supported by Changwon National University Faculty Research Grant 2009

II. DATA BASE

KDSD v1.0 was used as disordered database.[3] It consist of 4 kinds of voice, i.e. normal, 4 kinds of benign cases, malignant cases. We used normal voice and benign voice (Nodule and Polyposis) for the experiment. Malignant cases include much noise and it is not proper to analyze at this stage and excluded. In the database consists of the vowels /a/, /e/, /i/, /u/, /o/ and some sentences. In this experiment only the center portion of the /a/ vowels is used. Signal is sampled in 50KHz, 16 bit resolution.

III. SOURCE ANALYSIS

First, glottal volume velocity wave form was computed from the inverse filtered voice using IAIF method. IAIF method is an iterative method from recursive linear predictive analysis[4]. Analysis procedures are as follows. Vowel /a/ signal is sampled from the data and one pitch period is framed to compute the parameters of the FL model. Parameter optimization method is used to find the best fit ones. Figure 1 shows the used FL model.

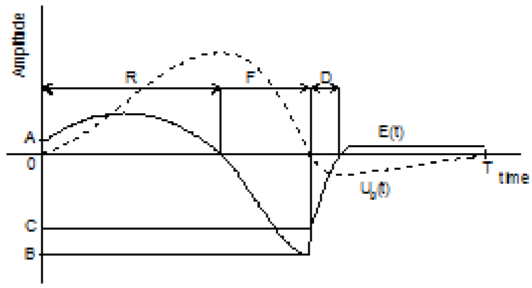

Figure 1. Fujisaki-Ljungqvist Source Model Ug(t)and derivative E(t) parameters

<Parameters>
A: Initial value of excitation signal
B: Minimum value of excitation signal
C: Start of the rising part
D: Rising part of glottal opening
R: Glottal closure
F: Falling part of glottal opening
T: Period

$$E(t) = \begin{cases} A-\frac{2A+R\alpha}{R}t+\frac{A+R\alpha}{R^2}t^2 & 0<t\leq R \\ \alpha(t-R)+\frac{3B-2F\alpha}{F^2}(t-R)^2-\frac{2B-F\alpha}{F^3}(t-R)^3 & R<t\leq W \\ C-\frac{2(C-\beta)}{D}(t-W)+\frac{C-\beta}{D^2}(t-W)^2 & W<t\leq W+D \\ \beta & W+D<t\leq T \end{cases} \quad (1)$$

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

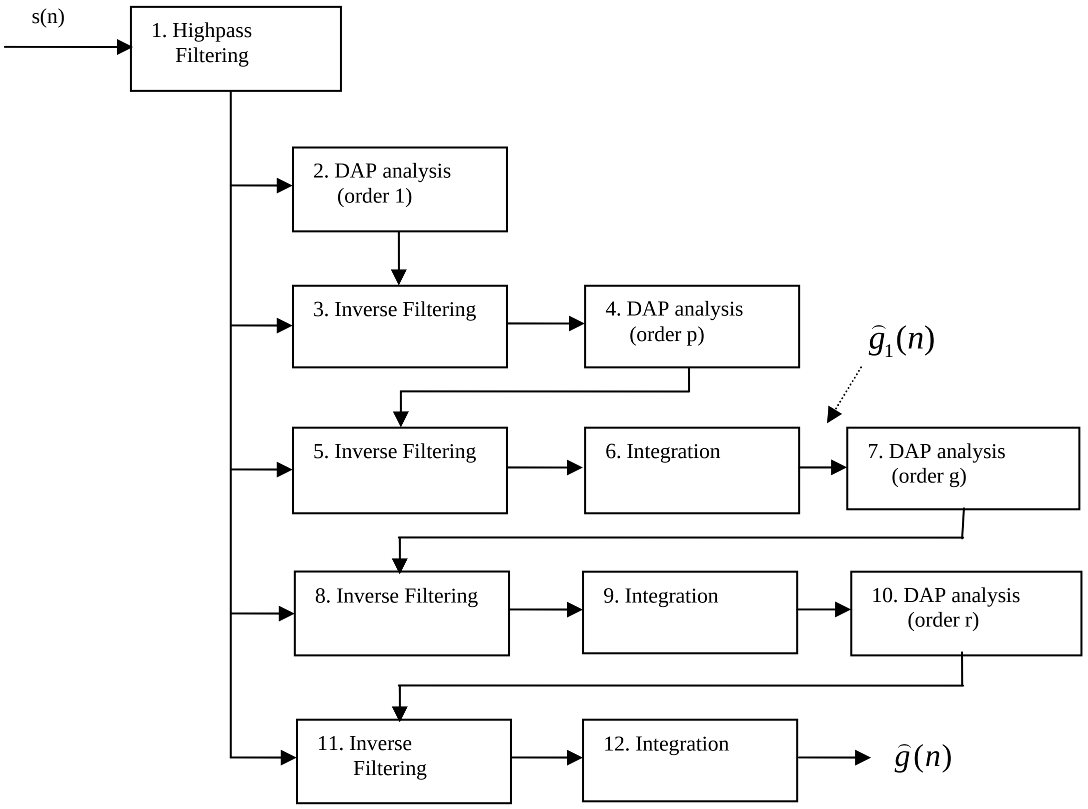

Figure 2 IAIF Method

where W=R+F and a and b can be defined as follows.

$$\alpha = \frac{4AR + 6FB}{2R^2 - F^2}, \qquad (2)$$

$$\beta = \frac{CD}{D - 3(T - W)}. \qquad (3)$$

Analysis procedures
1. Compute the coefficient from linear predictive analysis.
2. Find residual signal from the analysis
3. Estimation of glottal wave from IAIF method
4. Compute residual signal from derivative of glottal wave
5. Parameter estimation from FL model and optimization process

Matlab's fmincon function to find the parameters using optimization. The conditions of the optimization are as follows.
<Boundary condition>
D>0, R>0, F>0, A>B, C>B, T-(D+R+F)>0, R+F-D>0

To compare the difference between benign and normal voice, parameters such as OQ, DS, RS, Jitter, Shimmer, NHR, SPI, APQ, RAP were computed.
OQ is the ratio between open interval and pitch period. DS is defined as B over F, which means how fast the vocal folds close. RS is defined as C over D, which means how fast the vocal folds recover from complete closure. Other parameters are well-known parameters which are frequently used to measure the voice quality.

IV. ANALYSIS RESULTS

Figures 3,4,5 show the analysis results from normal, nodule, polyposis. From the 7 parameters obtained, Open Quotient(OQ), Down Slope(DS) values are computed. Open Quotient is defined as the ratio of open interval over the period from the source model. DS is the ratio of minimum value over the interval when the residual signal drops to the minimum value, and can be a measure of how fast the vocal folds open. If there are disorders on the vocal folds, the value can reflect the change. Among the data, in 2 benign cases and malignant case it is not possible to compute the excitation signal due to the severe noise component and they are excluded.

Figures6,7,8 are graphs whos horizontal axis is OQ, vertical axis is DS. From the scatter grams, OQ distribution is smaller than that of DS. In nodule random distribution is observed. Polyposis shows relatively constant DQ over OQ. From those distributions, we found that OQ-DS combination has some relation to the vocal folds disease. Figure9 shows the analyzed glottal wave and excitation signal.

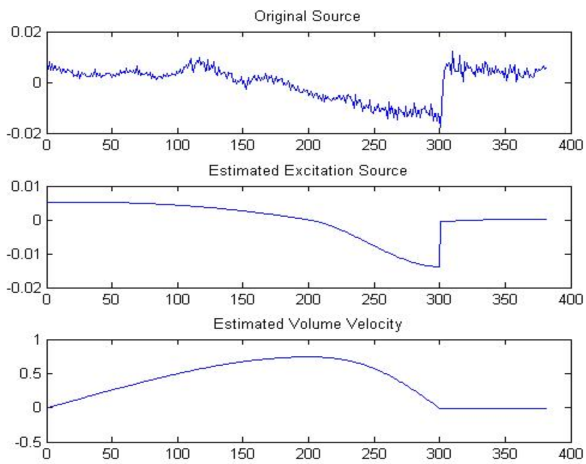

Figure 3. Analyzed Source from Normal Voice

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

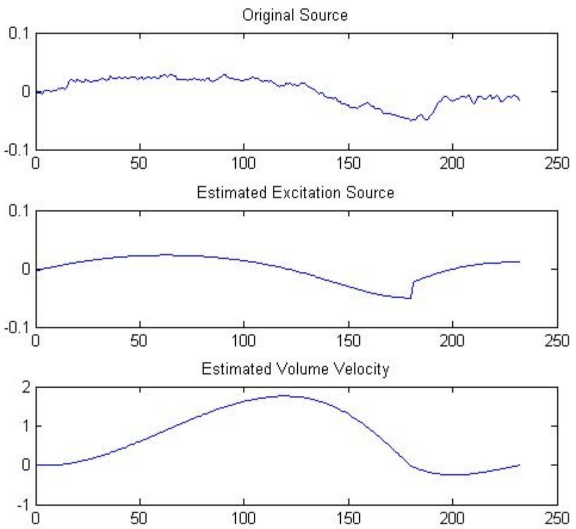

Figure 4. Analyzed Source from Benign Nodule

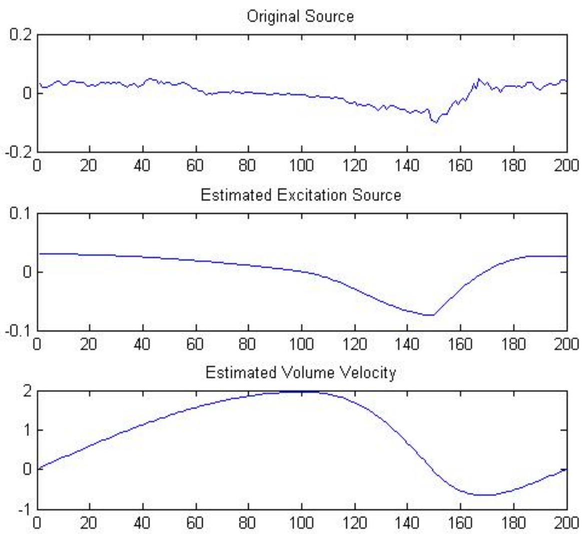

Figure 5. Analyzed Source from Benign Polyposis

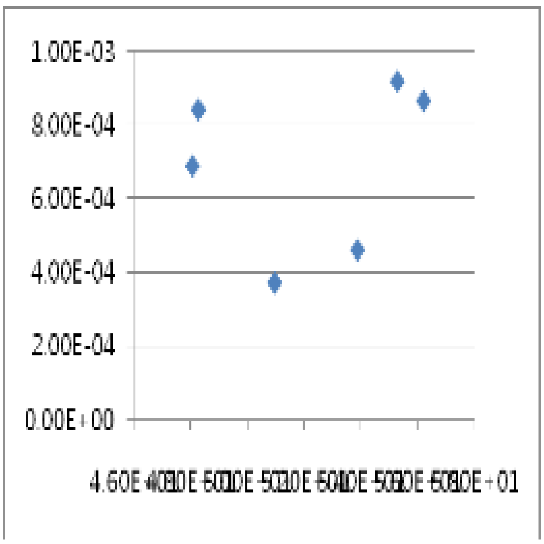

Figure 6. Distribution of OQ-DS from Nodule

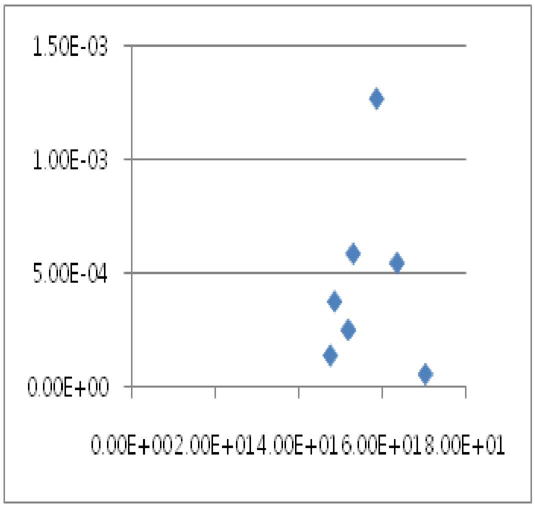

Figure 7. Distribution of OQ-DS from Normal

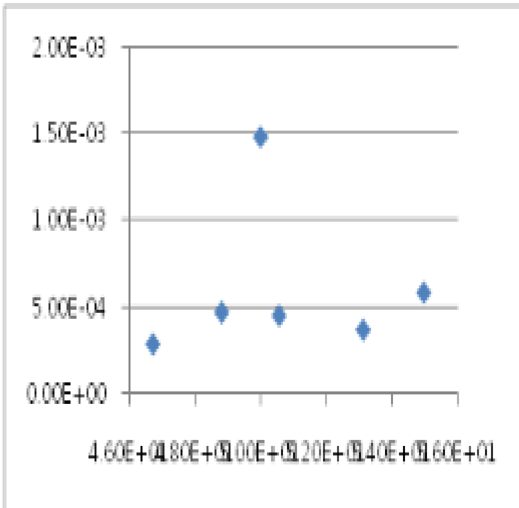

Figure 8. Distribution of OQ-DS from Polypoid

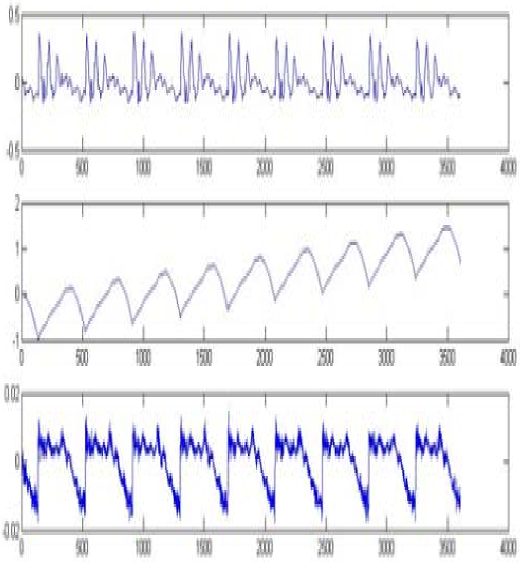

Figure 9. Source Excitation Signal by IAIF method

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

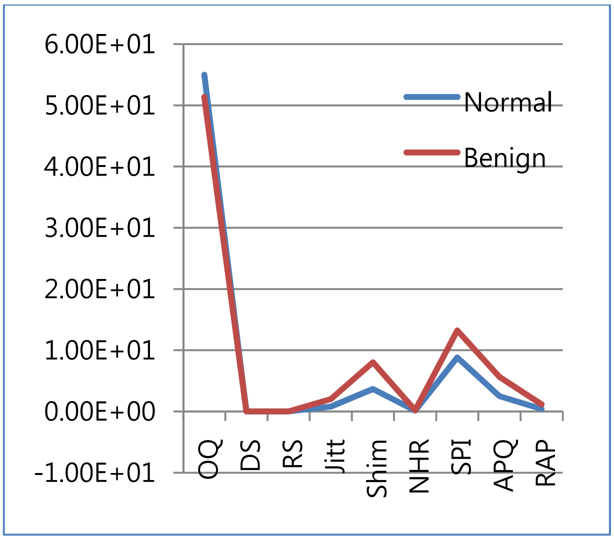

Figure 10. Variations of Parameters Computed

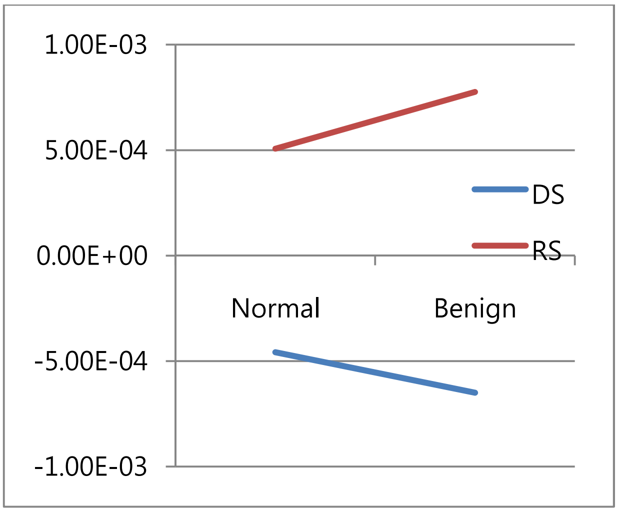

Figure 11. Comparison between DS and RS

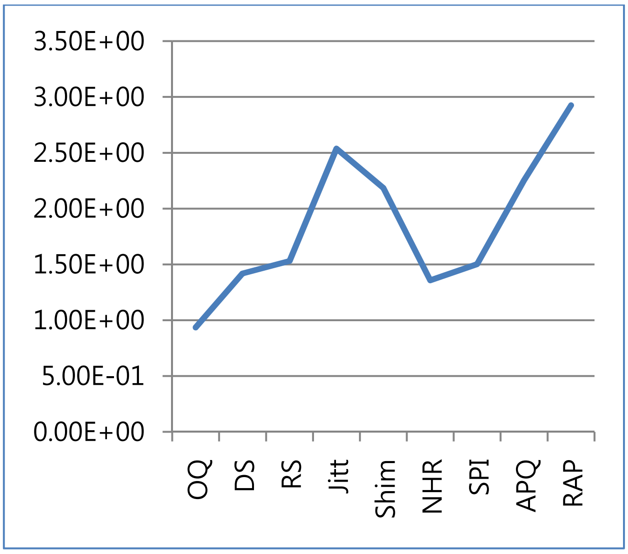

Figure 12. Normalized Values

Figure 9 shows the analyzed source waveforms from the analysis. Original wave, glottal wave from IAIF method, estimated residual signal from differentiating glottal wave.

Figure 11 is the comparison between slope related parameters. You can observe definite separation of two cases based on these parameters.

Figure 10 is the graph which compares the parameters on two cases. Because of the values of the absolute number ranges differ among the parameters, we tried to normalize those parameters based on normal cases. Simple normalization, which divided the parameter values by normal values of each parameter, is used.

Figure 12 shows the normalized distribution of the values. You can observe the Jitter and RAP most dominant compared to other parameters.

Based on the results, we can say that Jitter and RAP is better parameter to discriminate the benign and normal cases. But the parameters, which is based on glottal model, is less sensitive one compared to the conventional numerical parameters.

V. Conclusion

In this paper we tried to find out the discrimination of the parameters from the glottal source model between disordered voice and normal voice. Distribution of the parameters is compared between two groups.

In the distribution of OQ-DS, some discrimination among the three classes is possible and it is shown that those parameters can be a candidate of the discriminative parameters. But those are less sensitive ones compared to other conventional numerical parameters and it is required to modify the model when the acoustic change is severe and the signal cannot be analyzed by conventional methods.

In the future research, we need to investigate the way of improving the source model which can reflect the changes on the disordered voice better.

References

[1] Chytil,P. & Pavel, M. "Estimation of Vocal Fold Characteristics using a Parametric Source Model", Eleventh Australian International Conference on Speech Science and Technology, Auckland, NewZealand, 2006

[2] Korean Disordered Voice Database ver.1.0, Changwon National Universtiy, Korea, 1999

[3] Fujisaki, H. & Ljunqvist, M. "Proposal and Evaluation of Models for the Glottal Source Waveform", IEEE International Conference on Audio, Speech and Signal Processing, Tokyo, Japan, 1986.

[4] Alku, P. "Glottal wave analysis with pitch synchronous Interactive Adaptive Inverse FIltering", Speech Communication, 11, pp.109-118, 1992

ISBN: 978-988-18210-4-1
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)