Estimation of Multiple Source Component using Genetic Algorithm

Cheolwoo Jo, Jaehee Kim

***Abstract*—Source of speech signal consists of voiced part and unvoiced part. In conventional source-filter model, those two sources are considered to be independent. But in real situation it is difficult to segregate the source into voiced and unvoiced part. Actual source consists of mixture of two sources and the ratio varies according to the contents or intention of the speaker. In this paper we tried to segregate the components of voiced and unvoiced while considering source models. Source signals are modeled based on residual signal measured from inverse filtering. Two kinds of source models are assumed. Each model parameters is optimized to the original speech signal using genetic algorithm. The resulting parameters were compared in terms of the mel-cepstral distance to the original signal, spectrogram and spectral envelope from the synthesized signal.**

***Index Terms*—model, optimization, synthesis, voice-source**

I. Introduction

VOICE source can be utilized in various areas such as speech synthesis, speech recognition, pathological voice processing, speech coding etc. In speech synthesis, for example, voice source is very important because it has big effect on the quality of the synthesized speech in terms of naturalness, intelligibility and emotional expression. In other case, to measure the parameters of the disordered voice, there are many parameters which are related to voice source. There have been many previous researches which tried to measure the source information from the speech signal [1] [2] [3] [4].

Voice quality can be measured in various ways. The most precise way to observe the vocal folds is biological measurements. But it is not easy and not convenient. So naturally the indirect measurement from acoustic speech signal is preferred. But because of some limits on the mathematical analysis methods, there is no single way to extract voice source parameters. One simple way is to estimate the source component from the numerical analysis. Multiple source estimation was carried out in previous researches in the area of speech coding and speech synthesis. But their method focused on the approximate estimation of source by frame based analysis method and the purpose was not on finding exact ratio from the specific source model. So our aim for this research is to estimate the two components and obtain the numerical ratio of the two sources from the analysis of speech signal considering specific source model.

The authors of this paper were partly supported by the Second Stage of Brain Korea21 Project.

Cheolwoo. Jo. Author is with the School of Mechatronics, Changwon National University, Changwon 641-773, Korea (e-mail: cwjo@changwon.ac.kr).

Jaehee. Kim. Author is with School of Mechtronics, Changwon National University, Changwon 641-773, Korea (e-mail: porsche618@changwon.ac.kr).

II. Procedure for Paper Submission

In this paper, we use the inverse filtering from the linear predictive analysis to estimate the voice source. LP(linear prediction) method is a well-known method which models a signal or a system into a form of mathematical function. It is the best method measuring residual signal from LP analysis to estimate the glottal activities [5].

According to the source-filter theory, voice source consist of impulse train, which represents voiced part, and random noise, which represents unvoiced part. In simple source model speech signal is divided into voiced/unvoiced/silence part on temporal basis. Only one kind of the three can be possible in simple model. But in real situation, voiced and unvoiced part cannot be clearly separated. So in mixture source model, two types of source are considered at the same time. Yegnanarayana et.al.[6] used an iterative algorithm to separate the periodic and aperiodic components based on spectral decomposition.

In our research, we used a genetic algorithm to find an optimal level of noise sources in addition the voiced source, which is estimated from the residual signal. And we used a voice source model simulator to analyze the speech signal.

III. Voice Source Model and Basic Analysis Procedure

In this research, we considered two types of voice source signal. First one is unipolar source model, the other one is Klatt source model.

Unipolar source model is a simplified impulse model from residual signal. For each pitch period of the residual signal, only the highest peak is chosen as a candidate of the source signal. The remaining pulses are set to zero.

Klatt source model is a model which resembles to the shape of the actual glottal volume velocity. This model corresponds to the integration of the excitation, so this model

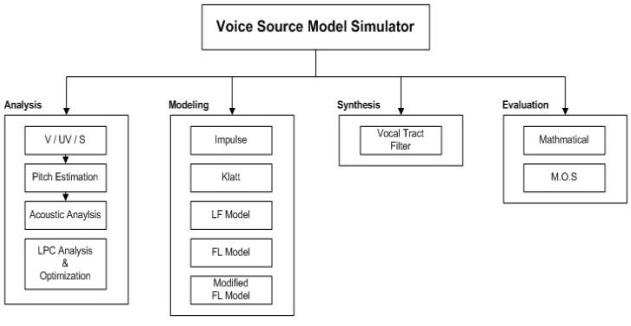

Fig. 1. Functions of the simulation

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

can be compared to the integrated residual signal.

Analysis of speech is done by conventional linear predictive analysis procedure. Residual signal is the reference source model which can be used to parameterize the voice source.

The residual signal is used to generate approximated source signal based on pitch and amplitude information of the residual signal. In each excitation position, unipolar and Klatt source shape is located.

IV. Source Optimization

Genetic algorithm was used to find the optimal noise level of the source component. Genetic algorithm uses the random and statistical method to optimize cost function. Table I shows options for GA algorithm is this research. Maximum number of iterations are set to 700. These parameters for genetic algorithm are chosen by trial and error method.

Figure 2 shows the flow of the optimization process. Based on original residual signal, noise component ratio is optimized to reduce the error between original speech and re-synthesized speech. On genetic algorithm, algorithm is iterated until the error becomes smaller than pre-specified range.

As a cost function to be minimized, the following functions were used.

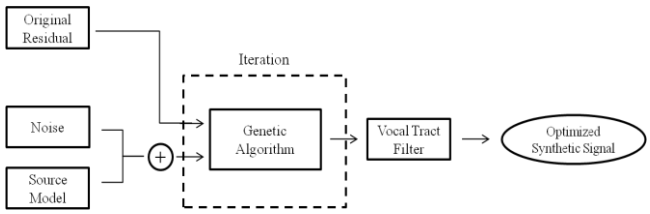

Fig. 2. Flow of optimization process

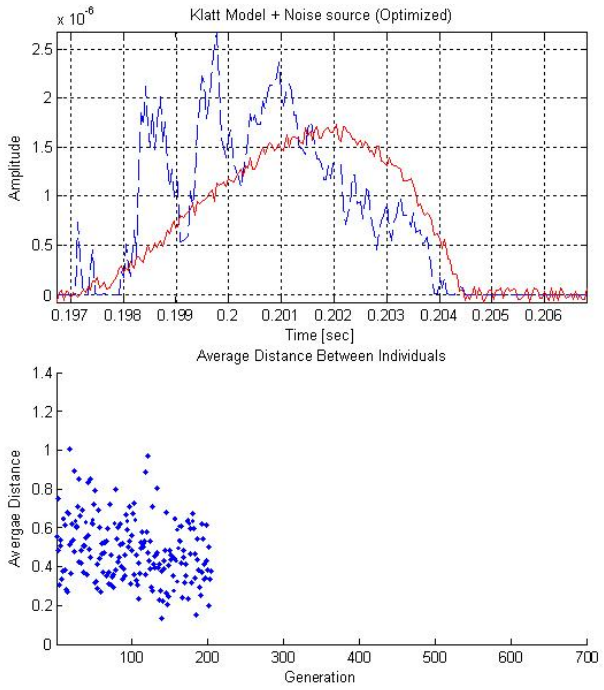

Fig. 3. Error minimization from GA

TABLE I
Options for GA algorithm

Options	Values
Population Type	Double Vector
Population Size	200
Creation Function	Uniform
Crossover Function	Scattered
Generation	700
Hybrid Function	fminunc
Mutation Function	Gaussian
Elite Count	5
Stall Generation Limit	100
TolFun	1/inf
TolCon	1/inf

For Klatt source model,

$$g(t) = \begin{cases} at^2 - bt^3, & (0 < t < O_q T_0) \\ 0, & (O_q T_0 < t < T_0) \end{cases} \quad (1)$$

$$err = |A - g(t)|$$

$$err = |\{A - (D + x(1)C)\}| \quad (2)$$

For unipolar residual,

$$err = |A - (x(1)B + x(2)C)| \quad (3)$$

Where A is the original residual, B is the modified unipolar impulse source model, C is the random noise signal and D is the optimized Klatt source model only voiced part.

Figure 3 shows the process of optimization, one pitch, the Klatt source model used as voiced source model and white random noise used as unvoiced source component.

V. Result

Figure 4, 5 and 6 show examples of comparing the original speech and the resynthesized speech in terms of time and frequency domain. And then figure 7, 8 and 9 show original signal and the result of the resynthesized signal after optimization in terms of time and frequency domain.

Figure 9 shows us modified unipolar residual and Klatt model give us close similarity in terms of spectral component. But in terms of the melcepstral distance, optimized version of the modified residual signal showed the closest to the original signal. In case of Klatt source, optimization process reduced the distance the distance considerably and it is useful to estimate noise component of the source in this way.

VI. Conclusions

In this paper we tried to estimate voice source components by applying optimization procedure to estimate the voiced and unvoiced components from the speech signal. We used genetic algorithm as an optimization method.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

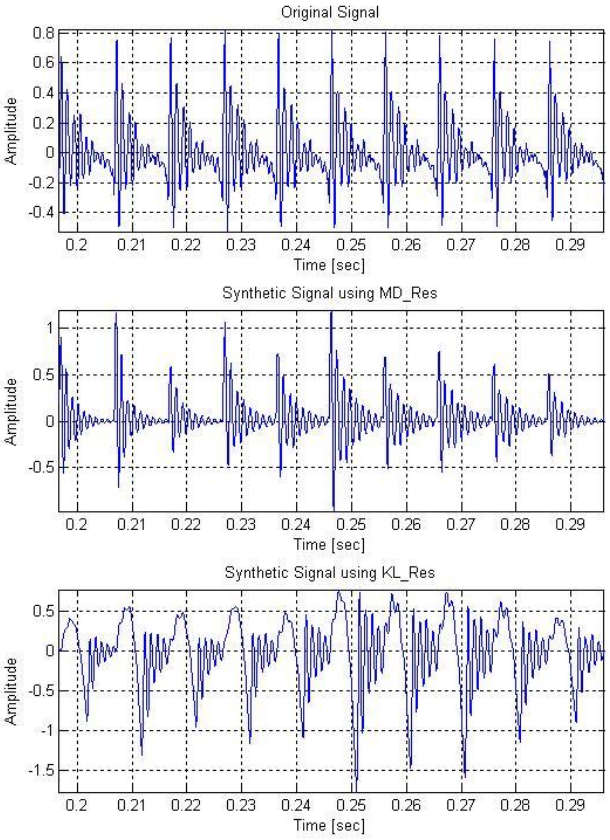

Fig. 4. Synthesized signal from each model

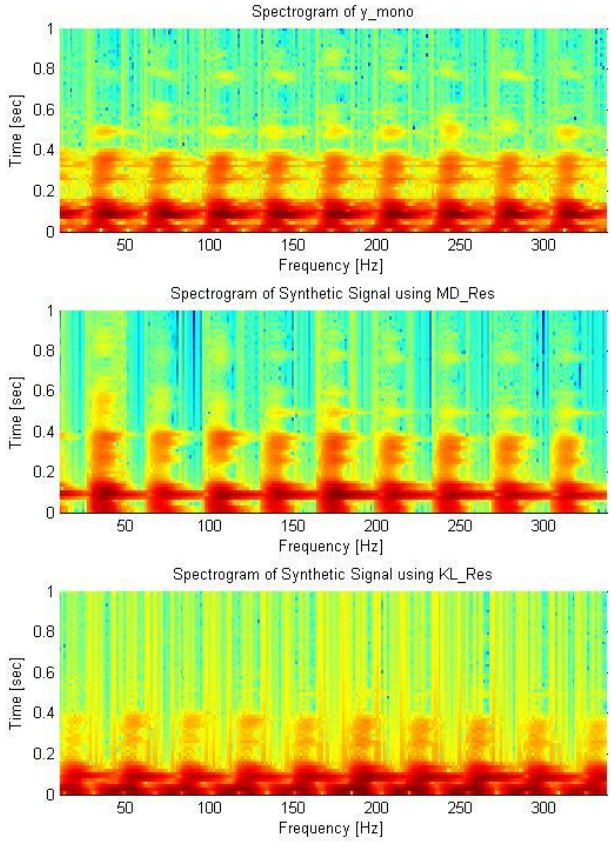

Fig. 5. Spectrogram from synthesized signal

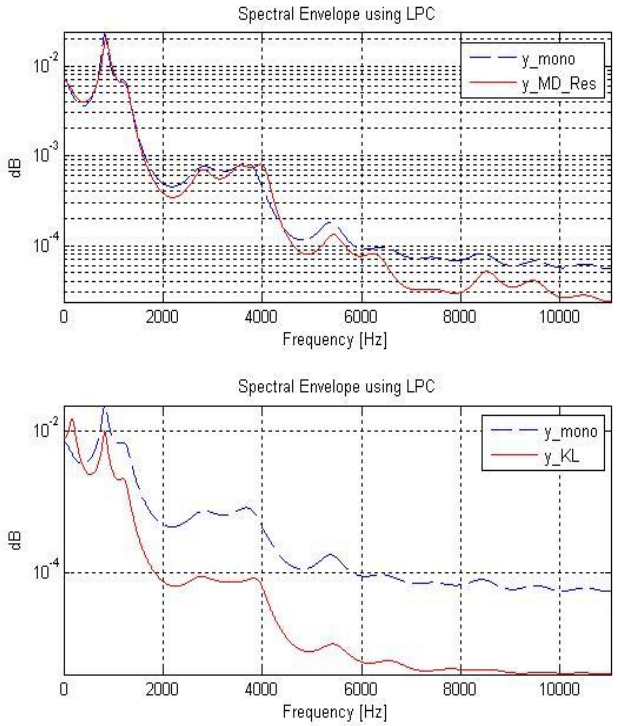

Fig. 6. Spectral envelope comparison for non-optimized case

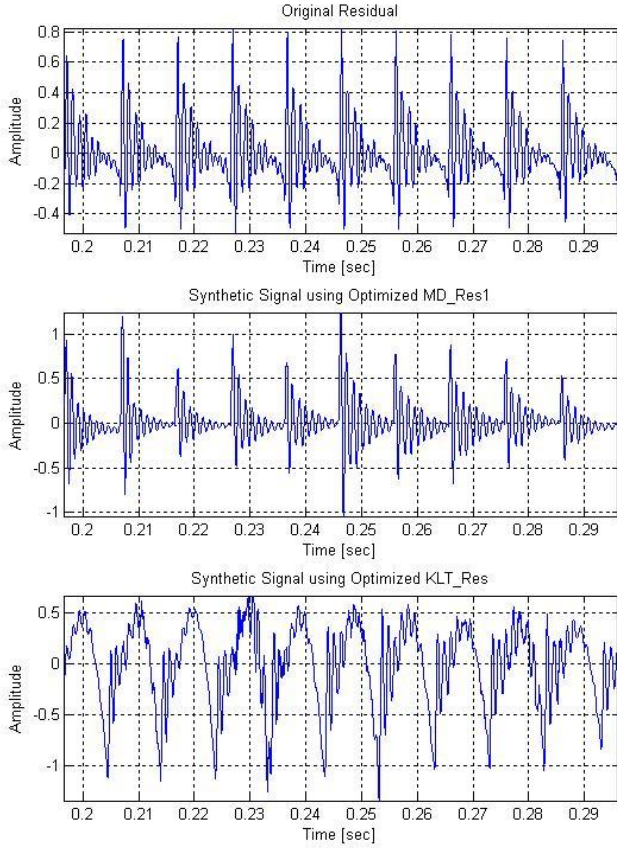

Fig. 7. Synthetic signal for optimized source models

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

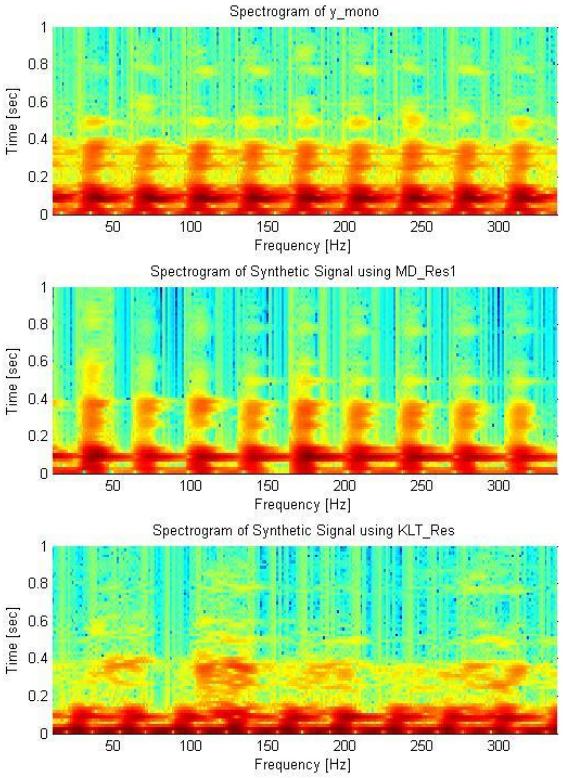

Fig. 8. Spectrogram from optimized synthetic signal

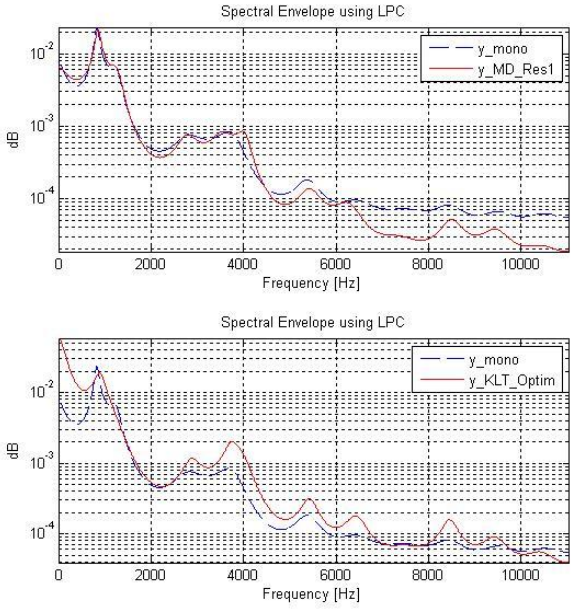

Fig. 8. Spectrogram from optimized synthetic signal

TABLE II
OPTIONS FOR GA ALGORITHM

Source Model	Distance
Modified Residual(MD_Res)	26.4
MD_Res1 (Optimized)	25.3
Klatt	181.5
Optimized Klatt	148.6

It is found out that addition of noise components with optimization procedure reduced error between the original signal and the synthesized signal when we use voice source models for research purposes. Analysis process for Klatt source model with additional noise with optimization process can be useful for the analysis of speech in various occasions such as speech synthesis or voice quality analysis or pathological voice analysis etc.

In future research, it is required to reduce the spectral distance while adding noise components to the voice source in multiple frequency bands.

REFERENCES

[1] P. Chytil and M. Pavel, "Estimation of Vocal Fold Characteristics using a Parametric Source Model," presented at Eleventh Australian International Conference on Speech Science and Technology, Auckland, Newzealand, 2006.
[2] A. Forcin and E. Abberton, "Phonetics & measurement of voice quality," *VOQUAL'03*, pp. 1-27, Aug. 2003.
[3] P. Mokhtari, H.R. Pfitzinger and C.T. Ishi, "Principal components of glottal waveform: towards parameterisation and manipulation of laryngeal voice-quality," *VOQUAL'03*, pp. 133-138, Aug. 2003.
[4] P. Alku, "Glottal wave analysis with pitch synchronous interactive adaptive inverse filtering," presented at Speech Communication, 11, 1992, pp. 109-118.
[5] J.D Markel and A.H. Gray, *Linear Prediction of Speech Communication*, Springer-Verlag, 1976.
[6] B. Yegnanarayana, d'Alessandro Christophe and D. Vassilis, "An iterative algorithm for decomposition of speech signals into periodic and aperiodic components," *IEEE Trans. Speech and Audio Processing*, vol. 6, no. 1, pp. 1-11, Jan. 1998.
[7] D.H. Klatt, "Software for a Cascade/Parallel formant synthesizer," JASA, vol. 67, no. 3, pp. 971-994, Mar. 1980.

ISBN: 978-988-18210-9-6
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)