말소리와 음성과학 제3권 제2호 (2011)
pp. 89~96

Simulink를 이용한 음원모델 시뮬레이터 구현

Implementation of Voice Source Simulator Using Simulink

조철우[1] · 김재희[2]
Jo, Cheolwoo · Kim, Jaehee

ABSTRACT

In this paper, details of the design and implementation of a voice source simulator using Simulink and Matlab are discussed. This simulator is an implementation by model-based design concept. Voice sources can be analyzed and manipulated through various factors by choosing options from GUI input and selecting pre-defined blocks or user created ones. This kind of simulation tool can simplify the procedure of analyzing speech signals for various purposes such as voice quality analysis, pathological voice analysis, and speech coding. Also, basic analysis functions are supported to compare the original signal and the manipulated ones.

Keywords: voice, source, simulator, Simulink, Matlab

1. 서론

음원은 음성의 특징을 규정짓는 매우 중요한 요소의 하나이다. 음성인식, 음성합성, 코우딩, 장애음성처리 등 모든 분야에서 음원의 역할은 중요하게 여겨지고 있다. 음성합성에서는 다양한 음질의 음성을 만들어 내기 위해서는 음원에 대한 제어가 필수적이다. 음성인식에서는 화자의 의도, 정서상태 등 까지 인식하는 단계에 들어선다면 음원정보는 필수적인 파라미터의 하나가 된다. 장애음성처리에서는 성대장애를 판별해 내는 데 기본적인 파라미터가 음원의 변동에 관한 파라미터가 된다[1][2][3][4].

그러나 음원을 직접적으로 측정하는 것은 쉽지가 않다. 생체신호를 물리적으로 측정하는 것이 가장 이상적인 음원측정의 방안이 될 수 있는데 실제로 생체에서 음원 신호를 측정하는 것은 별도의 부가장치를 사용해야 하기 때문에 어렵고 부자연스럽다. 이에 차선책으로 이용되고 있는 것이 음성신호를 역필터링한 여기신호를 통해 간접적으로 추정하는 방식이다. 정상음성의 경우 여기신호에 의한 음원추출 방법은 큰 오차없이 작동하는 방법으로 알려져 있다[5].

1) 창원대학교 cwjo@changwon.ac.kr, 교신저자
2) 창원대학교 porsche618@gmail.com

접수일자: 2011년 2월 1일
수정일자: 2011년 4월 6일
게재결정: 2011년 4월 11일

선형예측 역필터링을 통해 음원 정보를 추정하는 방법에 대한 연구는 지금까지 여러 방면으로 수행된 바가 있다. 그러나 음원에 대한 연구를 수행하는 과정에서 다양한 음성의 종류에 따른 음원을 분석하기 위해서는 통상적인 분석방법을 수시로 변형시켜가며 분석을 적용할 필요가 있다. 음원의 모델에 있어서도 기존의 정상음성을 기반으로 제안된 여러 가지 모델들이 있지만 성대의 형태 변화 등으로 인해 발성조건이 달라지는 장애음성의 경우 잘 들어맞지 않는 경우가 많다. 따라서 음성의 종류에 따라 다른 모델을 설계하고 적용해 볼 수 있는 시뮬레이터 형태의 분석 및 합성이 가능한 시스템이 필요하게 된다. 유사한 사례로는 음성 모핑을 목적으로 만들어진 STRAIGHT라는 시스템이 있다[6].

본 연구에서는 시뮬레이션 전용 프로그램인 시뮬링크와 Matlab을 이용하여 다양한 음성의 음원을 분석하고 변경이 용이한 음원 시뮬레이터를 제안하고, 구현한 사례를 소개하고자 한다.

2. 음원 분석 과정

음성신호의 음원을 측정하는 방법은 기본적으로 선형예측분석을 통한 역필터링 과정을 통해 간접 분석한다. 선형예측분석에서 구해진 여기신호는 원래의 음성신호의 여기신호와는 약간의 오차는 있지만 통상적으로 음원신호와 같이 간주되고 있으며 음성신호만에 의한 분석에서는 지금까지 음원을 추정할 수

있는 유일한 방법이 되고 있다.

본 논문에서 구현한 시뮬레이터에서는 선형예측 오차신호를 기반으로 여기신호를 구하여 다양한 형태로 변형시킨 음원신호를 생성하고 원래 음성신호로부터의 변화를 수치적, 음향적으로 비교할 수 있게 하고자 하였다.

시뮬레이터에 적용한 음원신호의 형태는 크게 두 가지로 나눌 수 있다. 여기신호로부터 변형 가공된 음원신호와 인위적으로 만들어낸 여기신호이다. 변형 가공된 음원신호는 (1) 필터링된 여기신호 (2) 여기신호에 동기된 임펄스 신호 (3) 여기신호에 동기된 다양한 음원모델(Klatt, LF, FL)에 기반한 음원신호 (4) 사용자가 정의한 음원모델 이 해당된다. 인위적으로 만들어낸 신호는 임의의 주파수와 형태로 생성한 신호로 (1) 임펄스 신호 (2) 톱니파, 정현파 등의 임의신호 들로 구성된다. 시뮬레이터에서는 이들 중 선택한 음원신호에 따라 음성을 다시 합성하여 보여주게 된다.

음원모델이 선택되면 변경된 음원모델을 그대로 적용하던지 최적화 과정에 의해 원래 음성과 최소 오차를 보여주도록 성도필터 모델을 재계산하여 설정할 수 있다.

다음은 음원모델로 사용한 여러 가지 모델에 대한 정의를 설명한다.

<그림 1>은 Klatt음원모델의 정의이다. 이 모델은 D.H.Klatt에 의해 제안된 모델로 가장 단순한 파라미터를 제공한다. 식(1)은 Klatt음원모델의 정의식이다. 이 모델은 음성의 다른 모델들과는 달리 오차신호의 적분값에 대응하여 적용한다[7].

$$g(t)=\begin{cases} at^2-bt^3, & (0<t<O_qT_0) \\ 0 \quad , & (O_qT_0<t<T_0) \end{cases} \tag{1}$$

$$a=\frac{27AV}{4T_0O_q^2} \quad , \quad b=\frac{27AV}{4T_0^2O_q^3}$$

<그림 2>는 LF음원모델이다. 이 모델은 Liljencrantz와 Fant에 의해 제안된 모델이다. 식(2)는 이 LF모델의 정의식이다[8].

$$g(t)=\begin{cases} E_0e^{at}\sin(\omega_g t) \ , & (0\le t\le t_e) \\ -\dfrac{E_e}{\varepsilon t_a}\left\{e^{-\varepsilon(t-t_e)}-e^{-\varepsilon(t_c-t_e)}\right\}, & (t_e\le t\le t_c\le T_0) \end{cases} \tag{2}$$

$$\int_0^T g(t)dt=0, \ \omega_g=\frac{\pi}{t_p}, \ \varepsilon t_a=1-e^{-\varepsilon(t_c-t_e)}$$

$$E_0=-\frac{E_e}{e^{at_e}\sin(\omega_g t_e)}$$

<그림 3>은 FL음원모델이다. 이 모델은 Fujisaki와 Ljunqvist에 의해 제안된 모델로 가장 정상음성의 음원 형태에 근접한 모델로 여겨지고 있다. 식(3)은 FL모델의 정의식이다[9].

$$g(t)=\begin{cases} A-\dfrac{2A+R\alpha}{R}t+\dfrac{A+R\alpha}{R^2}t^2, & (0\le t\le R) \\ \alpha(t-R)+\dfrac{3B-2F\alpha}{F^2}(t-R)^2-\dfrac{2B-F\alpha}{F^3}(t-R)^3, & (R<t\le W) \\ C-\dfrac{2(C-\beta)}{D}(t-W)+\dfrac{C-\beta}{D^2}(t-W)^2, & (W<t\le W+D) \\ \beta, & (W+D<t\le T) \end{cases} \tag{3}$$

$$W=R+F$$

$$\alpha=\frac{-4AR-6FB}{F^2-2R^2}$$

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

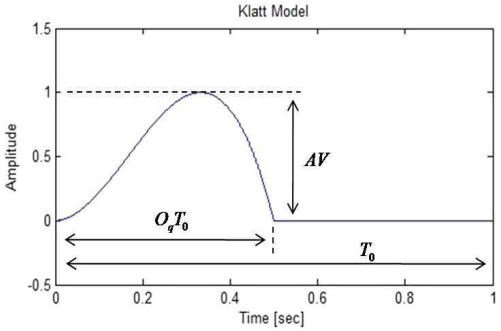


그림 1. Klatt 음원모델
Figure 1. Klatt source model

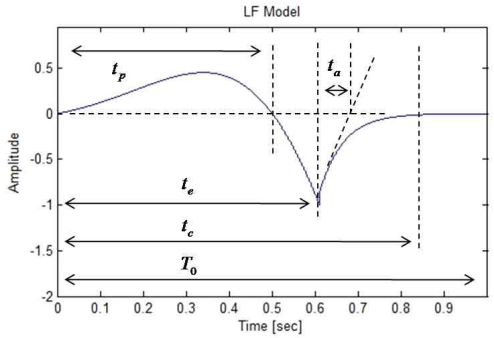


그림 2. LF 음원모델
Figure 2. LF source model

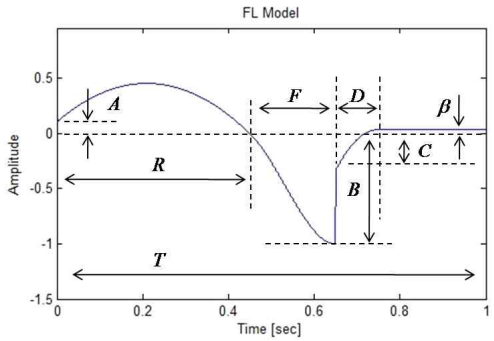


그림 3. FL 음원모델
Figure 3. FL source model

이들 모델들의 파라미터들은 모두 다르기 때문에 필요한 요소들을 지정하여 입력신호에 맞게 지정하고 변형할 수 있도록 하였다.

사용자 정의 음원모델은 기존의 모델로 대표되지 않는 음성의 경우 별도로 설정한 모델을 이용하여 음성을 분석할 수 있도록 정의한 시뮬링크 블록을 통해서 구현한다.

4. 시뮬레이터 구성

<그림 4>에 시뮬레이터의 구성을 나타내었다. 크게 분석부, 모델링부, 합성부, 평가부로 나뉜다. 분석부에서는 기본적인 분석을 수행하여 성도 및 음원정보를 선형예측 분석에 의해 계산한다. 모델링부에서는 원하는 형태의 음원모델을 지정하거나 정의하는 기능을 한다. 합성부에서는 앞에서 계산된 성도필터와 변형된 음원모델을 통해 원래 음성 또는 음원이 변형된 음성을 합성해 낸다. 평가부에서는 원래 음성과 합성된 음성과의 차이를 스펙트로그램 정보 및 음향정보를 통하여 확인할 수 있게 해 준다.

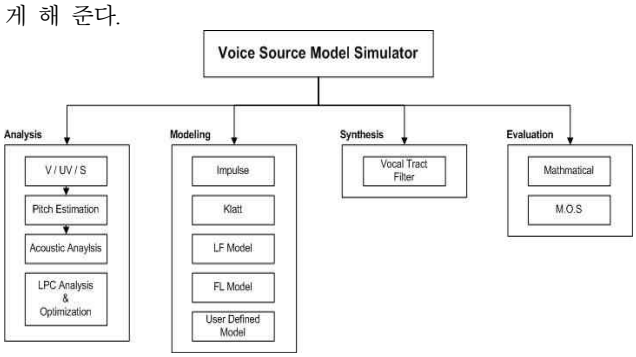


그림 4. 제안한 시뮬레이터의 구성
Figure 4. Structure of proposed simulator

<그림 5>는 시뮬레이터의 동작 흐름도이며 <그림 6>은 시뮬링크를 통해 실제로 구현한 음원 시뮬레이터의 구성도이다. <그림 5>의 음향파라미터 분석부에서는 음원의 특성을 나타내는 대표적인 파라미터인 Jitter, Shimmer, NHR(Noise-to-Harmonic ratio)을 계산하여 향후 합성할 음성에서 참조하도록 한다. <그림 5>의 모델 추정기에서는 선형예측 오차 신호로부터 원하는 형태의 음원모델을 선정하고 해당 모델의 파라미터를 구하는 부분이다. 합성 및 평가 부분에서는 구해진 음원 및 성도 파라미터로부터 음성신호를 재합성하여 들려주고 스펙트럼 분석 등을 통해 원음성과 비교할 수 있도록 한다. Residual selector는 Matlab으로 구현된 GUI(Graphic User Interface)로부터 선택사항을 받아들여 시뮬링크에서 음원을 선택하게 한다.

<그림 7>, <그림 8>은 Matlab으로 구현한 시뮬레이터의 GUI를 보여준다. 시뮬레이터에서 주요 처리과정은 시뮬링크를 통해 구현하고 Matlab에서는 초기 파라미터 입력부와 결과 출력을 처리한다. <그림 8>은 파라미터 입력부를 <그림 7>은 결과 출력부를 나타낸다. 파라미터 입력부에서는 분석할 대상 파일을 입력하고 선형예측 분석 파라미터를 지정한 다음 음원모델을 선택한다. <그림 7>은 출력부를 보여준다. 출력부에서는 비교할 두 가지 대상 신호를 선택하고 파형, 스펙트로그램을 비교할 수 있으며 직접 청취하여 비교할 수 있다. 이는 향후 MOS 방법에 의해 음질을 비교할 수 있도록 하기 위한 것이다.

<그림 9>에서는 시뮬레이터의 각 부분에 사용된 서브블록들을 설명하고 있다. (a)의 임펄스열 발생기는 음원으로 사용할 임펄스열을 정현파로부터 발생시키는 블록이다. 주어진 주파수의 정현파 신호 발생기로부터 영교차 점을 검출하여 폭이 1의 값을 갖는 임펄스 열을 발생시킨다. (b)의 피크 강조기는 여기신호로부터 피크를 강조하여 음성신호의 유성음 부분에서의 피크를 강조해 주는 부분이다. 여기신호의 최대 진폭값을 검출하

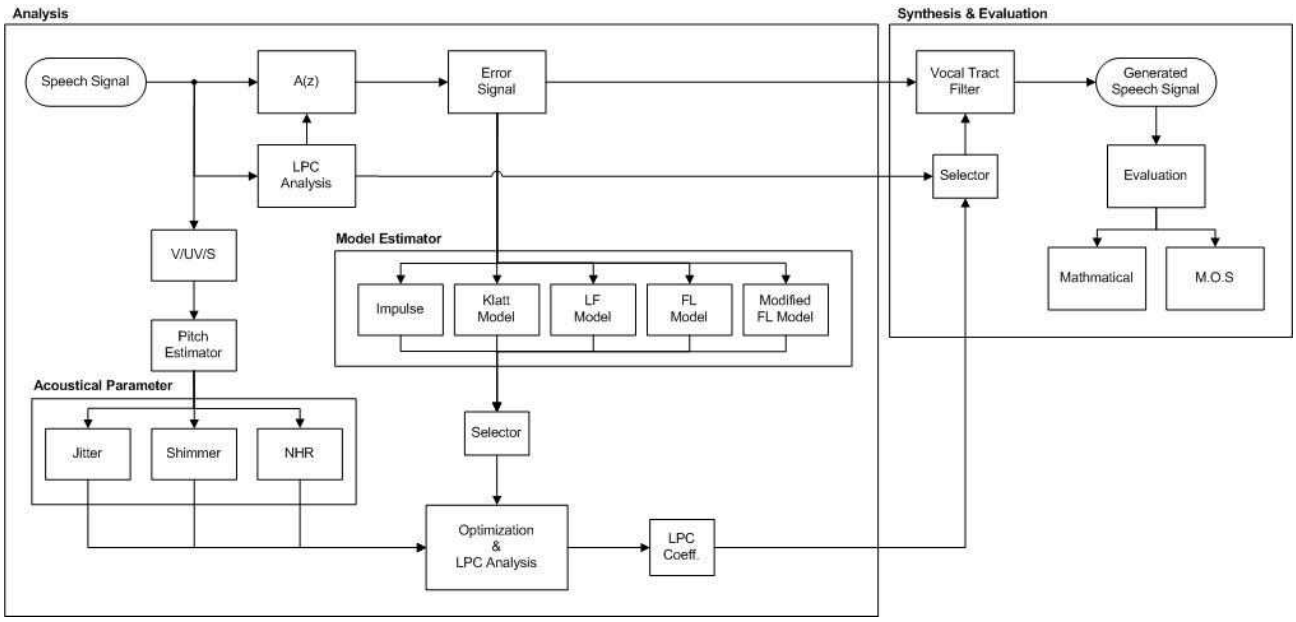


그림 5. 시뮬레이터의 동작 흐름도
Figure 5. System flow of simulator

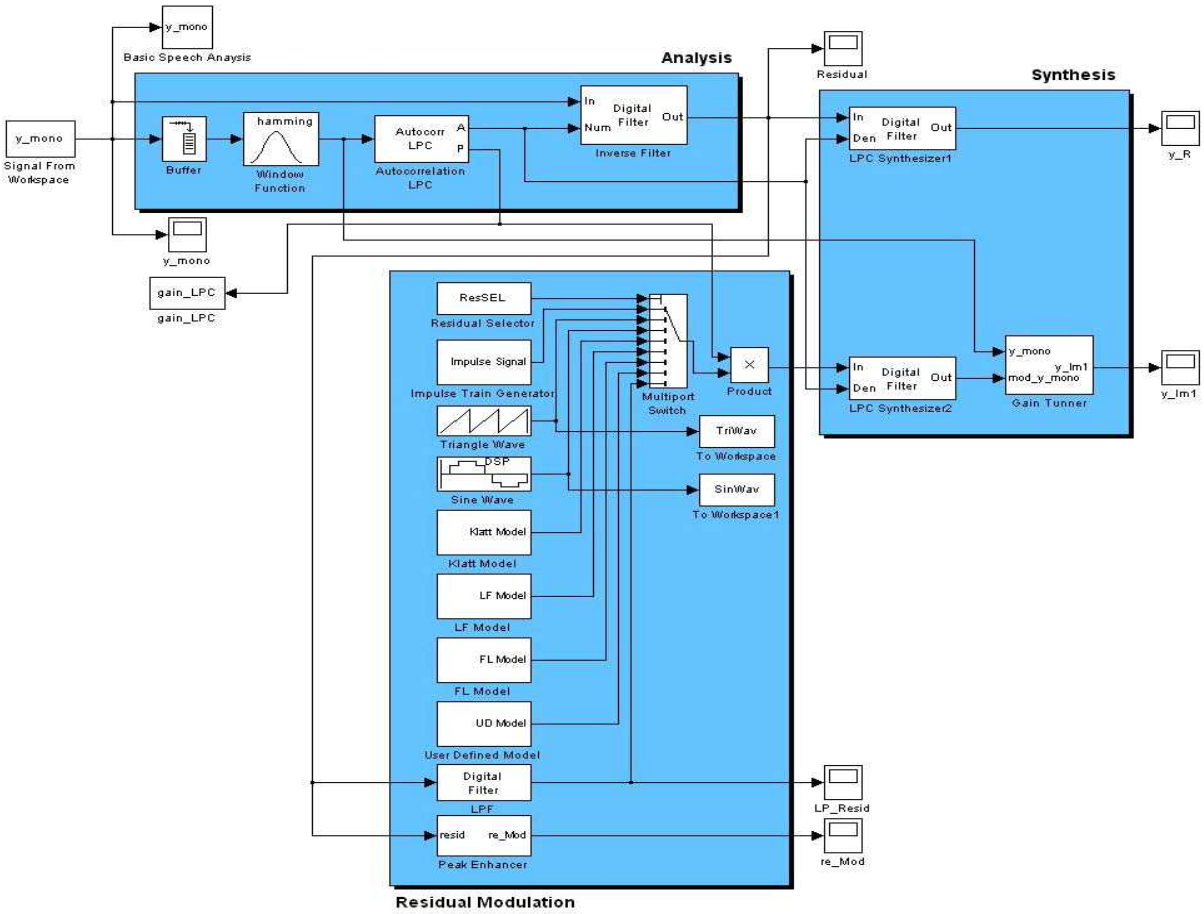


그림 6. 시뮬레이터의 시뮬링크 구현

Figure 6. Simulink implementation of simulator

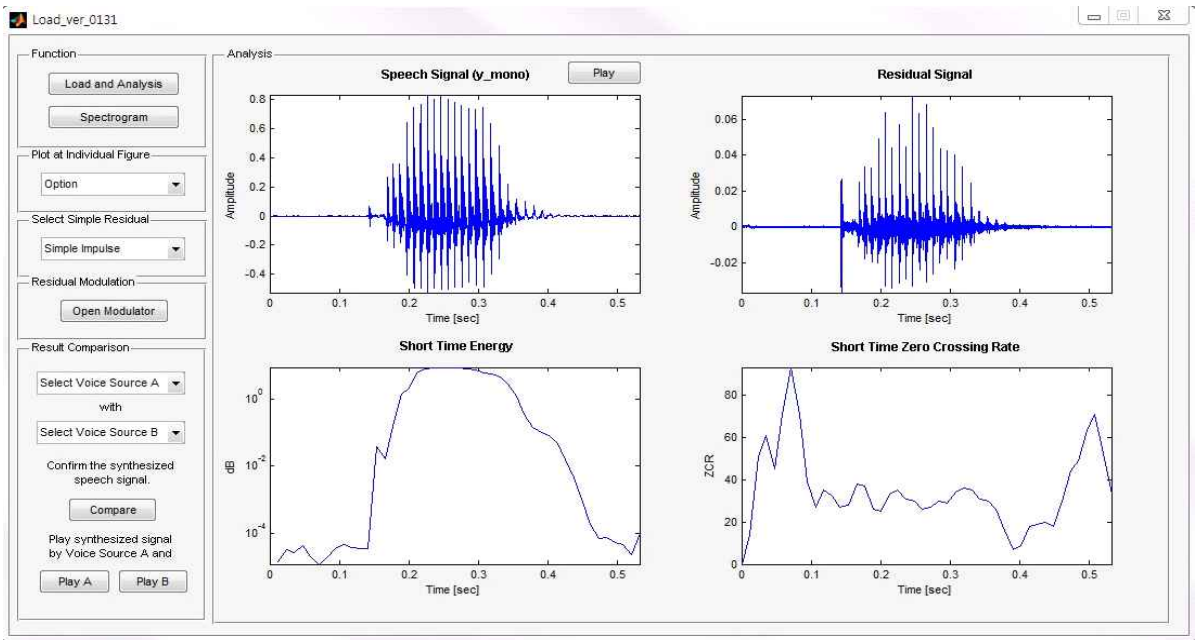


그림 7. 시뮬레이터 출력부

Figure 7. Output of simulator

여 해당 위치의 펄스를 강조해 주는 기능을 한다. 임베디드 Matlab함수를 적용하여 여러 가지 알고리즘으로 변경하여 적용이 가능하다. (c)의 신호강도 조절기는 여기신호의 크기로부터 구한 신호의 크기로부터 합성할 신호의 강도를 추정하는 부분이다. 기본적으로 선형예측분석에서 측정한 이득값을 바탕으로 합성음 생성 시 적절한 이득 값을 환산해 준다. 자세한 처리절차는 서브시스템으로 구현이 되어 있다. (d)는 변경된 여기신호로부터 합성을 행하는 부분이다. 변경된 여기신호는 원 신호로부터 추출한 여기신호를 원하는 방식으로 변형하여 성도필터에 입력한다. 시뮬링크는 시간 단위로 시뮬레이션이 진행되므로 복잡한 처리를 할 경우 입력 신호와 시간지연이 발생하게 되는데 이러한 시간지연과 무관하게 합성을 진행하기 위하여 일단 출력된 여기신호를 외부에서 Matlab프로그램을 통하여 처리한 뒤 다시 별도의 합성블록으로 합성하는 과정을 거친다.

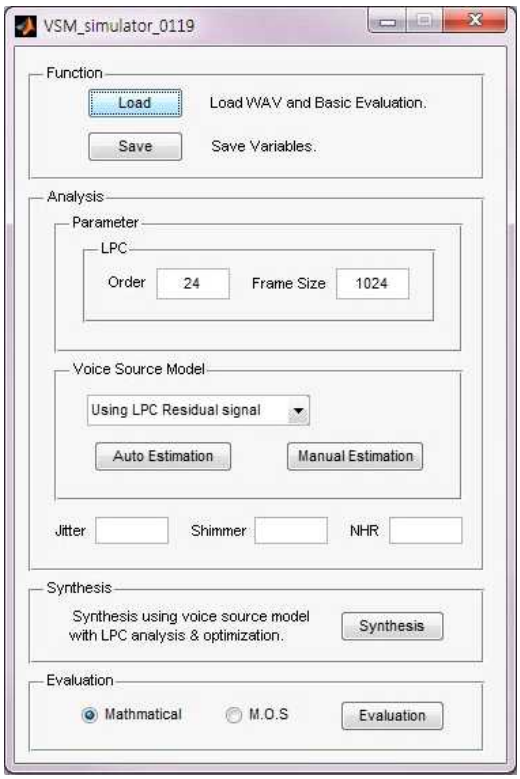
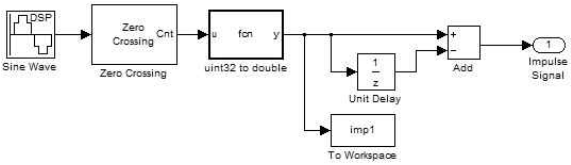
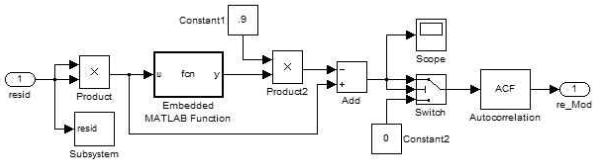


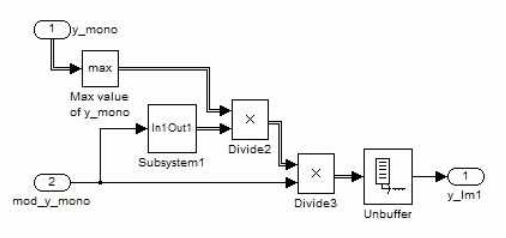
그림 8. 초기 파라미터 입력 GUI
Figure 8. GUI for parameter input



(a) 임펄스열 발생기(Impulse train generator)



(b) 피크 강조기 (Peak enhancer)



(c) 신호강도 조절기 (Gain Tuner)

그림 9. 시뮬레이터의 여러 가지 서브블록들
Figure 9. Various subblocks in simulator

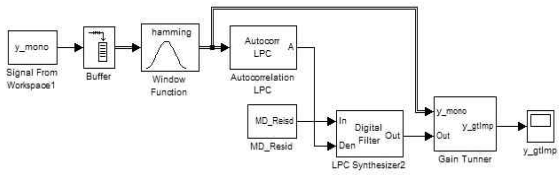


그림 10. 변경한 여기신호로부터의 합성기
Figure 10. Synthesizer from modified residual

<그림 10>에서는 변경한 여기신호로부터 음성을 합성하는 부분을 별도의 시뮬링크 블록으로 구현한 것을 보여준다. 변형된 음원으로부터 합성하는 절차는 시뮬링크의 구성상 별도로 존재해야 하므로 따로 구성하였다.

5. 실험 결과

이 장에서는 시뮬레이터로부터 구한 결과들에 대해서 설명한다. 이 실험결과들은 구현된 시뮬레이터를 통한 시험적 분석결과이기 때문에 음성 자체의 특성에 대한 분석은 상세히 다루지 않았고 다만 원하는 양상으로 음원의 변환이 제대로 이루어졌는가에 중점을 두고 결과를 비교하였다. 시험에 사용한 음성은 지속모음 /아/를 사용하였고 별도의 유, 무성, 묵음의 구분은 하지 않았다.

<그림 11>은 오차신호로부터 합성한 음성의 파형과 스펙트로그램을 보여준다. <그림 12>는 오차신호로부터 여기신호의 피치 시작점을 자동 검출하여 단순화시킨 음원모델을 생성하고 이에 따른 음성을 합성한 경우이다. 오차신호의 펄스 위치와 크기를 검출한 파형을 바탕으로 음원신호를 생성하였다. 합성된 음은 스펙트로그램 상으로 포먼트 성분이 강조된 형태를 보였고 청취 상으로는 기계적인 느낌을 나타내었다. <그림 13>은 변형된 음원의 하나로 임의의 주파수의 정현파를 입력으로 가

한 경우이다. 일반적으로 쉰 소리의 음원이 정현파에 잡음성분이 더해진 형태로 가해진다는 면에서 정현파 음원을 시험신호에 포함하였다. 현재는 유, 무성음 구분이 되지 않은 상태이지만 출력 합성음은 임펄스로 근사화한 음원에 비해 부드러운 소리를 들려주었다. <그림 14>는 Klatt모델에 의해 주어진 음성을 변환한 경우이다. 임펄스와 같은 단순 음원에 비해 보다 자연스러운 형태로 높은 주파수 대역이 감소하는 형태를 보이고 있다. 이는 Klatt음원모델이 저역통과 특성에 기인하는 것으로 음원모델의 특성이 시뮬레이터에서 정확히 구현되었음을 알 수 있다. <그림 15>는 지금까지 발표된 정상 음성의 음원모델 중 가장 정교하다고 여겨지는 FL모델에 대해 적용한 것으로 스펙트로그램과 청취실험상으로는 Klatt모델과 유사한 형태였고 가장 자연스러운 형태의 음원임을 확인하였다.

본 논문의 시뮬레이터는 동일한 신호의 재생을 목표로 하는 코우딩과는 목적이 다르기는 하지만 음원의 변화에 따른 신호의 변화정도를 살펴본다는 측면에서 수치적인 거리척도를 이용하여 원 음성과 변형된 음성 간의 차이를 측정하고 비교한 사례를 결과로 제시한다. 거리측정 척도로는 Melcepstrum 거리척도를 사용하였다.

표 1. 각 음원에 의해 합성된 음성간의 멜켑스트럼 거리

Table1. Acoustical melcepstral distance between original and synthesized voice according to different source type

	Distance
Residual	2.3×10^{-27}
Modified Residual (Pitch Detected)	131.8
Klatt	78.1
LF	126.7
FL	95.1

<표 1>에서는 원래의 음성과 변형된 음성이 어느 정도 다른가를 Mepcepstrum척도에 의해 구한 값들이다. 여기신호에 의한 합성음은 원래 음성과 거의 차이가 없지만 다른 음원을 사용한 경우 각각 다른 거리 값을 보이고 있다. 동일한 음원이라 하더라도 서로 다른 음원 파라미터 설정에 의한 합성음과 원래 음성을 이와 같은 방법으로 비교할 수 있다.

현재 시뮬레이터의 구현은 유성음원에 한정하여 구현하였으나 향후 유성음원과 무성음원의 복합적인 구성으로 분석 및 합성이 가능하게 된다면 음성의 자연성에 대한 인자를 분석하는 데도 음원시뮬레이터가 유용하게 사용될 수 있을 것으로 생각된다.

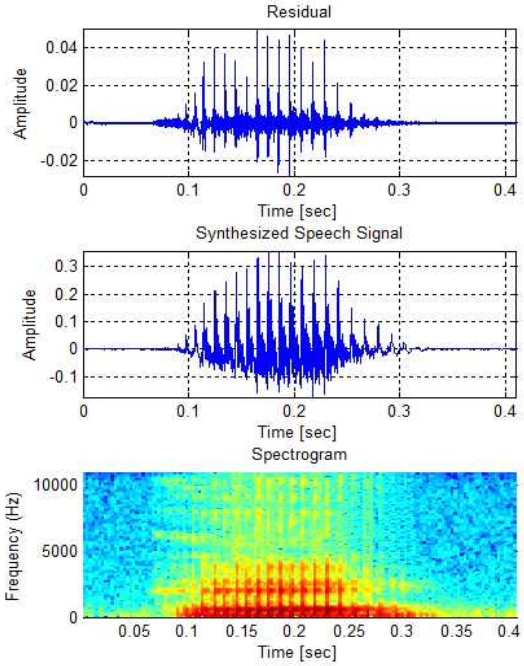


그림 11. 여기신호로부터의 합성음성

Figure 11. Synthetic speech from residual signal

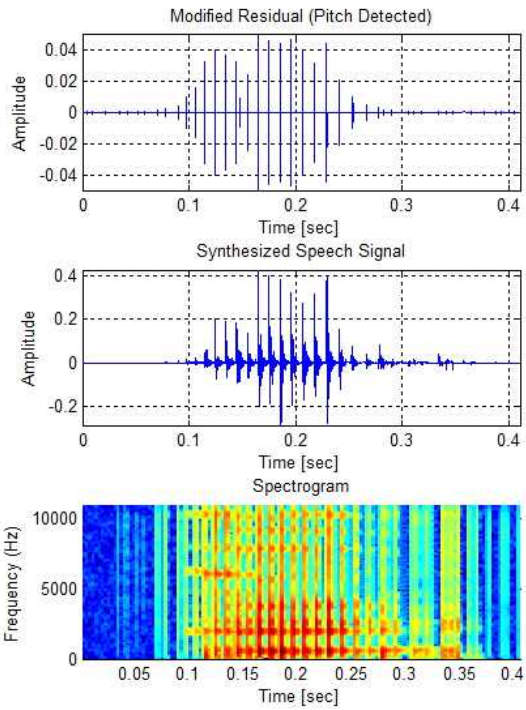


그림 12. 변형된 여기신호로부터의 합성음

Figure 12. Synthetic speech from altered residual

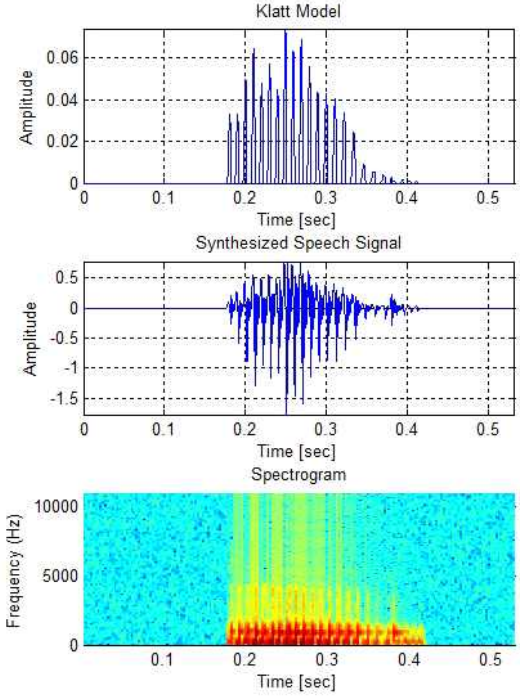


그림 14. Klatt음원에 의한 합성음
Figure 14. Synthetic speech from Klatt model

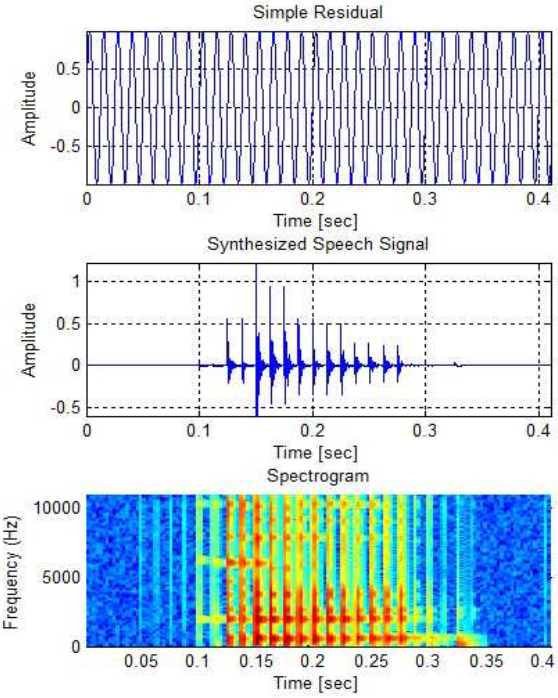


그림 13. 정현파 음원에 의한 합성음성
Figure 13. Synthetic speech from sinusoidal source

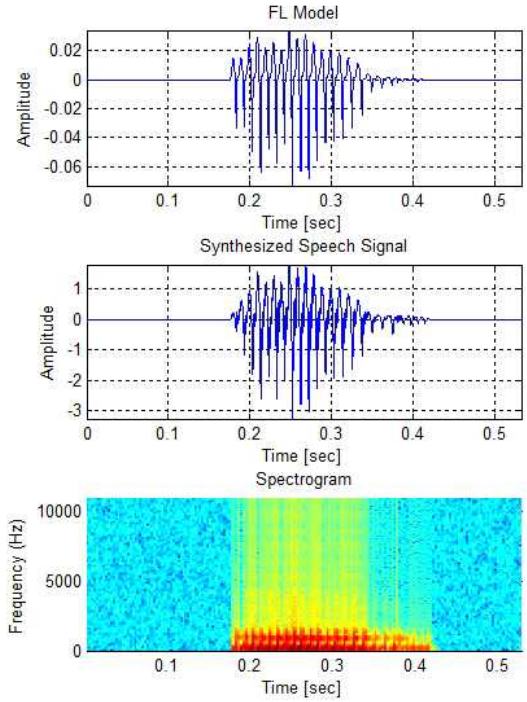


그림 15. FL음원에 의한 합성음
Figure 15. Synthetic speech from FL model

6. 결 론

본 논문에서는 음성신호를 분석하여 음원 정보를 추출하고 다양한 음원을 선택하여 합성음을 만들어 음원에 의한 효과를 측정하는 작업을 용이하게 하는 음원 시뮬레이터를 설계하고 구현하였다.

기존의 음원모델 및 가상적인 음원 모델에 대하여 원 음성을 수정하는 과정을 시뮬레이터에 의해 성공적으로 수행할 수 있었으며 변화된 음성과 원래 음성간의 차이를 비교하는 과정도 효과적으로 수행할 수 있었다.

향후의 연구에서는 기존의 음원모델을 다양화, 정교화 하도록 보완하여 다양한 발성환경에 대해 적용할 경우 생기는 현상들을 구현할 수 있도록 하고 음원모델 시뮬레이터를 이용하여 여러 형태의 음성을 분석하고 연구하는 데 적용함으로써 장애음성 및 합성음성, 음성변환 등의 분야에서 음원에 대한 여러 현상을 연구하는 데 유용한 도구로 사용할 수 있을 것으로 생각된다.

감사의 글

이 논문은 2009-2011년도 창원대학교 연구비에 의해 연구되었음. 이 연구에 참여한 연구자는 「2단계 BK21 사업」의 지원비를 받았음

참고문헌

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

[2] Forcin, A. & Abberton, E. (2003). “Phonetics & measurement of voice quality”, *VOQUAL '03*, 1-27.

[3] Mokhtari, P., Pfitzinger, H. R. & Ishi, C.T. (2003). “Principal components of glottal waveforms: towards parameterisation and manipulation of laryngeal voice-quality”, *VOQUAL '03*, 133-138.

[4] Alku, P. (1992). “Glottal wave analysis with pitch synchronous interactive adaptive inverse fIltering”, *Speech Communication*, Vol. 11, 109-118.

[5] Markel, J.D. & Gray, A.H. (1976). *Linear Prediction of Speech*, Springer-Verlag.

[6] Kawahara, H. (2003). “Exemplar-based Voice Quality Analysis and Control using a High Quality Auditory Morphing Procedure based on STRAIGHT”, *VOQUAL '03*, 109-114.

[7] Klatt, D.H. (1980). “Software for a Cascade/Parallel formant synthesizer”, *JASA*, Vol. 67, No. 3, 971-994.

[8] Fant, G., Lijencrants, J. & Lin, Q-g. (1985). “A four-parameter model of glottal flow”, *The French-Swedish Symposium*, Grenoble.

[9] Fujisaki, H. & Ljungqvist, M. (1986). “Proposal and evaluation of models for the glottal source waveform”, *ICASSP '86*, 1605-1608.

• **조철우 (Jo, Cheolwoo)**, 교신저자
창원대학교 제어계측공학과
경남 창원시 의창구 소나무5길 65
Tel: 055-213-3662 Fax: 055-262-5064
Email: cwjo@changwon.ac.kr
관심분야: 음성신호처리, 장애음성처리 및 식별
현재 제어계측공학과 교수

• **김재희 (Kim, Jaehee)**
창원대학교 제어계측공학과
경남 창원시 의창구 소나무5길 65
Tel: 055-213-3669 Fax: 055-262-5064
Email: porsche618@gmail.com
관심분야: 음성신호처리
현재 제어계측공학과 석사과정