음원 모델에 기초한 합성음의 피치 조절

Pitch Modification based on a Voice Source Model

최용진* · 여수진 · 김진영** · 성굉모***
(Yong-Jin Choi · Su-Jin Yeo · Jin-Young Kim · Koeng-Mo Sung)

ABSTRACT

Previously developed methods for pitch modification have not been based on the voice source model. Therefore, the synthesized speech often sounds unnatural although it may be highly intelligible. The purpose of this paper is to analyze the alteration of a voice source signal with pitch period and to establish the pitch-modification rule based on the result of this analysis. We examine the alteration of the interval of closing phase, closed phase and open phase using the excitation waveform as the pitch increases. In comparison to the previous methods which performed directly on the speech signal, the pitch modification method based on a voice source model shows high intelligibility and naturalness. This study might benefit the application to the speaker identification and the voice color conversion. Therefore the proposed method will provide high quality synthetic speech.

Keywords: pitch modification, Laryngograph signal, voice source model

1. 서 론

음성을 사용한 인간-기계 인터페이스(man-machine interface: MMI)기술은 컴퓨터 및 디지털 신호처리 기술의 획기적인 발전과 함께 이십여 년동안 급속한 발전을 거듭해왔다. 음성을 통한 MMI 기술은 그 편리함 때문에 멀티미디어 통신이라는 고부가가치 통신시대가 도래함에 따라 더욱더 필수적으로 되어가고 있는 실정이다. 음성 MMI 기술은 크게 음성인식 기술과 음성합성 기술로 구성되는데, 이중 음성합성 기술은 인간의 음성발생메커니즘을 이해하고 분석하여 임의의 문장을 음성으로 변환하는 text-to-speech 기술이다. 음성합성 기술은 미국, 일본, 유럽 등의 선진국에서 활발히 연구가 진행되어 왔으나 언어라는 특수성 때문에 한국어에는 적합하지 못하여 국내에서도 십여 년간 연구가 지속적으로 이루어져왔다. 그 결과 몇몇의 한국어 무제한단어 음성합성 시스템이 연구 완료되어 사용되고 있으나 이들 시스템의 합성음을 들어보면 이해도의 측면에서는 어느 정도 양질의 성능을 보이고 있으나 자연도의 측면에서는 아직까지도 부족한 실정이다. 따라서 현재 한국어 음성합성기의 고품질화가 시급한 실정이며 고품질화가 이루어질 때 한국

* LG종합기술원
** 전남대학교 전자공학과
*** 서울대학교 전기전자공학부

어 합성시스템은 여러 응용분야에서 적극적으로 채용될 것으로 보인다.

음성합성기의 고품질화를 위해 합성음의 자연성을 향상시켜 주고 동시에 음색을 제어할 수 있는 피치조절 방법이 필요하다. 기존의 피치조절 방법들은 음성 파형만을 고려하여 피치를 조절했으므로 합성음의 자연성을 해치는 많은 문제점들을 가지고 있고 음성의 개인성 보존이나 음색제어와 같은 고품질의 음성합성기가 갖추어야할 요건들을 구현하기 어렵다. 그러므로 이러한 문제들을 극복하기 위해 음성 파형뿐만 아니라 음원 여기신호를 고려하여 피치를 조절하는 방법을 연구하게 되었다. 피치에 따른 음원 여기신호의 변화를 분석하여 이 결과를 피치조절 방법에 이용하였다. 만약 각 개인의 음원 여기신호의 파라메타들을 추출하여 원하는 음색의 파라메타로 바꾸어서 합성을 하면 음색이 제어된 합성음을 얻을 수 있으므로 음원 여기신호를 이용한 피치조절은 음색제어 분야에도 응용할 수 있다.

본 논문의 구성은 기존의 피치조절 방법들의 종류와 문제점을 알아보고 음성발생 원리와 Laryngograph(Lx)신호에 대해 간략히 설명한 다음, 음성신호와 Lx신호를 이용하여 음원 여기신호를 추정하고 피치에 따른 음원 여기신호의 특성을 조사한다. 또 피치조절을 위한 규칙을 세우고 이를 바탕으로 새로운 피치조절 방법을 제안하고 이에 대한 과정을 설명하며 기존의 피치조절 방법과 제안한 피치조절 방법에 의한 합성음들을 청취테스트를 하고 그 성능을 평가한다.

2. 음원 여기신호 분석

2.1 음성 발생 모델 및 기존의 피치조절 방법의 문제점

그림 1에 나타낸 일반적인 음성 발생 모델은 음원과 성도의 두 가지 선형필터의 직렬 연결로 모델링할 수 있다. 그런데 기존의 피치조절 방법들은 이러한 음성 발생 모델에 기반하지 못하여 합성음의 피치조절 시에 음질의 저하를 야기한다.

예를 들어 현재 사용되는 피치조절 방법들 중 프랑스의 CNET에서 개발한 PSOLA 조절방법이 주류를 이루고 있다.[3] 가장 먼저 개발된 TD-PSOLA(Time Domain Pitch Synchronous OverLap Add) 방법은 시간영역에서 음성신호에 pitch synchronous하게 창함수를 곱하여 short-term(ST)신호의 열을 만들어 피치를 높일 때에는 ST신호의 간격을 작게 배열하고 피치를 낮출 때에는 간격을 크게 배열하여 겹치는 부분은 단순히 더하면 된다. 또 음성신호를 DFFT하여 주파수상에서 피치를 조절하는 FD-PSOLA가 있고 LPC 계수를 이용하여 잔차신호를 만든 후 이 잔차신호를 TD-PSOLA와 같은 방식으로 피치를 조절하여 다시 합성필터를 통과시킴으로써 합성음을 얻는 LP-PSOLA방법이 있다. 또한 LPC 계수를 이용하는 Wavelet 방식이 있다[2].

그런데 TD-PSOLA와 FD-PSOLA는 파형 자체만을 이용하여 피치조절을 하므로 음원과 성도를 구분하지 않으며, TD-PSOLA와 Wavelet 방식은 음원과 성도를 LPC 분석을 통하여 구분하지만 추출된 잔차신호가 실제의 음원과 차이가 난다는 문제점이 있다.

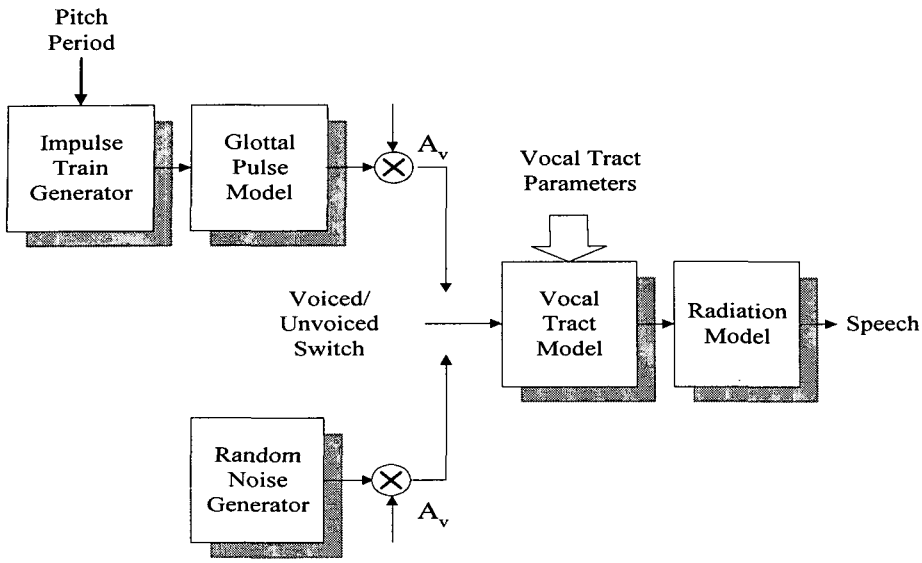


그림 1. 일반적인 음성발생 모델

이와 같이 기존에 개발된 여러 가지 피치조절 방법들은 이해도면에서 어느 정도 양질의 소리를 내지만 음질을 해치는 여러 가지 문제점들을 가지고 있다. 가장 큰 문제는 주파수상에서의 감쇠 및 소리 울림 현상 등으로 인하여 음질이 저하되는 점이다. 이러한 문제점들의 원인은 음성이 발성되기 전의 음원 여기신호를 고려하지 않고 단지 음성 파형만을 고려하였기 때문이다. 그러므로 피치에 따라 음원 여기신호를 분석하고 이를 피치조절에 이용하고자 한다.

2.2 음원 여기신호 추정 및 분석

2.2.1 음원 여기신호 분석 방법

그림 2에는 음원 여기신호 분석의 각 단계가 설명되어 있다. 먼저 프리엠퍼시스 단계에서는, 음성신호가 방사될 때 고주파성분이 감쇠 되므로 이를 보상해 주기 위하여 음성신호를 전달함수가 $H(z)=1-0.98\,z^{-1}$인 고역통과 필터에 통과시킨다. 그 다음 프리엠퍼시스된 음성신호를 이용하여 10차 공분산 방법에 의해서 LPC 계수를 구한다. 이때 음원과 성도간의 상호영향을 피하기 위해 성문이 닫히는 시점에서 열리기 전까지 구간의 음성신호를 이용한다. 이후에 LPC 계수를 이용하여 성도필터를 구현하며 프리엠퍼시스되지 않은 음성신호를 전달함수가 성도필터의 역과 같은 필터에 통과시킴으로써 음원 여기신호를 추정할 수 있다.

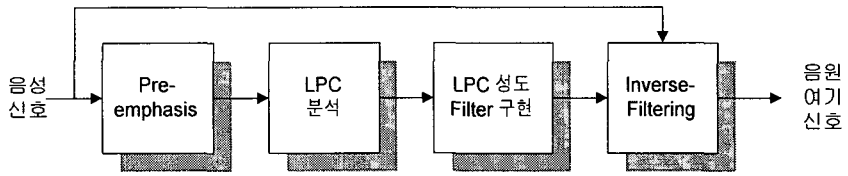


그림 78. 음원 여기 신호의 분석 구성도

2.2.2 Laryngograph(Lx) 신호를 이용한 음원 구간 분석

성도필터를 구현함에 있어서 가장 큰 어려움은 음성신호에서 성문의 닫힌 구간의 위치를 알아내는 것이다. 이 문제를 해결하기 위해 성대의 진동을 관측하기 위해 만들어진 방법중 하나인 Laryngograph(Lx)신호를 이용하였다. 성대가 위치한 후두 좌우에 2개의 금속전극을 부착시키고 일정한 값을 갖는 약한 고주파 전류나 전압을 인가하면 성대의 진동에 따라 얇은 막을 통과하는 전류의 일부분이 변조되어 두 전극사이에 임피던스가 변하게 되는데 그 변화를 전기신호로 검출한 것이 Lx신호이다. 유성음을 발성할 경우 Lx신호의 파형은 성대의 진동운동에 의해 단조증가 및 단조감소 현상을 나타내게 된다. Lx신호가 비록 음성과 성대의 접촉면적과의 대응관계를 명확하게 설명하기에는 충분하지 못하지만 매 주기에서 경사진 편향을 나타내는 간단한 파형은 음성의 피치검출과 성문의 열리고 닫히는 구간을 측정하는데 매우 적합하다.[14][15]

만일 음성의 피치가 매우 높을 경우 또 다른 문제가 발생하게 되는데 그러면 성문의 닫힌 구간이 매우 짧게 되어 성도필터를 구현하는데 필요한 LPC 계수를 성문의 닫힌 구간에서 추정하기가 매우 어려워진다. 그러므로 분석을 위한 음성 중 비음의 경우는 제외시킨다. 실험에 사용된 음성DB는 비음이 아닌 남성화자에 의해 발성된 모음으로 하였고 샘플링 주파수는 11025 Hz이다. 음성신호와 Lx신호를 동시에 녹음할 때 마이크와 발성화자의 거리 차로 음성신호가 Lx신호에 대해 약간의 지연이 발생하므로 녹음 후 이를 보상해 주었다.

그림 3에서는 Lx신호를 이용하여 성문의 닫혀지는 구간과 닫힌 구간, 열린 구간을 찾는 방법과 음원 여기신호를 이용하여 구간들을 찾는 방법이 비교되어 있다. 주기적인 변화를 하는 Lx신호에서 한 주기는 진폭이 급격히 증가하는 부분에서 시작한다.

- **성문이 닫혀지는 구간**

Lx신호에서 단조증가 하는 부분으로 주기의 시작에서 진폭 최대가 되는 점까지를 구간으로 잡고 Lx신호를 미분한 파형에서는 0을 지나는 점까지를 구간으로 했으며 음원 여기신호에서는 단조증가를 하다가 값이 일정하게 유지되기 전까지를 구간으로 잡는다.

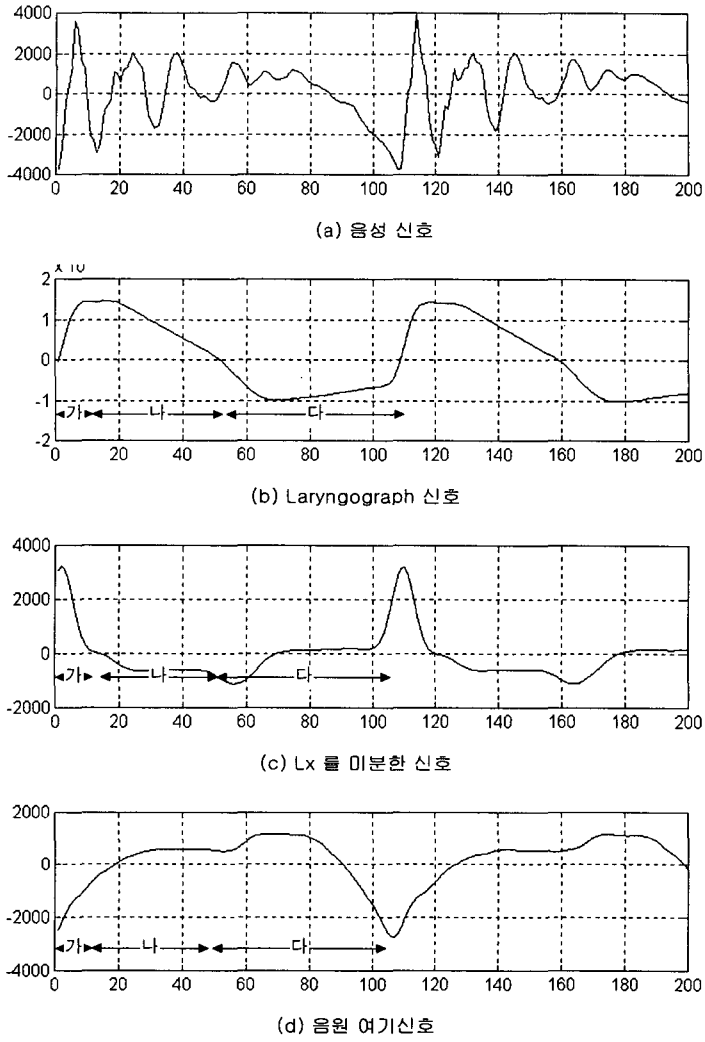
- **성문이 닫힌 구간**

Lx신호에서 단조감소를 하는 부분으로 최대점에서 변곡이 일어나는 점까지를 구간으로 했으며 Lx신호를 미분한 파형에서는 0을 지나는 점에서부터 최소값 이전에서 급격한 감소를 보이는 점(두 번 미분했을 때 0을 갖는 점)까지로 값의 변화가 적은 부분이며 음원 여기신호에서도 0에 가까운 거의 일정한 값을 갖는 부분이다.

- **성문이 열린 구간**

각 파형 들의 나머지 부분들로 가장 넓은 구간을 차지한다. 음원 여기신호에서는 성문이 닫힌 구간 다음으로 값이 증가하기 시작하는 점을 구간의 시작으로 본다.

그림 2에서 보는 바와 같이 Lx신호를 이용하여 찾은 구간과 음원 여기신호를 이용하여 찾은 구간이 일치함을 알 수 있다. 그러나 음원 여기신호에서 닫힌 구간의 값이 항상 일정한 것이 아니라 증가나 감소를 하는 경우가 있는데 이럴 경우 Lx신호를 이용하면 쉽게 구간을 찾을 수 있다



가-닫혀지는 구간, 나-닫힌 구간, 다-열린 구간

그림 79. 성문의 열리고 닫히는 구간

2.3 피치에 따른 음원 신호의 특성

피치와 성문이 열리고 닫히는 관계를 알기 위해 피치의 변화에 따른 성문의 열리고 닫히는 구간의 변화를 조사하였다. 녹음방법은 각각의 화자가 각각의 모음에 대해 낮은 피치부터 높은 피치까지 한번에 연속 발음하도록 하여 조사하였다.

- **성문이 닫혀지는 구간**

음원 여기신호나 Lx신호의 첫 번째 구간으로 다른 구간에 비해 간격이 짧고 피치에 따른 간격의 변화가 거의 발생되지 않는다.

● **성문이 닫힌 구간**

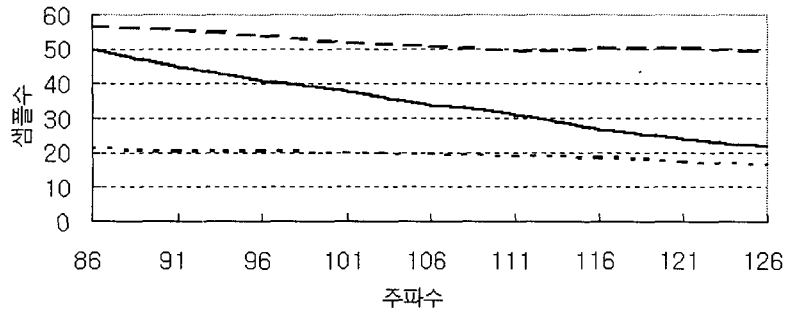
음원 여기신호와 Lx신호의 두 번째 구간으로 피치가 증가됨에 따라 즉 한 주기의 간격이 좁아짐에 따라 구간의 간격이 급격히 감소됨을 보였다.

● **성문이 열린 구간**

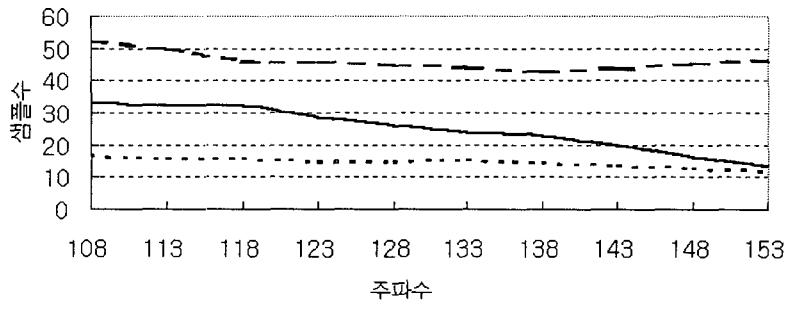
전체 구간 중 가장 넓게 차지하는 구간으로 피치가 증가됨에 따라 약간의 감소는 보이지만 비율로 따져볼 때 피치변화에 크게 영향을 미치지 않는 것으로 여겨진다.

동일한 피치에서 각 부분의 간격을 화자별로 비교해 보면 간격이 각각 다르게 나타난다. 그 원인은 화자의 음색에 달려 있다고 볼 수 있다. 그러므로 음색제어 연구나 화자의 개인성요인의 연구에서 이에 대한 상관관계를 분석하는 것도 중요할 것이다. 그림 3은 피치의 변화에 따른 각 부분들의 간격변화를 5명의 화자에 대해 나타내었다. 그림에서 보는 것처럼 모든 화자에 대해서 거의 비슷한 추세가 나타나는데 닫힌 구간의 변화가 피치가 높아짐에 따라 급격히 감소함을 알 수 있고 나머지 구간에서는 변화가 별로 일어나지 않는다. 그러나 아주 낮은 피치와 아주 높은 피치를 비교해 본다면 열린 구간도 약간의 변화가 있음을 알 수 있다. 또한 이것은 /아/ 모음뿐만 아니라 다른 모음에서도 이와 같은 결과를 얻을 수 있다.

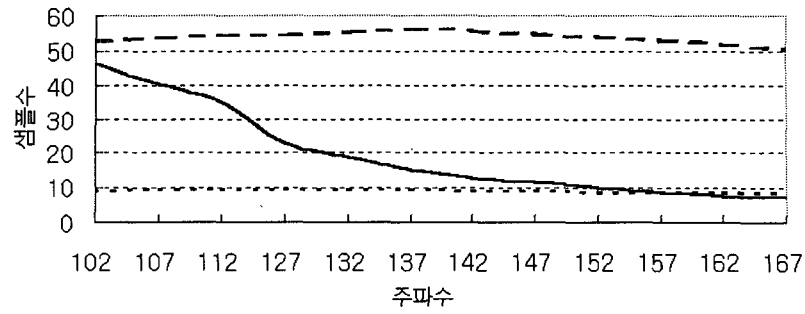
따라서 피치의 높낮이는 성문의 닫힌 구간의 간격과 관련이 있음을 알 수 있다. 즉 피치가 높아질수록 음원 여기신호의 닫힌 구간의 간격이 줄어든다. 이러한 닫힌 구간에 대한 특성을 이용하여 피치를 조절하는 방법을 제안하고 그 성능을 평가하여 보겠다.



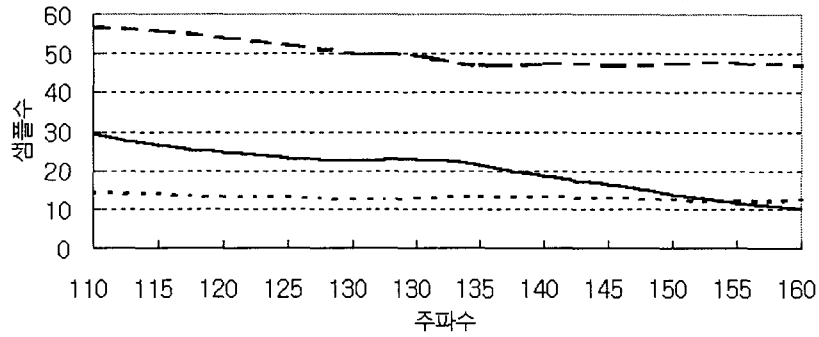
(a) 화 자 A



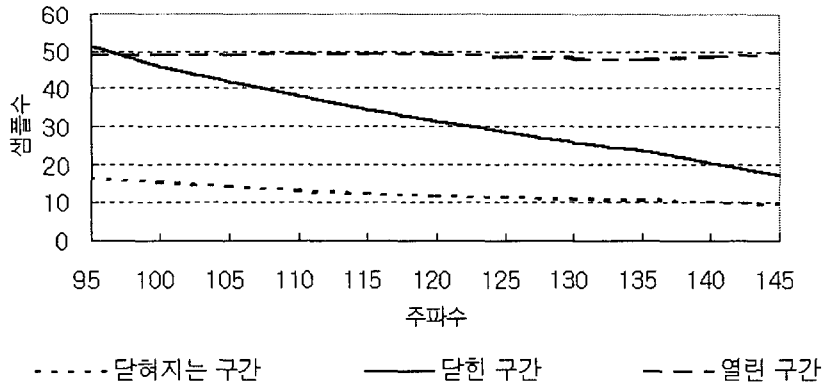
(b) 화 자 B



(c) 화 자 C



(d) 화 자 D



(e) 화 자 E

3. 음원 여기신호를 이용한 피치조절

3.1 피치조절에 사용되는 규칙

피치조절을 위한 입력신호로써 음원 여기신호를 사용한다. 분석의 결과에 의하면 음원 여기신호의 닫힌 구간의 변화가 피치의 증가와 감소에 가장 큰 영향을 주고 있으므로 이 구간을 피치조절을 위한 대상으로 삼는다. 이전의 피치조절 방법과 가장 큰 차이점은 원래의 음성신호를 그대로 사용하지 않고 피치조절을 위한 입력신호로써 음원 여기신호를 사용했으며 특히 그 중에서

도 닫힌 구간만을 고려했으므로 피치조절을 위한 전체적인 알고리즘이 간단해졌다. 피치가 조절된 합성음을 얻기 위한 전체적인 구성도는 그림 5와 같다.

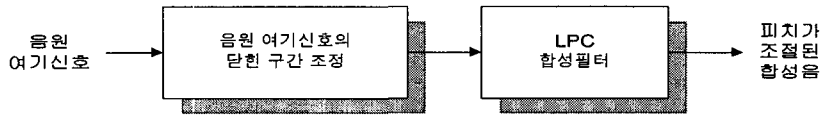


그림 5. 제안한 방법의 전체 블록도

3.2 음원 여기신호 조절방법

피치를 낮출 경우 즉 간격을 늘릴 경우 원래 파형을 그대로 주파수상에서 처리하여 다시 시간영역으로 바꾸어 음원 여기신호의 열린 구간과 연결할 경우 연결부분에서 파형이 너무 작은 값을 갖게 되므로 이러한 문제를 해결하기 위해 닫힌 구간의 파형을 미분하여 이를 주파수상에서 처리하고 시간영역으로 바꾼 후 다시 적분하면 이러한 문제를 해결할 수 있다. 닫힌 구간의 파형을 전달함수가 $H(z)=1-z^{-1}$ 인 필터에 통과시키면 미분파형을 갖게 되는데 이 파형은 원래 파형에 비해 그 값이 더 작게 되고 거의 0에 가깝게 된다. 이 필터를 주파수영역에서 해석하면 고역통과 필터로써 저주파성분이 큰 닫힌 구간의 신호를 통과시키면 주파수상에서 전체적으로 진폭이 일정해진다.

다음 단계에서는 이 미분한 파형을 주파수 영역으로 바꾸기 위해 DFFT를 해준다. 실험에서는 64점 DFFT를 해주었다. 분석에 사용되었던 음성의 닫힌 구간의 샘플 수는 약 30개 정도였는데 64개 보다 적을 때는 나머지 부분을 0으로 채워주었다. 128점 DFFT와 256점 DFFT를 해서 얻은 결과를 비교해 보면 주파수상에서 거의 차이를 나타내지 않았으므로 64점 DFFT를 해주었다.

Interpolation/Decimation 단계에서는 FD-PSOLA와 유사한 방법을 사용한다. 간격을 늘릴 경우, 즉 피치를 낮출 경우에는 주파수영역에서 interpolation을 해준다. 즉 n개를 늘린 경우 64+n점이 만들어진다. 간격을 줄일 경우는 decimation을 해준다. n개를 줄일 경우 64-n점이 만들어진다. 만일 $64\pm n$점이 홀수인 경우는 짝수로 만들기 위해 $64\pm n+1$점으로 한다. 64점 DFFT된 값에서 2점부터 32점까지와 34점부터 64점까지의 값은 서로 켤레 복소수의 관계이므로 1점부터 33점까지의 값을 가지고 interpolation/decimation을 한 후 34점부터 64점까지의 값은 대칭을 시켜서 만들어 준다. Interpolation이나 decimation을 통해서 만들어지는 복소계수는 아래의 식과 같다.

복소스펙트럼의 k번째 계수를 $S_1(k)$라고 하고 interpolation이나 decimation을 통해서 만들어진 계수 $S_2(l)$을 구하기 위해 아래의 식과 같은 과정을 거친다.

원하는 갯수가 $64\pm n$ 일 경우 전체 간격을 $64\pm n$ 개의 등간격으로 나눌 때 그 때의 비는 $R=\frac{63}{(64\pm n-1)}$ 이 되고 위치는 식 1과 같다.

$$m(1)=1$$
$$m(j)=1+(j-1)R \qquad j=2,3,4,\ldots,64\pm n \qquad (1)$$

$m(j)$는 항상 정수가 아니므로 $m(j)$의 소수점 이하를 버린 수를 $\widetilde{m}(j)$ 라고 하면 $S_2(l)$은 식 2와 같다.

$$S_2(l) = (1-\alpha)\ S_1(\widetilde{m}(j)) + \alpha\ S_1(\widetilde{m}(j)+1)\ ,$$
$$\alpha = m(j) - \widetilde{m}(j)\ , \qquad l = j = 1,2,3,\ldots,64 \pm n \qquad (2)$$

IDFFT를 하면 주파수상에서 조절된 값들을 시간영역으로 바꾸어 준다. 이렇게 시간영역에서 발생된 값에서 원하는 샘플 개수만큼만 취하면 간격이 조절된 닫힌 구간을 구할 수 있는데 미분 파형을 다시 적분하기 위해 전달함수가 $H(z) = \frac{1}{1-z^{-1}}$ 인 필터에 통과시킨다. 이 값을 다시 음원 여기신호의 열린 구간과 연결해 주면 피치가 조절된 음원 여기신호를 얻을 수 있다.

피치가 조절된 후의 음원 여기신호와 원래의 음원 여기신호를 시간상에서 비교해 보면 모양이 거의 비슷하고 열린 구간과도 자연스럽게 연결이 된다. 주파수상에서 비교해 보면 피치가 조절됨에 따라 발생되는 포만트들의 위치가 약간씩은 변화하지만 스펙트럼의 모양은 원래의 파형의 스펙트럼과 거의 비슷하며 각 프레임 별로 비교해 보아도 갑작스런 포만트들의 변화가 생기지 않고 포만트들의 연결이 부드럽다. 전체적인 알고리즘은 그림 6에 설명되어 있고 주파수영역에서의 조절과 피치가 조절된 음원 여기신호는 각각 그림 7과 그림 8에 나와 있다.

피치조절로 인해서 발생되는 주파수상에서의 왜곡을 극복할 수 있고 또한 이를 시간 영역에서 피치가 조절된 음원 여기신호를 살펴보아도 피치가 조절되기 전과 파형이 거의 비슷하다.

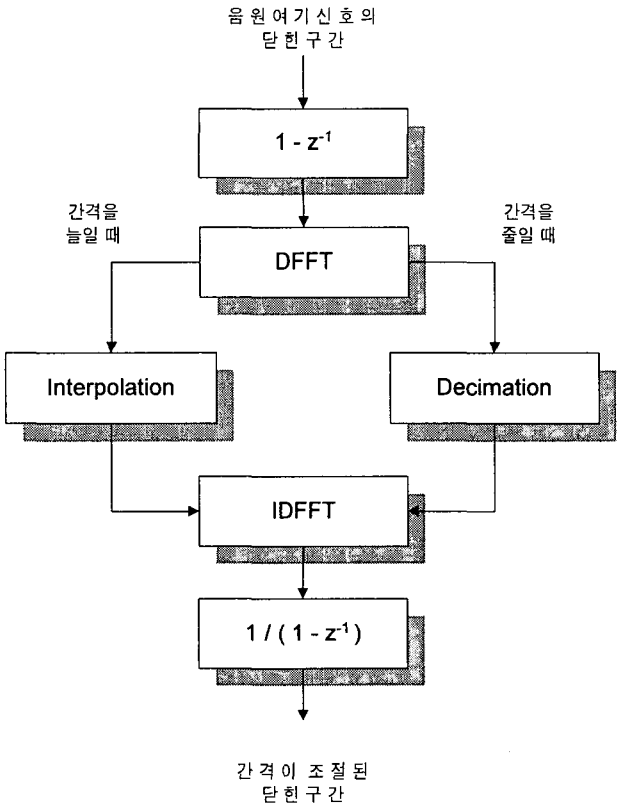
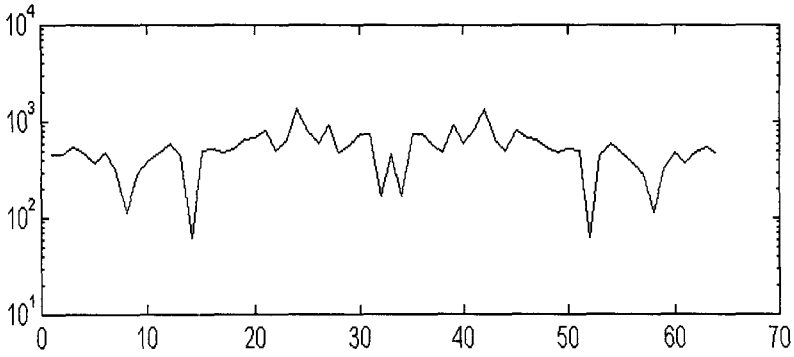
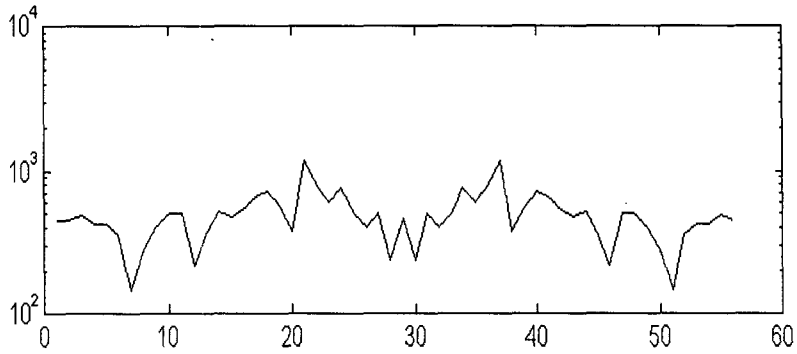


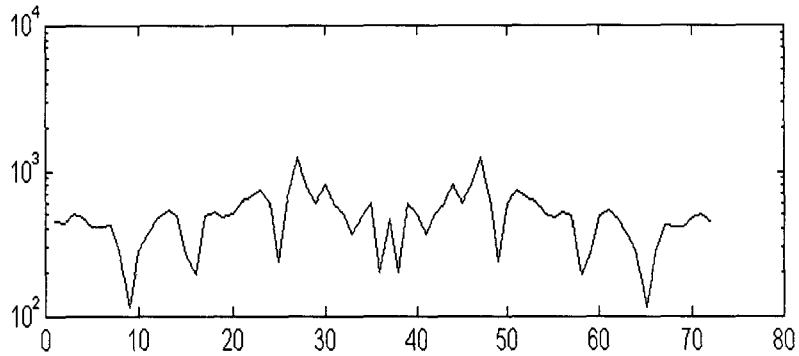
그림 6. 주파수영역에서 interpolation/decimation 알고리즘



(a) 닫힌 구간의 64점 DFFT

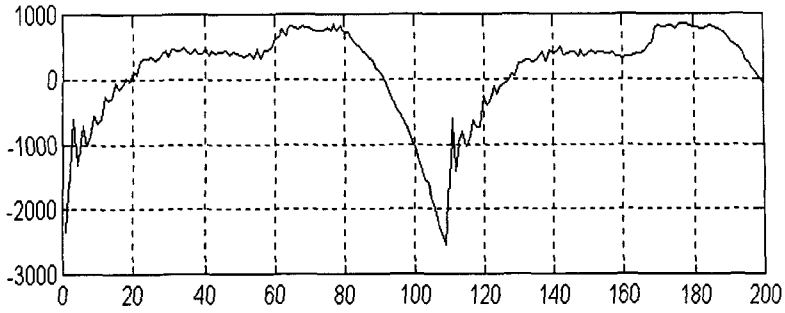


(b) Decimation (64 − 8 점 DFFT)

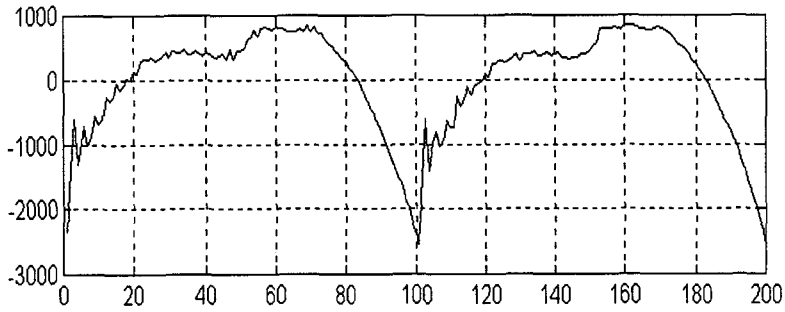


(c) Interpolation (64 + 8 점 DFFT)

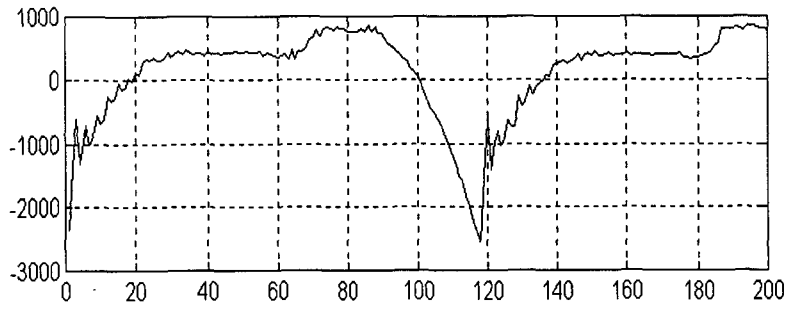
그림 7. 주파수상에서 decimation과 interpolation



(a) 음원여기신호



(b) 피치를 높일때 음원여기신호



(c) 피치를 낮출때 음원여기신호

그림 8. 피치가 조절된 음원 여기신호

닫힌 구간이 조절된 음원 여기신호는 LPC 합성필터에 통과시킴으로써 피치가 조절된 합성음을 얻을 수 있다. 그림 9에 피치가 조절된 합성음이 나와있다. 여기에 사용되는 LPC계수는 분석에서 구한 LPC계수를 이용한다. 음원 여기신호의 열린 구간에서 최대값과 최소값은 화자와 피치에 따라 각각 변화를 보였으므로 이를 음색제어에 이용하는 것도 가능할 것이다. 포만트주파수를 구하는데 사용되는 LPC 계수를 원하는 포만트주파수에 해당하는 LPC 계수로 바꾼 후 LPC 합성필터에 통과시킴으로 합성음의 음색 또한 제어할 수 있다. 이는 합성방법이 음원 여기신호를 고려했기 때문에 가능하다. 실험을 통해서 포만트주파수를 바꾸어 주거나 대역폭을 조절하므로 합

성음의 음색이 변함을 알 수 있었다. 기존의 방법들은 피치를 조절하는데 음성 파형만을 이용했기 때문에 합성음의 음색제어가 불가능하며 각 화자의 개인성요인을 나타내지 못하고 있다. 그러므로 고품질의 음성합성기를 구현하기 위해서는 음성 파형 자체를 조합하는 합성방식보다는 이와 같이 적은 양의 DB를 이용하여 여러 가지 음색을 제어할 수 있는 합성방식이 적합할 것이다.

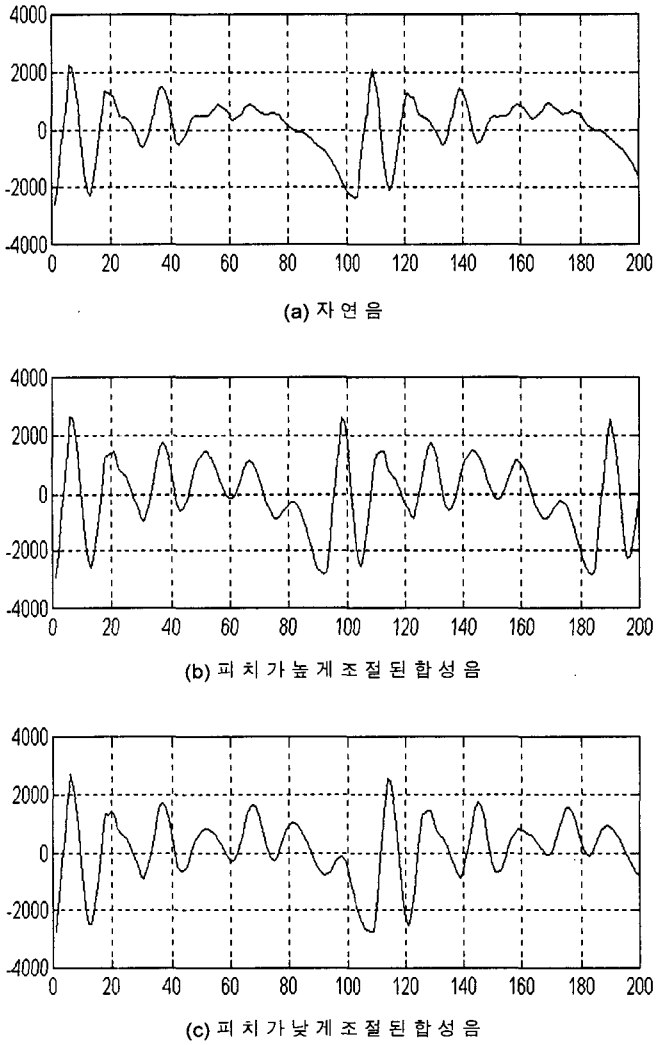


그림 9. 제안한 방법에 의해 피치가 조절된 합성음

4. 평 가

제안한 방법으로 피치가 조절된 합성음의 성능을 평가하기 위해 기존의 피치조절 방법들에 의해서 만들어진 합성음들과 비교한다. 각각의 합성음에 대해 청취테스트를 하고 파라메타의 값을 비교한다.

■ 비교대상이 되는 방법

현재 많이 사용되고 있는 PSOLA방식 중 가장 먼저 개발된 TD-PSOLA와 최근에 개발된 LP-PSOLA, Wavelet 방법, 본 논문에서 제안한 방법이다.

■ 비교대상이 되는 파라메타

스펙트럼과 포만트주파수와 포만트주파수의 대역폭과 포만트주파수에서의 진폭으로 하였다. 이는 포만트가 음성을 특성 지어주는 가장 중요한 파라메타 중의 하나이며 진폭은 주파수상에서의 성분의 증가와 감소를 측정할 수 있다.

합성음의 성능을 평가하기 위해 다음과 같은 방법으로 실험을 하였다.

● 실험1

동일화자가 같은 모음을 낮은 피치와 높은 피치로 각각 녹음하여 낮은 피치로 녹음된 음성을 높은 피치로 녹음된 음성의 피치와 같도록 각각의 방법에 의해 조절하였다. 실험방법은 그림 10에 나와 있고 높은 피치의 자연음과 각각의 방법에 의한 합성음의 스펙트럼 비교는 그림 11에 나와 있다.

그림 10과 같은 실험방법으로 목표가 되는 자연음과 각각의 방법들에 의해서 만들어진 합성음을 스펙트럼 상에서 비교해 본 결과 TD-PSOLA방법은 주파수의 모든 영역에서 크기가 감소되고 LP-PSOLA와 제안한 방법의 합성음은 자연음의 스펙트럼과 거의 비슷하게 나타난다.

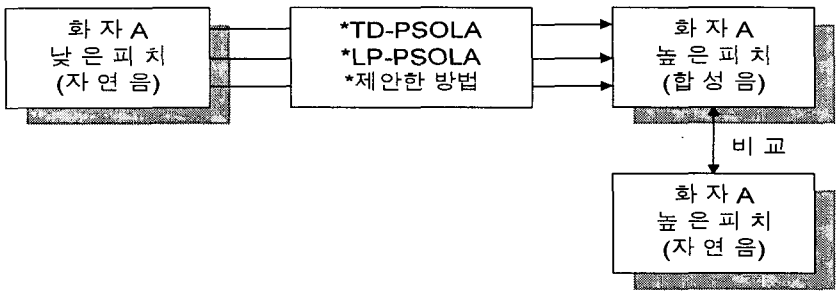


그림 10. 실험1의 블록도

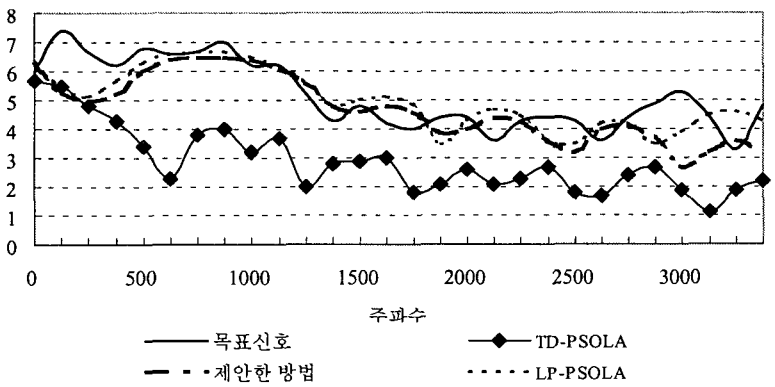


그림 11. 피치조절 방법들의 스펙트럼 비교

● 실험 2

신문기사에서 발췌한 각 10 단어 내외의 5개의 문장에 대해 각각의 방법으로 피치를 조절하여 19명(남자 15명, 여자 4명)의 청취자들을 대상으로 청취테스트를 실시하였다. 자연음의 피치에 비례하여 높거나 낮게 조절한 합성음과 동일한 피치로 조절한 합성음을 각각의 방법에 의해서 만들었다. 각각의 합성음을 MOS(mean opinion scores)테스트를 하여 성능에 따라 1에서 5까지의 점수(1:very low, 2:low, 3:fair, 4:high, 5:very high)를 주었는데 표 1은 이 점수들의 평균을 나타낸 것이다.

표 1. 합성음들의 MOS테스트 결과

	Wavelet	TD-PSOLA	LP-PSOLA	제안한 방법
피치를 높일때	4.0	3.7	3.9	4.3
피치를 낮출때	4.1	3.9	3.9	4.3
동일한 피치	3.7	3.6	3.7	4

청취 테스트를 한 결과, 모든 청취자들이 제안한 방법의 성능을 가장 높게 평가하였고 Wavelet 방법의 경우도 어느 정도 양질의 성능을 보였으나 동일한 피치로 조정한 경우에는 피치 변화폭이 클 경우보다 그 성능이 많이 떨어진다. TD-PSOLA와 LP-PSOLA의 합성음의 구별은 어렵지만 LP-PSOLA가 성능이 약간 우수하다고 할 수 있다.

5. 결 론

본 논문에서는 합성음의 피치조절 방법을 음원모델에 기초하여 제안하였다. 기존의 피치조절 방법들과 달리 음성 파형만을 고려하지 않고 피치의 높고 낮음의 원인을 음성의 발생과정에서 찾아서 음원 여기신호를 피치조절의 대상으로 삼았다. 주파수상에서의 왜곡을 최소화하기 위해 신호를 주파수영역에서 고려해 주었다. 피치조절을 위한 기존의 방법들에 비해서 포만트주파수의 왜곡이 줄어들고 스펙트럼 진폭이 감소되지 않았으며 청취테스트에서도 더 나은 성능을 보였다. 음원 여기신호 중 닫힌 구간만을 이용했으므로 조절방법이 간단하다. 또한 음원 여기신호에서 파라메타들이 각 화자의 개인성 요인을 나타내므로 파라메타들을 추출하여 음색과 파라메타들과의 상관관계에 대한 규칙을 만들어서 이를 음색제어에 이용할 수 있다. 고품질의 합성기를 구현하기 위해서는 피치조절 이외에 음성의 개인성 요인이나 음색을 고려해 주어야 하므로 제안한 방법은 이러한 요건들을 만족시킬 수 있을 것이다.

참 고 문 헌

[1] 한국전자통신연구소. 1993. *자동통역전화요소기술연구 93년 연구보고서*. ETRI.

[2] 한국통신. 1993. *자동통역전화요소기술연구 93년 연구보고서*. 한국통신 S/W연구소.

[3] 오영환. 1994. "*PARCOR 합성기를 이용한 한국어 Text-to-Speech 시스템의 구현*". 93 통신학술 연구과제, 1-67.

[4] E. Moulines, F. Charpentier. 1990. "Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphone", *Speech Communication*, 9, 453-467.

[5] T. Dutoit, H. Leich. 1992. "Improving the TD-PSOLA text-to-speech synthesizer with a specially designed MBE resynthesis of the segments database", *EUSIPCO 92*, 343-347.
[6] Donald G. Childers, Chun-Fan Wong. 1994. "Measuring and modeling vocal source-tract interaction" *IEEE*, 41(7), 663-671.
[7] Dale E. Veeneman, Spencer L. Bement1985. "Automatic glottal inverse filtering from speech and electroglottographic signals", *IEEE*, ASSP-33(2), 369-376.
[8] Paavo Alku, Erkki Vilkman. 1995. "Effects of bandwidth on glottal airflow waveforms estimated by inverse filtering", *J. Acoust. Soc. Am*, 98(2), 763-767.
[9] C. Hamon, E. Moulines, F. Charpentier. 1989. "A diphone synthesis system based on time-domain prosodic modifications of speech", *Proc. Int. Conf. Acoust., Speech, Signal Proc.*, Glasgow, 238-241.
[10] F. M. Gimenez de los Galanes, M. H. Savoji, J. M. Pardo. 1994. "New algorithm for spectral smoothing and envelope modification for LP-PSOLA synthesis" *IEEE, Proc.* I-573 - I-576.
[11] F. J. Charpentier, M. G. Stella. 1986. "Diphone synthesis using an overlap-add technique for speech waveforms concatenation", *IEEE ICASSP*, 38.5.1-38.5.4.
[12] D. G. Childersand, Ke Wu, D. M. hicks. 1989. "Voice conversion", *Speech Communication*, 8, 147-158.
[13] P. H. Milenkovic. 1993. "Voice source medel for continuous control of pitch period", *J. Acoust. Soc. Am*, 93(2), 1087-1096.
[14] L. R. Rabiner and R. W. Schafer.. 1978. *Digital Precessing Speech Signals*, Englewood Cliffs, NJ: Prentice-Hall.
[15] A. M. Kondoz. 1994. *Digital speech(coding for low bit rate communications systems)*, John Wiley & Sons.
[16] D. G. Childers and C. K. Lee. 1991. "Vocal quality factors: analysis, synthesis, and perception", *J. Acoust. Soc. Am*, 90(5), 2394-2410.

접수일자 : '98. 2. 20.
게재결정 : '98. 3. 11.

▲ 최용진
서울특별시 서초구 우면동 16
LG 종합기술원 정보기술연구소 CT Gr. (우 : 137-140)
Tel : (02) 526-4575 (O), (0342) 705-7647 (H)
Fax : (02) 526-4852
e-mail : yongjin@lgcit.com

▲ 김진영
광주광역시 북구 용봉동 300
전남대학교 전자공학과(우 : 500-757)
Tel : (062) 530-1757 (O), (062) 572-0508 (H)
Fax : (02) 530-0472
e-mail : kimjin@dsp.chonnam.ac.kr

▲ 여수진
광주광역시 북구 용봉동 300
전남대학교 전자공학과(우 : 500-757)
Tel : (062) 530-0472 (O) Fax : (02) 530-0472
e-mail : sjyeo@dsp.chonnam.ac.kr

▲ 성굉모
서울특별시 관악구 신림동 산 56-1
서울대학교 전기공학부(우 : 151-742)
Tel : (02) 880-8403 (O) Fax : (02) 880-8207
e-mail : kmsung@acoustics.snu.ac.kr