한국음향학회지 제23권 제1호 pp. 82~89 (2004)

음성신호의 실시간 피치변경에 관한 연구

A Study on Real Time Pitch Alteration of Speech Signal

김 종 국*, 박 형 빈*, 배 명 진*
(Jong Kuk Kim*, Hyung Bin Park*, Myung Jin Bae*)

*숭실대학교 정보통신공학과

(접수일자: 2003년 11월 11일; 채택일자: 2003년 12월 30일)

고음질 합성을 하면서도 다양한 음색을 갖도록 하기 위해서는 파형부호화를 이용한 합성법에 적용할 수 있는 피치 변경법이 필요하다. 따라서 본 논문에서는 스펙트럼 왜곡률을 최소화하는 영교차 단위의 시간축 조절에 의한 피치 변경법과 피치 동기분석이 용이하고 다른 영역으로의 변환 과정이 불필요한 피치시점 검출법을 제안함으로써 고음질을 유지하면서 시간영역에서만 처리됨으로써 계산량을 줄이고 스펙트럼 왜곡률을 최소화하고 위상을 그대로 보존할 수 있는 시간영역에서의 피치 변경법을 제안하였다. 결과적으로 전체 피치 변경율에 대해서는 기존의 방법에 비해서 제안한 방법의 스펙트럼 왜곡률이 0.73% 개선되었고 피치 압축시에는 제안한 방법의 스펙트럼 왜곡율이 2.18% 개선되었다.

핵심용어: 스펙트럼 왜곡, 피치변경, PSOLA, 피치시점
투고분야: 음성처리 분야 (2.4)

This paper describes how to reduce the effect of an occupation threshold by that the transform of mixture components of HMM parameters is controlled in hierarchical tree structure to prevent from over-adaptation. To reduce correlations between data elements and to remove elements with less variance, we employ PCA (principal component analysis) and ICA (independent component analysis) that would give as good a representation as possible, and decline the effect of over-adaptation. When we set lower occupation threshold and increase the number of transformation function, ordinary MLLR adaptation algorithm represents lower recognition rate than SI models, whereas the proposed MLLR adaptation algorithm represents the improvement of over 2% for the word recognition rate as compared to performance of SI models.

Keywords: *Spectrum distortion, Pitch alteration, PSOLA, Pitch point*
ASK subject classification: *Speech signal processing (2.4)*

I. 서 론

음성합성은 합성 방식에 따라 파형부호화, 신호원부호화, 혼성부호화를 이용한 합성으로 분류된다. 특히 고음질합성을 위해서는 파형부호화를 이용한 합성방식이 적합하다. 그렇지만 파형 부호화를 이용한 합성법은 여기성분과 여파기 성분을 분리하지 않고 처리하기 때문에 음원을 변경하기가 상당히 어렵다. 따라서 고음질 합성을 하면서도 다양한 음색을 갖도록하기 위해서는 파형 부호화를 이용한 합성법에 적용할 수 있는 피치 변경법이 필요하다. 피치 변경율에 따라 스펙트럼 왜곡율이 크게 증가하지 않게 하려면 포만트 주파수와 기본 주파수를 정수배로 유지시키는 것이 필요하다. 따라서 피치주기를 압축하는 경우에는 F1과 F0의 율을 정수배로 유지시켜서 스펙트럼 왜곡을 최소화하는 영교차 단위의 시간축 조절에 의한 피치 변경법을 적용하였다. 또한 피치주기를 신장하는 경우에는 기존의 PSOLA 합성방식에 의한 피치 변경법을 적용하였다[4]. 파형부호화법에서 화자의 개성과 명료성을 유지하려면 발성자의 중심이 되는 피치를 기준으로 하여 피치 변경이 이루어져야 한다. 따라서 피치 변경을 수행하기 위해서는 그 발성자의 정확한 피치시점을 검출할 수 있어야 한다. 더불어 피치 시점을 정확히 검출함으로써 음성분석 시 피치 동기된 분석을 할 수 있고, 음성합성 시에는 여기원의 위상특성을 파악할 수 있으므로 개성이 강조된 합성음을 얻을 수 있다[2]. 따라

책임저자: 김종국 (kokjk@hanmail.net)
156-743 서울시 동작구 상도5동 1-1
숭실대학교 정보통신공학과
(전화: 017-237-7545 ; 팩스: 02-825-0018)

서 본 논문에서는 스펙트럼 왜곡률을 최소화하는 영교차 단위의 시간축 조절에 의한 피치변경법과 피치시점 검출법을 제안하였다.

본 논문은 모두 6장으로 구성되어 있으며 제2장에서는 피치 변경 시스템의 처리 과정과 기존에 제안되었던 피치 변경법들의 장·단점 그리고 피치 변경시에 고려하여야 할 사항에 대하여 살펴본다. 제3장에서는 스펙트럼 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있는 피치 변경율에 따른 최적의 피치 변경법을 제안한다. 제4장에서는 피치를 변경하기 전에 수행되어야 할 피치 검출에 대하여 살펴보고 기존의 방법과는 달리 분석구간별로 얻어지는 성문의 진폭특성과 주기특성을 적용하여 제안한 피치시점 검출법에 대해서 설명한다. 제5장에서는 제안한 피치 변경법의 성능을 측정하기 위한 실험 과정과 그 결과를 분석하였고 마지막으로 6장에서 종합적인 평가 및 검토로 결론을 맺는다.

II. 피치 변경법

2.1. 피치 변경 시스템

피치 변경 시스템은 그림 1과 같이 구성된다. 피치 변경 시스템의 분석단에서는 마이크로폰으로 입력된 원 신호와 목적 신호의 피치를 검출하여 변경 규칙 생성단에 넘겨준다. 변경 규칙 생성단에서는 이를 이용하여 피치 변경율과 그에 적합한 피치 변경법을 결정한다[1]. 이러한 피치 변경 규칙은 실제 피치 변경단에 제공되어 원 신호의 피치를 선정된 피치 변경법을 적용하여 변경율만큼 피치를 변경하고 합성단에서는 이를 이용하여 음성이 변경된 합성음을 생성한다. 이러한 과정에는 정확한 피치 검출 기법과 함께 왜곡이 적은 피치 변경 기법을 필요로 한다.

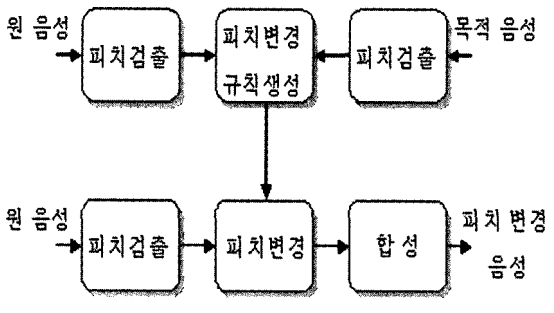

그림 1. 피치 변경 시스템의 블록도
Fig. 1. Diagram of pitch alteration system.

2.2. 기존의 피치 변경법

2.2.1. 시간 영역법

지금까지 제안된 시간 영역 피치 변경법으로는 Multi-Pulse법, LPC 신장법, 피치반분법 등이 있다. Caspers와 Atal은 MPLPC에서 멀티 펄스 사이에 영을 삽입하거나 삭제함으로써 피치를 변경하는 방법을 제안하였으나[10], MPLPC상의 멀티 펄스의 위치는 원 신호와 합성 신호와의 오차가 최소가 되도록 최적의 위치에 선정되므로 펄스의 위치를 바꾸는 것은 합성음의 스펙트럼 왜곡을 초래한다.

Varga와 Fallside는 LPC 계수를 이용한 피치연장법을 제안하였다[11]. 그러나, 이 방법 역시 피치주기를 줄이는 경우 파형의 일부분을 소거하고 평활화하는 방법을 사용하고 있기 때문에 스펙트럼의 왜곡이 심하다. 피치 반분법은 변경하려고 하는 목적 피치의 2배 피치를 갖는 파형을 LPC 신장법에 의해 생성한 후 데시메이션에 의해 주기를 반분하는 피치 변경법이다[12].

2.2.2. 주파수 영역법

Quatieri와 McAulay는 주파수 영역 피치 변경법으로 음성 신호의 진폭 스펙트럼과 위상 스펙트럼을 분리하여 별도로 처리하는 방법을 제안하였다[13]. 진폭 스펙트럼에 대해서는 두드러진 스펙트럼 고조파들을 추출하여 이것을 피치 변경율 (ρ)만큼 인터폴레이션하여 진폭 스펙트럼의 피치를 변경시킨다. 위상 스펙트럼에 대해서는 시간 영역에서 구한 피치 개시시간 (pitch onset time)에 해당하는 위상을 제거하고 피치가 변경되었을 때의 새로운 피치 개시시간의 위상을 더해줌으로써 새로운 위상을 구성하게 된다. 이 방법은 피치 변경시에 피치 주기와는 별도로 피치 개시시간을 공급해 주어야 하고, 또한 진폭 스펙트럼상에서 두드러진 고조파 위주로 인터폴레이션을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

다른 주파수 영역 피치 변경법으로는 평탄화 기법에 의해 포만트와 기본 주파수의 고조파를 분리하여 기본 주파수를 선형적으로 스케일링함으로써 피치를 변경하는 방법이 있다[14]. 이 방법은 스펙트럼을 원래의 음성스펙트럼으로 대신하는 방법이나 스펙트럼 상에서 고조파를 스케일링시킴으로서 창함수의 특성도 변경되어 시간 영역에서 위상을 복원하기가 어렵게 된다.

2.2.3. 혼성 영역법

시간-주파수 혼성법으로는 켑스트럼의 특징을 이용하

여 켑스트럼값이 거의 영이 되는 부분에서 영값을 삽입하거나 삭제함으로써 피치를 변경하는 방법이 있다[15]. 그러나 이 방법 역시 위상의 보존이 어렵다는 문제점을 가지고 있다. Takagi와 Miyasaka가 제안한 시간-주파수 혼성법은 시간영역에서 피치 변경을 하였을 때에 나타나는 스펙트럼왜곡을 스펙트럼 영역상에서 LPC 포락을 통해 수정하는 방법이다. 이 방법은 LPC 스펙트럼 포락이 갖는 극점에 치중된 시스템 전달 특성 때문에 모든 유성음을 만족하지 못한다는 한계성을 갖는다.

2.2.4. 피치 변경시의 고려사항

음성 발성 모델에 따라 음성 신호를 분석해 보면 인간의 개성과 감정을 나타내는 여기 (excitation)정보와 의사 내용을 나타내는 성도 여파기의 포만트 정보로 구성되어 있음을 알 수 있다. 이는 주파수 영역 상에서 여기 스펙트럼과 포만트 스펙트럼으로 나타나게 된다. 피치 변경시에 포만트 스펙트럼이 왜곡되면 성도의 여파기 정보가 왜곡되므로 의사 내용을 제대로 보존할 수 없게 된다. 또한, 위상이 왜곡되면 인근 프레임간의 진폭 레벨의 변동이 커져서 음소간의 연결이 부자연스럽게 된다. 따라서 피치를 변경할 때에는 위상을 보존하면서도 스펙트럼 왜곡을 최소화 할 수 있는 피치 변경 방법이 필요하다[3,9].

III. 피치변경율에 따른 최적의 피치 변경법

3.1. 영교차 단위의 시간축 조절에 의한 피치 변경법

유성음의 한 피치주기 동안의 파형 모양은 성문펄스 파형과 제1포만트 파형이 서로 컨벌루션 되어 나타나기 때문에 첫 파형의 봉우리는 성문파의 모양과 유사하다. 그러므로 이 파형은 인근 봉우리 파형에 비해 영교차 간격이 길고 봉우리의 진폭도 높게 나타난다. 따라서 이 파형 봉우리를 성문의 영향이 지배적인 봉우리라고 생각하여 G(glottal)-peak라 정의한다. 또한 제1포만트의 에너지가 다른 포만트들에 비해 두드러지고 대역폭을 가지기 때문에 한 피치 주기내에서의 파형은 거의 제1포만트의 주기로 감쇄진동을 하게 된다[5,7].

음성스펙트럼은 기본주파수의 고조파로 이루어진 여기스펙트럼과 포만트 공명봉우리가 곱해진 형태를 띠게 된다. 여기스펙트럼에서 고조파 봉우리들의 대역폭은 포만트 봉우리에 비해 아주 좁은 편이다. 따라서 피치 변경시의 여기스펙트럼이 고조파 간격을 변경시키면 음성스펙트럼에서 제1포만트의 위치 F_1은 기본주파수 F_0의 거의 정배수를 나타내게 되고, 고조파의 위치에 따라 추정되는 포만트의 위치는 다음 범위 내에서 변화를 관찰할 수 있다.

$$F_1' = F_1 \pm \frac{1}{2} F_0 \tag{1}$$

이것을 시간영역에서 설명하면 한 피치주기 내에서 제1포만트 파형은 피치주기에 거의 정배수가 되고 피치를 변경하는 경우에도 피치주기와 제1포만트의 주기는 정배수로 일치시키는 것이 필요하다. 시간영역에서 음성 파형에 대해 쉽게 처리할 수 있는 것은 영교차 단위로 피치주기를 일치시키는 것이다. 유성음 파형에서 영교차 간격은 제1포만트의 파형 주기의 1/2 정도를 차지하게 된다. 먼저 피치주기를 압축시키기 위해 음성 파형의 일부분을 삭제시키면 한 피치주기 파형의 모서리에서 스펙트럼 누설 현상이 발생되기 때문에 파형의 명료성이 저하된다. 이것은 피치주기와 파형에 포함된 제1포만트의 주기가 불일치 되어 나타나는 현상이다. 이러한 경우에 파형 모서리의 스펙트럼누설을 제거하려면 피치 변경된 파형에 대해 시간축 조절을 수행하여 파형의 영교차 점을 변경된 피치주기에 일치시키면 된다. 이렇게 하면 시간축 조절에 의해 포만트의 주파수가 일부 변경되지만, 피치주기와 포만트의 주기가 거의 정약수배로 일치하게 된다. 다음의 그림 2는 영교차 단위의 시간축 조절에 의해서 피치주기를 압축한 경우의 예를 나타낸 것이다.

3.2. PSOLA 합성방식

3.2.1. 분석과정

원래 음성 파형이 유성음인 경우에는 피치단위로 분해한 다음 윈도우 함수를 곱하여 ST신호의 열로 만든다. 무성음인 경우에는 10 ms의 주기로 일정하게 분석한다. 분석 윈도우 함수에는 다음과 같은 Hanning, Hamming, Blackman 등의 형이 쓰인다[4]. 이런 윈도우 함수를 원래의 음성 샘플에 곱함으로써 다음 식 (2)과 같은 피치 단위로 분해된 샘플열들을 얻는다.

$$S_{analysis}(n) = W_{analysis}(m-n)S(n) \tag{2}$$

$S_{analysis}(n)$: 피치주기 단위의 ST 신호
$W_{analysis}(n)$: 분석 윈도우 함수
m : m번째 피치
$S(n)$: 원 음성 파형

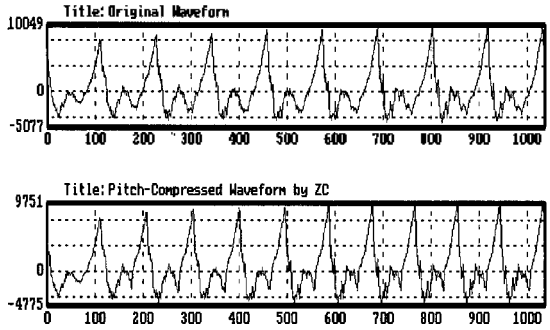

그림 2. 영교차 단위의 시간축 조절에 의한 피치 변경법
Fig. 2. Pitch alteration method by time scaling of zero-crossing unit.

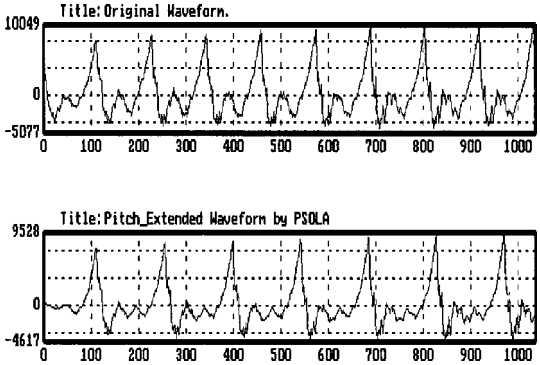

그림 3. PSOLA 합성방식에 의한 피치 변경법
Fig. 3. Pitch alteration method by PSOLA.

3.2.2. 운율 조절 및 합성과정

분석 과정에서의 ST신호의 열은 원래의 음성 샘플의 피치단위로 배열되어있다. 따라서 피치를 변경하기 위해서는 이 간격들을 변경할 피치 간격들로 재배열하면 된다. 다음 식 (3)은 피치가 변경된 신호를 나타낸 것이다.

$$S_{synthesis}(n) = S_{analysis}(n - m_a) \tag{3}$$

$S_{synthesis}(n)$: 피치가 변경된 ST신호

m_a : 변경할 피치 간격

따라서 피치를 높일 때는 ST 신호의 간격을 작게 배열하고, 피치를 낮출 때는 ST신호의 간격을 크게 배열하면 된다. 하지만 이런 순차적인 배열사이에서 정확한 피치 동기화를 유지하는 것이 중요하다. 이렇게 재배열된 ST 신호에서 겹쳐지는 부분을 더해주면 된다. 다음 그림 3은 PSOLA합성방식을 적용해서 피치주기를 신장한 경우의 예를 나타낸 것이다.

IV. 피치시점 검출법

4.1. 제안한 피치시점 검출법

운율정보 변환 시 사전에 피치검출과정이 수행되어져야만 한다. 하지만 분석프레임간 평균 피치정보는 음성신호에서 음소변화 특성 등을 잘 표현하기 어렵다. 따라서 정확한 피치시점을 검출할 수 있다면 피치동기 분석할 수 있고 운율정보의 변환이 용이하다. 음성생성모델의 관점에서 음성신호는 앞서 언급한 바와 같이 여기신호가 성도특성을 나타내는 필터를 통과함으로써 발생되는 신호로 볼 수 있다[7].

다음 식 (4)와 같이 선형예측계수로 표현되는 필터에 역으로 통과시킴으로써 여기신호 특성을 잘 나타내는 잔여신호 (Residual signal)를 얻을 수 있다. 결과적으로 음성의 여기원으로 볼 수 있는 주기적인 펄스열들을 얻을 수 있다.

$$e(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{5} \alpha_k s(n-k) \tag{4}$$

본 논문에서는 단구간 (short-term) 분석 잔여 신호열을 가지고 피치 동기된 분석을 통해서 피치시점 검출법을 제안하였다. 제안한 방법은 다음과 같이 크게 분석 과정, 예측과정, 피치시점 검출과정으로 나누어진다.

4.1.1. 분석 과정 (Analysis)

$$r_p(n) = s(n) - \tilde{s}(n) = s(n) - \sum_{k=1}^{order} \alpha_{prem\,k}\, s(n-k) \tag{5}$$

$\alpha_{prem\,k}$는 프리엠퍼시스 필터를 통한 고주파수 영역이 강조된 선형예측계수이다. $r_p(n)$는 $\alpha_{prem\,k}$계수로 표현되는 필터에 통과하여 얻어진 예측잔여신호이다.

$$r_p^N(n) = [\, r_p^N(0),\ r_p^N(1),\ r_p^N(2), \ldots, r_p^N(M)] \tag{6}$$

여기서 $0 \le M \le 45[ms]$, N은 전체 입력음성신호의 분석프레임 개수를 나타내고, M은 한 분석프레임구간에서의 예측잔여신호열의 개수를 나타낸다. 의 분석프레임 구간을 15 ms단위로 세 서브프레임구간으로 나눈다. 각 서브프레임구간에서 음(-)의 피크 값만을 가지고 예비 피치시점열을 구한다.

$$Mag_{\min}[r_p^{N_a}(n)] = [m_{r_p^{N_a}}(0), m_{r_p^{N_a}}(1), \ldots, m_{r_p^{N_a}}(M_a)]$$
$$Mag_{\min}[r_p^{N_b}(n)] = [m_{r_p^{N_b}}(M_{a+1}), \ldots, m_{r_p^{N_b}}(M_b)]$$
$$Mag_{\min}[r_p^{N_c}(n)] = [m_{r_p^{N_c}}(M_{b+1}), \ldots, m_{r_p^{N_c}}(M_c)] \quad (7)$$

각 서브프레임구간에서 예비 피치시점열들의 음(-)의 피크 값을 비교해서 올림차순으로 6개의 예비 피치시점열을 구한다.

구해진 최소값들의 피치시점위치열 $pitchP_{M_{\min}^{N_a}}(0)$, $pitchP_{M_{\min}^{N_b}}(0)$, $pitchP_{M_{\min}^{N_c}}(0)$을 기준 피치시점열로 정한다.

4.1.2. 예측 과정 (Prediction)

각 서브프레임에서 결정된 기준 피치시점 위치열 $pitchP_{M_{\min}^{N_a}}(0)$, $pitchP_{M_{\min}^{N_b}}(0)$, $pitchP_{M_{\min}^{N_c}}(0)$을 가지고 나머지 5개의 예측된 피치시점 위치열과의 간격을 조사하여 최소피치구간 $\Delta pitch_{\min.range}$ (2.5 ms)이상 간격을 가지는 것은 해당프레임구간에서의 피치시점 위치열로 결정한다. 다음은 첫 번째 서브프레임구간에서의 과정을 나타낸 것이다.

$$| pitchP_{Mag_{\min}[r_p^{N_a}(n)]}(i-1) - pitchP_{M_{\min}^{N_a}}(0) |$$
$$> \Delta pitch_{\min.range}$$
$$| pitchP_{Mag_{\min}[r_p^{N_a}(n)]}(i+1) - pitchP_{M_{\min}^{N_a}}(0) |$$
$$> \Delta pitch_{\min.range} \quad (8)$$

즉, $pitchP_{Mag_{\min}[r_p^{N_a}(n)]}(i-1)$와 $pitchP_{Mag_{\min}[r_p^{N_a}(n)]}(i+1)$는 피치시점 위치열로 정한다. 그런 다음 순차적으로 인근한 예측된 피치시점 위치열과 위와 같은 조건이 성립되는지 계속 비교한다. 하지만 위 조건을 만족하지 않다면 그 다음 인근한 예측된 피치시점 위치열과 조건이 성립되는지 비교를 한다. 위와 같은 과정을 다 수행한 다음 조건을 성립하는 경우가 없을 경우에는 이미 결정된 기준 피치시점 위치열 $pitchP_{M_{\min}^{N_a}}(0)$, $pitchP_{M_{\min}^{N_b}}(0)$, $pitchP_{M_{\min}^{N_c}}(0)$만을 피치시점 위치열로 결정한다.

$$pitchP_{Mag_{\min}[r_p^{N}(n)]}(i) =$$
$$[pitchP_{Mag_{\min}[r_p^{N}(n)]}(0), \ldots, pitchP_{Mag_{\min}[r_p^{N}(n)]}(k)] \quad (9)$$

여기서 $2 \le k \le 17$, k는 현재 분석 프레임구간에서 구해잔 피치시점 위치열의 개수를 나타낸다.

4.1.3. 피치시점 검출 과정 (Detection)

제안한 피치시점검출 방법은 피치 동기된 분석을 하기 때문에 원 음성신호에서 해당 분석프레임구간에서의 영교차 정보 $ZCI^{N}_{negative}(j)$와 $ZCI^{N}_{positive}(j+1)$를 적용해서 정확한 피치시점 위치열을 검출한다. 다음 조건을 만족하는 피치시점 위치열을 구한 다음 최종 피치시점과 피치를 결정한다.

$$ZCI^{N}_{negative}(j) < pitchP_{Mag_{\min}[r_p^{N}(n)]}(i) < ZCI^{N}_{positive}(j+1)$$
$$pitchP(pIndex) = ZCI^{N}_{positive}(j+1)$$
$$Pitch = | pitchP(pIndex) - pitchP(pIndex+1) | \quad (10)$$

제안한 방법은 시간영역에서 직접 처리하기 때문에 피치동기분석이 용이하고 다른 영역으로의 변환과정이 불필요하다[6]. 또한 기존의 피치시점검출 방법에서는 결정논리를 실험적인 문턱값이나 무게치를 적용하여 처리하는 반면에 제안한 방법은 분석구간별로 얻어지는 주기적인 성문특성을 적용하여서 정확한 피치시점을 검출할

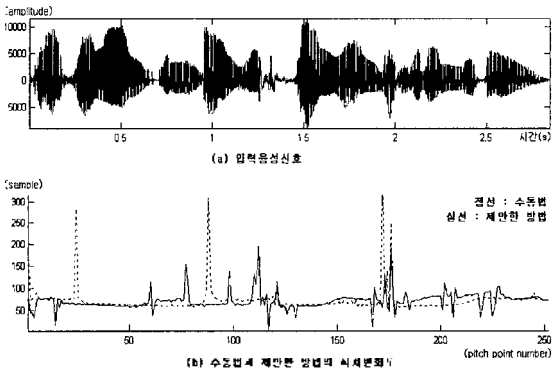

그림 4. 제안한 방법을 이용한 피치변화도 (남성화자)
Fig. 4. Pitch contour using proposed method (male speaker).

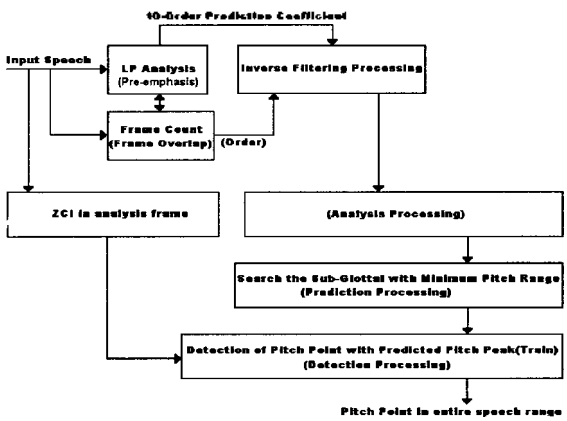

그림 5. 제안한 피치시점 검출방법의 블록도
Fig. 5. Diagram of pitch point detection using proposed method.

수 있었다. 그림 4는 제안한 방법을 적용한 경우 피치시점을 검출한 후 피치 변화도를 나타낸 것이며 그림 5는 제안한 방법의 블록 다이어그램을 나타낸 것이다.

V. 실험 및 결과

5.1. 음성 시료 구성 및 처리 과정

본 논문에서 제안한 방법을 실험하기 위하여 이용한 장비는 IBM-PC (Pentium-III) 시스템이며 여기에 음성신호를 입출력하기 위한 상용화된 16비트 AD/DA변환기를 인터페이스하여 8 kHz의 표본율로 데이터를 입력하였다. 각 시료에 대해 한 프레임의 길이를 45 ms으로 하고 예측차수(p)만큼 프레임 오버랩 과정을 수행하였다. 처리결과의 성능을 평가하기 위해서 대표적인 문장들을 연령층이 다양한 남녀화자가 발성하여 시료로 사용하였다. 제안한 방법을 구현하기 위해서 C-언어로 구현하여 수행하였으며 시뮬레이션에는 제안한 피치시점 검출방법을 사용하여 한 피치구간의 음성표본을 피치단위로 저장한 다음에 각 피치구간마다 음성표본을 저장한다. 그런 다음 피치 변경율에 따라 피치주기 압축 시에는 영교차 단위의 시간축 조절에 의한 피치 변경법을 적용하였다. 또한 피치주기 신장 시에는 PSOLA 합성방식에 의한 피치 변경법을 적용하였다[6,8]. 그림 6은 본 논문에서 제안한 피치 변경법의 블록도를 나타낸 것이다.

다음 그림 7과 8은 제안한 피치시점 검출법을 적용해서 스펙트럼 왜곡을 최소화하는 변경율에 따른 피치 변경을 수행한 결과를 시간 영역상에서 나타낸 것이다.

5.2. 결과 평가

5.2.1. 객관적인 평가

본 논문에서는 제안한 피치 변경 방법에 대한 성능평가를 위해서 피치주기를 60%에서 80%까지 변경시켰을 경우와 120%에서 200%까지 변경시켰을 경우에 대해서 스펙트럼 왜곡율을 측정하여 백분율로 환산하여 표 1에 제시하였다. 스펙트럼의 기준은 피치 변경되기 이전의 원래 음성의 스펙트럼이다. 피치를 변경시키면 원래의 스펙트럼과 비교할 수 없기 때문에 피치주기를 압축 및 신장시킨 다음 원래의 음성 스펙트럼과 고조파를 일치시켜서 왜곡율을 측정하였다.

본 논문에서 제안한 변경율에 따른 피치 변경법은 피치 압축시에는 기존의 PSOLA 방법과 제안한 방법인 영교차 단위의 시간축 조절을 비교하였다. 더불어 피치 신장시에는 기존의 PSOLA 합성방식을 그대로 적용하였다. 하지만 기존의 PSOLA 합성방식은 비대칭적인 음성파형에

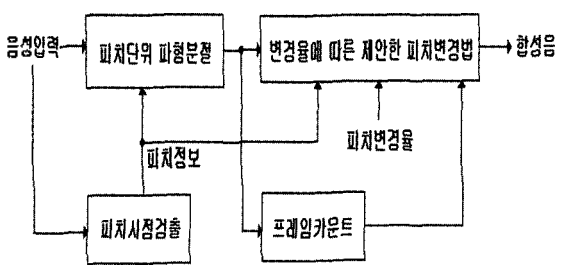

그림 6. 제안한 피치 변경법의 블록도
Fig. 6. Diagram of pitch alteration method using proposed.

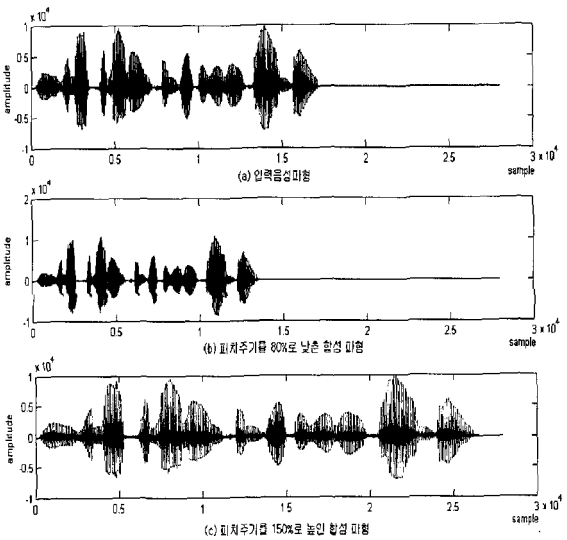

그림 7. 예수님께서 천지창조의 교훈을 말씀하셨다 (남성화자)
Fig. 7. Yesunimkeoseo cheonjichangjoeu gyohuneul malseumhasyuda (male speaker).

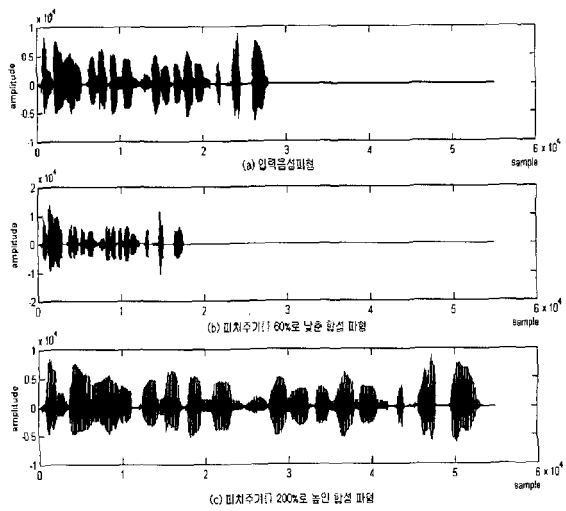

그림 8. 창공을 헤쳐 나가는 인간의 도전은 끝이 없다 (여성화자)
Fig. 8. Changgongeul hechye naganeun inganeu dojeoneun keuchieobda (female speaker).

표 1. 피치 변경율에 따른 스펙트럼 왜곡율 비교
Table 1. Compare spectrum distortion ratio to Pitch alteration ratio.

변경법 / 변경율	기존의 방법			제안한 방법		
	남성	여성	평균	남성	여성	평균
60%	8.51	11.69	10.1	6.37	8.91	7.64
80%	3.44	7.58	5.51	2.22	5.00	3.61
평 균	5.98	9.64	7.81	4.30	6.96	5.63
120%	0.98	2.06	1.52	0.98	2.06	1.52
150%	2.86	3.20	3.03	2.86	3.20	3.03
180%	4.66	5.62	5.14	4.66	5.62	5.14
200%	7.63	8.69	8.16	7.63	8.69	8.16
전 체 평 균	4.68	6.47	5.58	4.12	5.58	4.85

표 2. 발성시료에 대한 평균 MOS 비교
Table 2. Compare average MOS to utterence speech.

발성시료	5 Level MOS	
	수동 피치시점 검출법	제안한 피치시점 검출법
발 성 1 (남성)	4.05	3.80
발 성 2 (남성)	4.12	3.87
발 성 3 (여성)	3.86	3.12
발 성 4 (여성)	3.98	3.36
발 성 5 (남성)	4.20	3.90
평 균	4.04	3.61

대칭적인 윈도우 함수를 적용함으로써 야기되는 에너지 불균형 현상 때문에 피치 신장시 보다 압축시 왜곡을 발생시키는 단점을 가지고 있다. 표 1에 나와있는 결과와 같이 전체 피치변경율에 대해서는 기존의 방법에 비해서 제안한 방법의 스펙트럼 왜곡율이 0.73% 개선되었고, 피치 압축시에는 제안한 방법의 스펙트럼 왜곡율이 2.18% 개선되었다. 결과적으로 제안한 피치 변경율에 따른 피치 변경법은 상대적으로 스펙트럼 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있었다.

5.2.2. 주관적인 평가

본 논문에서는 수동적인 피치시점 검출과 제안한 피치시점 검출법에 의해서 화자간 피치 변경을 수행하였다. 제안한 피치시점 검출법의 성능평가를 하기 위해서 각각의 피치시점 검출법을 제안한 피치 변경법에 적용하여서 주관적인 음질 평가를 수행하였다. 표 2는 피치시점 검출법에 따른 제안한 피치 변경법을 적용하여 합성한 다음 무작위로 추출된 청취자 10명에게 들려주고 발성시료에 대한 평균 MOS (Mean Opinion Score)를 측정하여 나타낸 것이다. 표 2에 나와있는 결과와 같이 제안한 피치시점 검출법에 의한 피치 변경법이 수동 피치시점 검출법에 비해 크게 떨어지지 않는 상대적으로 우수한 결과를 얻을 수 있었다.

VI. 결 론

본 논문에서는 피치주기를 압축하는 경우에는 F1과 F0의 율을 정수배로 유지시켜서 스펙트럼 왜곡을 최소화하는 영교차 단위의 시간축 조절에 의한 피치 변경법을 적용하였다. 이 방법은 시간축 조절을 제1포만트의 주파수 범위 내로 제한하여 변경율이 높아질 때 발생하는 스펙트럼의 왜곡을 줄였다. 또한 피치주기를 신장하는 경우에는 기존의 PSOLA 합성 방식에 의한 피치 변경법을 적용하였다. 파형 부호화법에서 화자의 개성과 명료성을 유지하려면 발성자의 중심이 되는 피치를 기준으로 하여 피치 변경이 이루어져야 한다. 따라서 피치 변경을 수행하기 위해서는 그 발성자의 정확한 피치시점을 검출할 수 있어야 한다. 또한 피치시점을 정확히 검출함으로써 음성분석시 피치 동기된 분석을 할 수 있고, 합성 시에는 여기원의 위상특성을 파악할 수 있으므로 개성이 강조된 합성음을 얻을 수 있다.

따라서 본 논문에서는 스펙트럼 왜곡률을 최소화하는 영교차 단위의 시간축 조절에 의한 피치 변경법과 피치 동기분석이 용이하고 다른 영역으로의 변환 과정이 불필요한 피치시점 검출법을 제안함으로써 고음질을 유지하면서 시간 영역에서만 처리됨으로써 계산량을 줄이고 스펙트럼 왜곡률을 최소화하면서도 위상을 그대로 보존할 수 있는 시간 영역에서의 피치 변경법을 제안하였다

결과적으로 전체 피치 변경율에 대해서는 기존의 방법에 비해서 제안한 방법의 스펙트럼 왜곡율이 0.73% 개선되었고, 피치 압축시에는 제안한 방법의 스펙트럼 왜곡율이 2.18% 개선되었다. 제안한 피치 변경율에 따른 피치 변경법은 상대적으로 스펙트럼 왜곡율을 최소화하면서도 위상을 그대로 보존할 수 있었다. 또한 제안한 피치시점 검출법에 의한 피치 변경법이 수동 피치시점 검출법에 비해 크게 떨어지지 않는 상대적으로 우수한 결과를 얻을 수 있었다.

감사의 글

본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음.

참 고 문 헌

1. G. Bristow, *Electronic Speech Synthesis*, McGraw-Hill, 1984.
2. L. R. Rabiner and R. W. Schafer, Digital Processing of Speech Signal, Prentice Hall, 1978.
3. T. W. Parsons, *Voice and Speech Processing*, McGraw-Hill, 1986.
4. E. J. Yannakoudakis and P. J. Hutton, *Speech Synthesis and Recognition Systems*, Ellis Horwood Ltd., 1987.
5. A. N. Ince, *Digital Speech Processing - Speech Coding, Synthesis and Recognition*, Kluwer Academic Publishers, 1992.
6. 김종국, 조왕래, 배명진, "피치조절에 의한 G.723.1 음성부호화기의 전송률 감소에 대한 연구," 제15회 신호처리합동학술대회, **15** (1), 146, 2002.
7. 이해군, 배명진, 임운천, "G-Peak 검출에 의한 음성신호의 피치시점검출," 제6회 신호처리합동학술대회, **6** (1), 58-61, 1993.
8. 박형빈, 조왕래, 김종득, 박원, 심도식, 배명진, "피치변경율에 따른 최적의 피치 변경법에 관한 연구," 제15회 음성 통신 및 신호처리 워크샵 논문집, **15** (1), 460-464, 1998.
9. 김종국, 박원, 배명진, "스펙트럼상에서 하모닉스 파형의 피크-피팅을 이용한 정확한 피치 검출에 관한 연구," 한국통신학회, 하계학술발표대회, **23** (2), 1308-1311, 2001.
10. B. E. Caspers and B. S. Atal, "Changing pitch and duration in LPC synthesised speech using multipulse excitation," *J. Acoust. Soc. Amer.*, **73** (1), 55, 1983.
11. A. Varga and F. Fallside, "A technique for using multipulse linear predictive speech synthesis in text-to-speech type system," *IEEE signal processing*, **ASSP-35** (4), 586-587, April 1987.
12. M. BAE, H. YOON, and S. ANN, "On altering the pitch of speech signals in waveform coding-alteration method by the LPC and pitch halving-," *J. Acoust., Soc., Korea*, **10** (5), 11-19, Oct. 1991.
13. T. F. Quatieri, and R. J. McAulay, "Shape invariant time-scale and pitch modification of speech," *IEEE Trans., Signal Processing*, **40** (3), 497-510, March 1992.
14. M. Bae, "On a pitch alteration method using scaling the harmonics compensated with the phase for speech synthesis," *J., Acoust., Society, Korea*, 15 (6), 99-103, December 1996.
15. T. Takagi, and E. Miyasaka, "A speech prosody conversion system with a high quality speech analysis-synthesis method," Proc. EUROSPEECH'93, 995-998, September 1993.

저자 약력

● 김 종 국 (Jong Kuk Kim)

2002년: 숭실대학원 정보통신공학과 졸업 (공학석사)
현재: 숭실대학원 정보통신공학과 박사과정
※ 주관심분야: 음성코딩, 음성인식, 음성합성, 디지털 신호처리 등

● 박 형 빈 (Hyung Bin Park)

1998년: 숭실대학교 정보통신공학과 졸업 (공학사)
2001년: 숭실대학교 정보통신공학과 졸업 (공학석사)
※ 주관심분야: 음성코딩, 음성인식, 음성합성, 디지털 신호처리 등

● 배 명 진 (Myung Jin Bae)

현재: 숭실대학교 정보통신공학과 교수
한국음향학회지 제22권 제3호 참조