

평탄화된 여기 스펙트럼에서 캡스트럼 피치 변경법에 관한 연구

On a Pitch Alteration Technique by Cepstrum Analysis of Flatten Excitation Spectrum

조 왕 래*, 함 명 규*, 배 명 진*

(Wang Rae Jo*, Myung Kyu Ham*, Myung Jin Bae*)

요 약

음성합성은 합성방식에 따라 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다. 특히 고음질 합성을 위해서는 파형부호화를 이용한 합성방식이 적합하다. 그렇지만, 파형부호화를 이용한 합성법은 여기 성분과 여파기 성분을 분리하지 않고 처리하기 때문에 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다. 따라서 파형부호화법을 규칙에 의한 합성에 적용되도록 음원피치를 변경시키기 위한 피치 변경법이 필요하게 된다. 본 논문에서는 스펙트럼 왜곡을 최소화하기 위해 캡스트럼의 성질을 이용하여 피치를 변경하는 방법에 대하여 제안하였다. 이 방법은 주파수영역상에서 여기 스펙트럼과 여파기 스펙트럼을 분리하여 여기 스펙트럼을 여기 캡스트럼으로 변환한 후 영값 삽입이나 삭제에 의해 피치를 변경하고 스펙트럼영역에서 피치 변경된 스펙트럼을 재구성하는 기법을 적용하였다. 제안한 방법의 성능을 평가하기 위해 스펙트럼 왜곡율을 측정하여 본 결과 평균 스펙트럼 왜곡율은 평균 2.29%이하로 유지되었으며 주관적인 음질도 평균 3.74로 우수하였다.

ABSTRACT

Speech synthesis coding is classified into three categories: waveform coding, source coding and hybrid coding. To obtain the synthetic speech with high quality, the synthesis by waveform coding is desired. However, it is difficult to apply waveform coding to synthesis by syllable or phoneme unit, because it does not divide the speech into the excitation and the formant component. Thus, it is required to alter the excitation in waveform coding for applying waveform coding to synthesis by rule. In this paper we propose a new pitch alteration method that minimizes the spectrum distortion by using the characteristic of cepstrum. This method splits the spectrum of speech signal into excitation spectrum and formant spectrum, and transforms the excitation spectrum into excitation cepstrum. The pitch of excitation cepstrum is altered by zero insertion or zero deletion, and the pitch altered spectrum is reconstructed in the spectrum domain. As a result of performance test, the average spectrum distortion was below 2.29% and the MOS was maintained above 3.74.

I. 서 론

합성을 위해 출력되는 데이터의 처리 방식에 따라서는 메모리형 합성법과 전송형 합성법으로 구분할 수 있다. 메모리형 합성법은 분석한 데이터를 메모리에 저장시켜 두고 필요에 따라 다시 합성하는 방법으로 모저(Mozer)법, 기본 주파수단위 분절법 등이 있다[2]. 이 방법은 메모리에 저장되는 데이터 형식이 번지당 일정 비트수로 제한되지 않으며, 분석시의 시간은 합성시의 시간과 반드시 일치시킬 필요도 없다. 또한 메모리의 저장 오류율은

무시될 정도이기 때문에 이를 방지하기 위한 별도의 기술이 거의 요구되지 않는다.

전송형 합성법으로는 부호화 방식에 따라서 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다[1]-[4]. 파형부호화법은 파형 자체의 잉여성분을 제거한 후에 부호화 하는 방법이며, PCM, ADPCM, ADM 등이 제안되어 있다. 최근 다양해진 음성서비스 분야에서는 고음질의 합성음을 요구하고 있다. 이러한 고음질 합성방식으로는 파형부호화법이 바람직하다. 이 부호화법은 인간의 개성과 감정을 대별해 주는 여기 정보와 메시지전달을 나타내는 여파기 정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다. 또한, 파형부호화법을 사용하

* 숭실대학교 정보통신공학과
접수일자: 1998년 9월 24일

번 데이터베이스에 저장해야 할 메모리 규모가 방대하고 음원피치의 변경이 어렵다는 문제점이 발생한다. 부호화에 필요한 메모리 문제는 현재의 기술수준으로 충분히 극복 가능하며, 나머지 문제를 해결하기 위해서는 음원피치를 변경시킬 수 있어야 한다.

본 논문에서는 피치 변경시에 발생하는 스펙트럼 왜곡을 최소화하기 위해 스펙트럼영역에서 여기 스펙트럼을 분리하여 캡스트럼영역으로 변환한 후 피치를 변경하는 방법을 새로이 제안하였다. 캡스트럼 피치 변경법에서는 피치를 변경하기 위해 큐프런시 상에서 영값을 삽입하거나 삭제할 때, 삽입이나 삭제할 위치를 선정하기가 어렵다. 따라서, 잘못된 위치선정에 의해 스펙트럼 왜곡이 발생하는 단점을 보완하기 위하여 스펙트럼 상에서 여기 스펙트럼을 분리하여 여기 캡스트럼으로 변환한 후 영값 삽입이나 삭제에 의해 피치를 변경하는 기법을 제안하였다.

II. 기존의 피치 변경법

지금까지 제안된 피치 변경법은 처리영역에 따라 시간영역법, 주파수영역법, 시간-주파수 혼성영역법으로 나눌 수 있다.

시간영역 피치 변경법으로는 Multi-Pulse법, LPC 신장법, 피치 반분법 등이 있다. Caspers와 Atal은 MPLPC에서 멀티 펄스 사이에 영을 삽입하거나 삭제함으로써 피치를 변경하였다[5]. 그러나 MPLPC상의 멀티 펄스 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[6]. 이 방법은 피치 신장시에는 자연스럽게 피치 주기가 늘어나지만 피치 주기를 줄이는 경우에는 파형의 일부분을 소거하고 평활화하는 방법을 사용하고 있기 때문에 스펙트럼 왜곡이 심하다. 피치 반분법은 변경하려고 하는 목적 피치의 2배 피치를 갖는 파형을 LPC 신장법에 의해 생성한 후 데시메이션에 의해 주기를 반분하는 피치 변경법이다[7]. 그러나 이 방법은 시간영역에서만 수행되기 때문에 스펙트럼 왜곡이 발생하여 합성음의 명료성이 저하된다.

Quatieri와 McAulay는 주파수영역 피치 변경법으로 음성 신호의 진폭 스펙트럼과 위상 스펙트럼을 분리하여 별도로 처리하는 방법을 제안하였다[8]. 진폭 스펙트럼에 대해서는 두드러진 스펙트럼 봉우리들을 추출하여 이것을 피치 변경율(ρ)만큼 인터폴레이션하거나 데시메이션하여 진폭 스펙트럼의 피치를 변경시킨다. 위상 스펙트럼에 대해서는 시간영역에서 구한 피치 개시시간에 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 인터폴레이션 또는 데시메이션을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

다른 주파수영역 피치 변경법으로는 평탄화 기법에 의

해 포먼트와 기본 주파수의 고조파를 분리하여 기본 주파수를 선형적으로 스케일링함으로써 피치를 변경하는 방법이 있다[9]. 이 방법은 스펙트럼 상에서 고조파를 스케일링함으로써 창함수의 특성도 변경되어 시간영역에서 위상을 복원하기가 어렵게 된다.

시간-주파수 혼성영역법으로는 캡스트럼의 특징을 이용하여 캡스트럼값이 거의 영이 되는 부분에 영값을 삽입하거나 삭제함으로써 피치를 변경하는 방법이 있다[10]. 그러나 이 방법은 위상의 보존이 어렵다는 문제점을 가지고 있다. Takagi와 Miyasaka가 제안한 시간-주파수 혼성영역법은 시간영역에서 피치변경을 하였을 때 나타나는 스펙트럼 왜곡을 스펙트럼영역 상에서 LPC포락을 통해 수정하는 방법이다[11]. 이 방법은 LPC스펙트럼 포락이 갖는 극점에 치중된 시스템 전달 특성 때문에 모든 유성음을 만족하지는 못한다는 한계성을 갖는다.

피치 변경시에 여파가 스펙트럼이 왜곡되면 성도 정보가 왜곡되므로 의사 내용을 제대로 보존할 수 없게 된다. 또한, 위상이 왜곡되면 인근 프레임간 진폭 레벨의 변동이 커져서 음소간의 연결이 부자연스럽게 된다. 따라서 피치를 변경할 때에는 위상을 보존하면서도 스펙트럼 왜곡을 최소화 할 수 있는 피치 변경법이 필요하다.

III. 캡스트럼 피치 변경법

음성 신호의 캡스트럼 분석은 캡스트럼 상에서 여기 특성과 여파기 특성을 쉽게 분리할 수 있기 때문에 음성 신호의 분석에 많이 사용된다. 즉, 그림 3-1(a)의 음성 스펙트럼을 캡스트럼으로 나타내면 그림 3-1(b)와 같다[2]. 이때 캡스트럼의 낮은 큐프런시(quefrensy) 성분은 성도 여파기 특성을 나타내고, 높은 큐프런시 성분은 성대 여기 특성을 나타낸다. 따라서 식 (3.1)과 같은 창함수(window)를 이용하면 성도 여파기의 특성을 구할 수 있다.

$$k(n) = \begin{cases} 1, & |n| < n_0 \\ 0, & |n| \geq n_0 \end{cases} \quad (3.1)$$

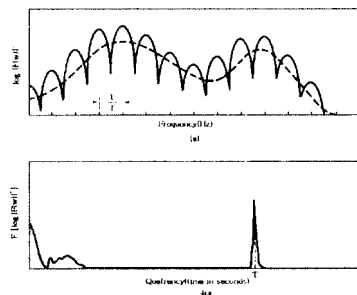


그림 3-1. 음성 신호의 스펙트럼과 캡스트럼

- (a) 음성 신호의 스펙트럼
- (b) 음성 신호의 캡스트럼

Fig. 3-1. Spectrum and Cepstrum of speech signal
(a) Spectrum of speech signal
(b) Cepstrum of speech signal

여기서 n_0 는 피치 주기 N_p 보다 작게 선택된다. 이렇게 구해진 여파기 스펙트럼은 음성 신호의 공명 특성을 나타내며, 포먼트 스펙트럼과 같아진다. 또한, 캡스트럼상의 성도 여파기 특성은 쿠프런시가 증가함에 따라 급속히 감소하는 특성을 갖는다.

한편, 캡스트럼에 식 (3.2)와 같은 창함수(window)를 적용하면 음성 신호의 여기 특성을 구할 수 있다.

$$k(n) = \begin{cases} 0, & |n| < n_0 \\ 1, & |n| \geq n_0 \end{cases} \quad (3.2)$$

캡스트럼 분석에 의해 음성을 합성할 때는 낮은 쿠프런시의 캡스트럼을 여파기 특성으로 취하고 높은 쿠프런시의 캡스트럼을 여기 특성으로 취하여 이들을 권별무선함으로써 음성 신호를 합성한다. 이때 여기 특성을 변경하여 피치를 변경할 수 있게 된다.

캡스트럼의 특징은 대부분의 캡스트럼 값이 영(zero) 쿠프런시 부근에 존재하며, 이들 값은 쿠프런시 증가에 따라 급속히 감소하여 피치 주기 부근에서는 거의 영이 된다. 피치를 변경하기 위해서는 캡스트럼 값이 거의 영이 되는 부분에 변경하려는 주기만큼의 영 캡스트럼을 삽입하거나 삭제하게 된다[10]. 이러한 방법은 여파기 특성에는 영향을 주지 않으면서 여기 특성만을 변경시키기 위해 영값을 삽입하거나 삭제하기 위한 위치의 선정이 매우 중요하다. 현재 분석중인 음성구간의 피치를 사전에 알고 있다면 피치 주기 근방에서 영값을 삽입하거나 삭제하는 것이 바람직하다. 그러나 분석중인 창함수내에서 시간에 따라 피치 주기가 변화하고 있는 경우에는 피치 주기 근방의 캡스트럼 필스가 일정 폭을 유지하게 되어 영값을 삽입하거나 삭제하기 위한 위치의 선정에 어려움이 따르게 된다. 따라서, 잘못된 위치선정은 합성음질에 큰 열화를 초래하게 된다.

IV. 평탄화된 여기 스펙트럼에서 캡스트럼 피치 변경법

캡스트럼 피치 변경법에서 피치를 변경하기 위해 쿠프런시 상에서 영값을 삽입하거나 삭제하는 경우에 최적의 위치를 선정하기가 어렵고, 잘못된 위치선정에 의해 스펙트럼 왜곡이 발생하는 단점이 있다. 따라서 본 논문에서는 스펙트럼 상에서 음성 스펙트럼을 여파기 스펙트럼과 여기 스펙트럼으로 분리하여 처리하는 방법을 제안하였다.

먼저 음성신호를 푸리에 변환하여 진폭 스펙트럼과 위상 스펙트럼으로 분리하여야 한다. 푸리에 변환에 의해 얻어진 음성 스펙트럼은 식 (4.1)과 식 (4.2)와 같이 진폭 스펙트럼과 위상 스펙트럼으로 나타낼 수 있다.

$$M(K) = 10 \log S^2(K) \quad (4.1)$$

$$\varphi(K) = \tan^{-1} \frac{Im[S(K)]}{Re[S(K)]} \quad (4.2)$$

여기서 $Re[S(K)]$ 는 음성 스펙트럼의 실수성분이고, $Im[S(K)]$ 는 음성 스펙트럼의 허수성분을 나타낸다. 진폭 스펙트럼을 여기 스펙트럼과 여파기 스펙트럼으로 분리하기 위해 진폭 스펙트럼에 식 (4.3)과 같은 리프터함수를 적용하여 근사적인 포먼트 스펙트럼을 구한다.

$$H(K - \frac{K_0}{2}) = \frac{1}{K_0} \sum_{L=0}^{K_0} M(K - L) \quad (4.3)$$

이때 근사적인 여파기 스펙트럼, $E(K)$ 는 다음 식 (4.4)와 같이 구할 수 있다.

$$E(K) = S(K) - M(K) \quad (4.4)$$

이 여기 스펙트럼을 역푸리에 변환하여 캡스트럼영역으로 나타낸다. 이것은 여기 성분만의 캡스트럼이므로 피치 필스 위에는 거의 영값을 갖게 된다. 피치 주기를 변경하기 위하여 영-쿠프런시와 피치필스 사이에 변경하려는 주기만큼의 영값을 삽입하거나 삭제한다. 이때 영값을 삽입하거나 삭제하는 위치는 합성음질에 거의 영향을 미치지 않게 된다. 피치를 변경하고 다시 푸리에 변환하여 피치가 변경된 여기 스펙트럼 $E'(K)$ 를 구하여 식 (4.5)와 같이 피치가 변경된 진폭 스펙트럼을 구성한다.

$$M'(K) = H(K) + E'(K) \quad (4.5)$$

이 대수 진폭 스펙트럼에 지수함수를 적용하여 진폭 스펙트럼을 구하고 이와 동시에 피치가 변경된 위상 스펙트럼을 사용하여 역푸리에 변환하면 피치가 변경된 음성신호가 얻어진다. 이러한 과정을 그림 4-1에 나타내었다.

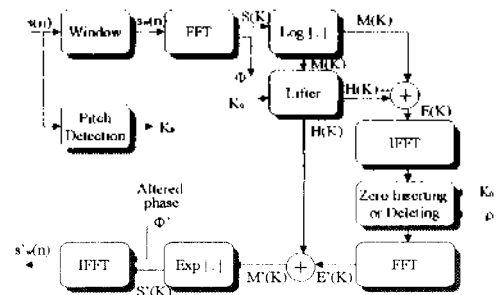


그림 4-1. 제안한 피치 변경법의 처리과정
Fig. 4-1. A block diagram of proposed method

V. 피치검출

신호원부호화법과는 달리 파형부호화법에서 피치를 변경하려면 사전에 발생자의 피치 주기를 알고 있어야 한다. 이것은 발생자의 억양이나 감정의 변화에 의하여 피치가 상대적으로 변화하기 때문이다. 특히 파형부호화에서는 발생자의 개성과 메시지 정보를 보존함으로써 음질

의 명료성이 우수하다. 이 때문에 피치 변경시에는 발생자가 주로 사용하는 피치 주기를 기준으로 피치를 변경시킬 필요가 있다. 따라서 정확한 피치 검출이 선행되어야 한다. 지금까지 제안된 피치 검출법은 크게 시간영역법, 주파수 영역법, 그리고 시간-주파수 혼성영역법으로 나눌 수 있다[1]-[3].

본 논문에서는 피치 검출법으로 시간영역의 면적비교법을 적용하였다. 그렇지만, 합성을 위해 파형을 편집하는 경우에는 피치의 추출이 반드시 자동화될 필요는 없으며 면적비교법[12]과 함께 눈으로 피치를 추출하는 반자동법이나, 눈으로 찾는 수동법으로 처리하여도 된다.

VI. 실험 및 결과

제안한 피치 변경법의 성능을 평가하기 위해 IBM-PC/pentium(150MHz)에 음성 입·출력용 16bit AD/DA 변환기를 인터페이스하여 사용하였고, 알고리즘은 C언어로 구현하였다. 음성 시료로는 24세의 남성, 30세의 남성, 27세의 여성, 29세의 여성에게 아래와 같은 문장들을 각각 5회씩 발성하게 하고 11kHz로 표본화하여 저장하였다:

- 발성1 : /인수네 꼬마는 천재소년을 좋아한다./
- 발성2 : /예수님께서 천지 창조의 교훈을 말씀하셨다./
- 발성3 : /여기는 음성합성 연구실입니다./
- 발성4 : /창공을 헤쳐나가는 인간의 도전은 끝이 없다./
- 발성5 : /이용해 주셔서 감사합니다./

알고리즘의 처리를 위한 분석 프레임의 길이는 256표본을 사용하였다. 먼저 한 프레임의 음성신호에 대해 면적비교법[12]을 사용하여 피치 주기를 결정하였다. 동시에 음성신호에 대해 해밍윈도우를 취하고 FFT하여 주파수영역으로 변환하였다. 여기에 대수 연산을 적용하여 대수 스펙트럼을 구성하고, 근사적인 포먼트 스펙트럼을 구하기 위해 기본 주파수를 차단 주파수로 갖는 리프터(lifter) 함수를 대수 스펙트럼에 적용하였다. 다음으로 원래의 대수 스펙트럼에서 근사적인 포먼트 스펙트럼을 빼내어 평탄화된 여기 스펙트럼을 구하였다. 이 평탄화된 여기 스펙트럼을 IFFT하여 캡스트럼 영역으로 변환하고, 여기 캡스트럼상에서 영값 삽입이나 삭제를 통하여 피치 주기를 변경하였다. 피치가 변경된 여기 캡스트럼을 FFT하여 피치가 변경된 여기 스펙트럼을 구성하고 근사적인 포먼트 스펙트럼을 다시 더하여 줌으로써 피치가 변경된 진폭 대수 스펙트럼을 구성하였다. 피치가 변경된 진폭 스펙트럼을 지수함수에 통과시켜 진폭 스펙트럼을 만들고, 피치 변경된 위상 스펙트럼[9]과 함께 IFFT를 수행하여, 피치 변경된 음성파형을 합성하였다.

제안된 피치 변경법의 성능평가는 객관적인 평가로서 스펙트럼 왜곡율과 주관적인 평가로서 MOS를 측정하였다. 음성신호의 피치 주기를 120%에서 200%까지 변경시키면서 원래음성 신호의 스펙트럼에 비해 나타나는 스펙트럼 왜곡율을 측정하여 백분율로 환산하여 표 6-1에 제

시하였다. 그림 6-1에는 피치 주기를 120%로 신장한 경우의 처리 결과 예를 나타내었다. 스펙트럼의 비교 기준은 피치가 변경되기 이전의 원래 음성의 스펙트럼을 사용하였다. 피치를 변경시키면 원래의 스펙트럼과 직접 비교할 수 없기 때문에 피치주기를 120%, 140%, 160%, 180%, 200%로 각각 신장시킨 다음에 83%, 71%, 62%, 55%, 50%로 각각 압축하여 원래의 음성 스펙트럼과 고조파를 일치시킨 다음에 에너지 왜곡율을 측정하였다. 표 6-1에 제시한 바와 같이 평균 스펙트럼 왜곡율은 기존의 성분분리형 피치 변경법의 2.47%에서 제안한 방법이 2.29%로 0.18%가 개선되었다.

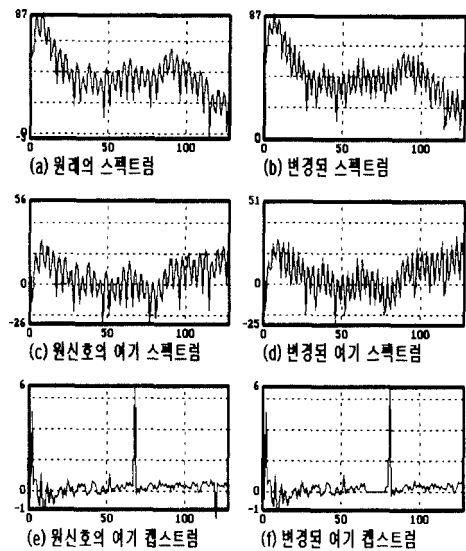


그림 6-1. 피치 주기를 120%로 신장한 경우의 처리 결과
Fig. 6-1. The result of 120% pitch alteration

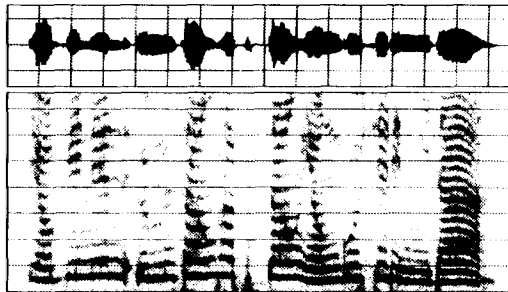
표 6-1. 피치 변경율에 따른 스펙트럼 왜곡율 비교
Table 6-1. The comparison of spectrum distortion by pitch alteration

| 변경율 | 기존의 방법 | | | 제안한 방법 | | |
|------|--------|------|-------|--------|------|-------|
| | 남성 | 여성 | 평균(%) | 남성 | 여성 | 평균(%) |
| 120% | 1.67 | 2.03 | 1.85 | 1.49 | 1.83 | 1.66 |
| 140% | 2.04 | 2.32 | 2.18 | 1.75 | 2.15 | 1.95 |
| 160% | 2.18 | 2.50 | 2.34 | 1.84 | 2.38 | 2.11 |
| 180% | 2.67 | 2.98 | 2.83 | 2.52 | 2.74 | 2.63 |
| 200% | 2.84 | 3.43 | 3.14 | 2.82 | 3.38 | 3.10 |
| 평균 | 2.28 | 2.65 | 2.47 | 2.08 | 2.50 | 2.29 |

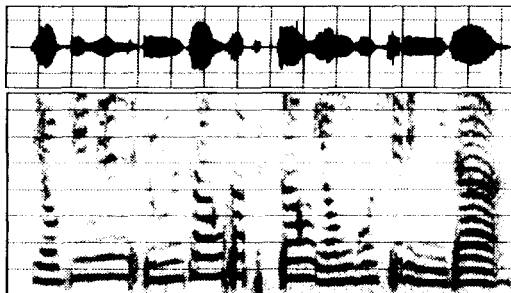
주관적인 음질 평가를 위해서 음성 시료로 사용된 5가지 문장을 기존의 스펙트럼 스케일링법과 제안한 피치 변경법을 적용하여 50%, 70%, 90%, 120%, 140%로 각각 변경시켜 합성하고, 무작위로 추출된 청취자 20명에게 들려주고

그에 대한 MOS를 측정하였다. MOS의 등급은 Excellent(5), Fair(4), Good(3), Poor(2), Unsatisfactory(1)로 구분하였다.

표 6-2에 주관적인 평가 결과를 제시하였고, 그림 6-2에는 발성 3의 피치주기를 80%로 변경시켜 합성한 합성 문장의 파형과 스펙트로그램을 나타내었다. 표 6-2에 나타난 바와 같이 주관적인 평가 결과는 기존의 스펙트럼 스케일링에 의한 피치 변경법의 MOS 3.34에 비해 3.74로 향상되었다.



(a) 원음성파형과 스펙트로그램
(a) Original waveform and spectrogram



(b) 피치 주기를 80%로 압축한 경우의 음성파형과 스펙트로그램
(b) Waveform and spectrogram of 80% pitch alteration

그림 6-2. /여기는 음성합성 연구실입니다./의 처리 결과
Fig. 6-2. The experiment result of /여기는 음성합성 연구실입니다./

표 6-2. 피치 변경에 따른 MOS 비교
Table 6-2. The comparison of MOS by pitch alteration

| 발 성 | 5 레벨 MOS | |
|------|----------|--------|
| | 기존의 방법 | 제안한 방법 |
| 발성 1 | 3.4 | 3.8 |
| 발성 2 | 3.2 | 3.6 |
| 발성 3 | 3.5 | 3.7 |
| 발성 4 | 3.1 | 3.7 |
| 발성 5 | 3.5 | 3.9 |
| 평 균 | 3.34 | 3.74 |

VII. 결 론

고음질 음성합성을 위해서는 파형부호화법이 바람직하

지만 파형부호화법을 사용하면 음원피치의 변경이 어렵다는 문제점이 발생한다. 이러한 문제의 해결을 위해서는 파형부호화법을 규칙에 의한 합성에 적용하기 위해 음원 피치를 변경시킬 수 있어야 한다.

따라서, 본 논문에서는 피치변경시 스펙트럼 왜곡을 최소화하기 위해 주파수영역에서 여기 스펙트럼을 분리하고 켈스트럼영역에서 피치를 변경하는 시간-주파수 혼성형 피치 변경법을 새로이 제안하였다. 제안한 방법은 기존의 켈스트럼 피치 변경법에서 피치를 변경하기 위해 쿼펀시 상에서 영값을 삽입하거나 삭제하고자 할 때 그 위치를 선정하기가 어렵고, 잘못된 위치선정에 의해 스펙트럼 왜곡이 발생하는 단점을 보완하기 위해, 스펙트럼상에서 여기 스펙트럼을 분리하여 여기 켈스트럼으로 변경한 후 영삽입이나 영삭제에 의해 피치를 변경하는 기법을 적용하였다.


제안한 방법에 의해 피치가 변경된 음성의 스펙트럼 왜곡을 측정하여 본 결과 평균 스펙트럼 왜곡율은 기존의 성분분리형 피치 변경법의 2.47%에서 제안한 방법이 2.29%로 개선되었다. 또한 주관적인 음질(MOS)도 기존의 성분분리형이 3.34인데 비해 3.74로 개선되었다. 본 논문에서 제안한 방법은 스펙트럼 왜곡특성 및 위상 왜곡특성이 우수하나 계산량이 많다는 단점이 있기 때문에 이에 대한 연구가 계속되어야 한다.

참 고 문 헌

1. L.R. Rabiner & R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.
2. Panos E. Papamichalis, *Practical Approaches to Speech Coding*, Prentice-Hall, 1987.
3. Thomas W.Parsons, *Voice and Speech Processing*, McGraw-Hill, 1986.
4. A.M. Kondoz, *Digital Speech Coding for Low Bit Rate Communications Systems*, John Wiley & Sons, 1994.
5. B.E. Caspers and B.S. Atal, "Changing Pitch and Duration in LPC Synthesised Speech using Multipulse Excitation," *J. Acoust. Soc. Amer.*, Vol.73, No.1, pp.55, 1983.
6. A. Varga and F. Fallside, "A Technique for Using Multipulse Linear Predictive Speech Synthesis in Text-to-speech Type System," *IEEE signal processing*, Vol.ASSP-35, No.4, pp.586-587, 1987.
7. 강동규, 김은재, 배명진, 안수길, "음성 파형의 halving 기법에 의한 파형코딩의 피치변경에 관한 연구," *한국음향학회 추계발표회(국제음향학회)논문집*, pp.107-111, 1990.
8. T.F. Quatieri, R.J. McAulay, "Shape Invariant Time-Scale and Pitch Modification of Speech," *IEEE Trans. Signal Processing*, Vol.40, No.3, pp.497-510, 1992.
9. 배명진, "위상보상된 고조파 스케일링에 의한 음성합성용 피치변경법," *한국음향학회, 한국음향학회지*, Vol.13, No.6, pp. 91-97, 1994.
10. M.J. Bae and S.H. Lee, "On a Cepstral Technique for Pitch Control in the High Quality Text-To-Speech Type System," *39'th Midwest Symposium on Circuits and Systems*, *Proceeding of MWSCAS'96*, pp.803- 806, 1996.


11. T. Takagi, E. Miyasaka, "A Speech Prosody Conversion System with a high Quality Speech Analysis-Synthesis Method," Proc. EUROSPEECH'93, pp.995-998, 1993.
12. 배명진, 안수길, "변적 비교법을 이용한 음성신호의 고속 피치 추출," 전자공학회지, Vol.22, No.2, pp.13-17, 1985.
13. 조왕래, 김동성, 배명진, "고음질 합성방식용 V/UV 스펙트럼상의 피치변경법에 관한 연구," 한국음향학회, 한국음향학회지, Vol. 15, No. 6, pp.99-103, 1996.

▲조 왕 래(Jo Wang Rae) 1967년 3월 25일생




1996년 2월 : 숭실대학교 정보 통신 공학과 졸업(공학사)
 1998년 2월 : 숭실대학교 대학원 전기공학과 석사과정 졸업(공학석사)
 현재 : 벽성대학 전자과 교수
 ※주관심분야 : 음성 신호처리, 음성 통신

▲함 명 규(Ham Myung Kyu) 1974년 5월 23일생



1997년 2월 : 숭실대학교 정보 통신 공학과 졸업(공학사)
 1997년 3월 ~ 현재 : 숭실대학교 대학원 정보통신공학과 석사과정
 ※주관심분야 : 음성 신호처리, 음성 통신

▲배 명 진(Bae Myung Jin)



현재 : 숭실대학교 정보통신공학과 교수
 (제14권 1E호 참조)