1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

평탄화된 여기 스펙트럼에서 켑스트럼 피치 변경법에 관한 연구

조왕래, 함명규(*), 배명진(*)
벽성대학 전자과, (*)숭실대학교 정보통신공학과

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

Wangrae Jo, Myungkyu Ham(*), Myungjin Bae(*)
Dept. of Electronics Byuksung College, (*)Dept. of Telecomm. Engr. Soongsil University
wrjo@www.byuksung-c.ac.kr, mjbae@saint.soongsil.ac.kr

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 excitation and 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 behavior of cepstrum. This method splits the spectrum of speech signal into excitation spectrum and formant spectrum and transforms the excitation spectrum into cepstrum domain. The pitch of excitation cepstrum is altered by zero insertion or zero deletion and the pitch altered spectrum is reconstructed in spectrum domain. As a result of performance test, the average spectrum distortion was below 2.29%.

1. 서 론

음성합성은 합성단위에 따라서 문장단위, 음절단위, 음소단위 등으로 나눌 수 있다. 예를 들면 가전제품의 사용법 안내나 녹음안내 사항에 대해서는 보통 문장 단위로 합성하고, 시간 및 기후안내에는 단어 단위의 합성법이 적용된다. 또한 상황이 복잡하고 모든 분야에 적용되어야 할 합성기법은 음절이하의 합성단위가 바람직하게 된다.

부호화 방식에 따라서는 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다[1-3]. 파형부호화법은 파형 자체의 잉여성분을 제거한 후에 부호화 하는 방법이며, PCM, ADPCM, ADM 등이 제안되어 있다. 이 부호화법은 인간의 개성과 감정을 대별해 주는 여가정보와 메시지전달을 나타내는 성도의 여과기정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다.

최근 다양해진 음성서비스 분야에서는 고음질의 합성음을 요구하고 있다. 이러한 고음질 합성방식으로는 파형부호화법이 바람직하다. 그렇지만 파형부호화법을 사용하면 데이터베이스에 저장해야 할 메모리 규모가 방대하고 음원피치의 변경이 어렵다는 문제점이 발생한다. 그러나 부호화에 필요한 메모리 문제는 현재의 기술수준으로 충분히 극복이 가능하다. 나머지 문제의 해결법으로는 파형부호화법을 규칙에 의한 합성에 적용되도록 음원피치를 변경시킬 수 있어야 한다.

본 논문에서는 스펙트럼 왜곡을 최소화하기 위해 주파수 영역에서 여기 스펙트럼을 분리하여 켑스트럼 영역에서 피치를 변경하는 피치변경법을 새로이 제안하였다. 제안한 방법은 켑스트럼 피치변경법에서 피치를 변경하기 위해 큐퍼런시 상에서 영값을 삽입하거나 삭제하는 위치를 선정하기가 어렵고 잘못된 위치선정에 의해 스펙트럼 왜곡이 발생하는 단점을 보완하기 위해 스펙트럼상에서 여기 스펙트럼을 분리하여 켑스트럼 영역에서 영삽입이나 영삭제에 의해 피치를 변경하는 기법을 적용하였다.

2. 기존의 피치 변경법

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

시간 영역 피치 변경법으로는 Multi-Pulse법, LPC 신장법, 피치반분법 등이 있다. Caspers와 Atal은 MPLPC에서 멀티 펄스 사이에 영을 삽입하거나 삭제함으로써 피치를 변경하는 방법을 제안하였으나[4], 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[5]. 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[6]. 그러나 이방법은 시간 영역에서만 수행되기 때문에 스펙트럼 왜곡이 발생하여 합성음의 명료성이 저하된다.

Quatieri와 McAulay는 주파수 영역 피치 변경법으로 음성 신호의 진폭 스펙트럼과 위상 스펙트럼을 분리하여 별도로 처리하는 방법을 제안하였다[7]. 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 스펙트럼 상에서 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[8]. 이 방법은 스펙트럼을 원래의 음성스펙트럼으로 대신하는 방법이나 스펙트럼 상에서 고조파를 스케일링시킴으로서 창함수의 특성도 변경되어 시간영역에서 위상을 복원하기가 어렵게 된다.

시간-주파수 혼성법으로는 켑스트럼의 특성을 이용하여 켑스트럼값이 거의 영이 되는 부분에서 영값을 삽입하거나 삭제함으로써 피치를 변경하는 방법이 있다[9]. 켑스트럼의 특징은 대부분의 켑스트럼 값이 영(zero) 큐퍼런시 부근에 존재하며 이들 값은 큐퍼런시 증가에 따라 급속히 감소하여 피치 주기 부근에서는 거의 영(zero)이 된다. 피치를 변경하기 위해서는 켑스트럼 값이 거의 영(zero)이 되는 부분에 변경하려는 주기만큼의 영(zero) 켑스트럼을 삽입하거나 삭제하게 된다[9]. 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 켑스트럼 펄스가 일정 폭을 유지하게 되어 영값을 삽입하거나 삭제하기 위한 위치의 선정에 어려움이 따르게 되며 잘못된 위치선정은 합성음질에 심각한 영향을 초래하게 된다.

성분분리형 피치변경법은 음성신호의 스펙트럼을 여파기 스펙트럼과 여기 스펙트럼으로 분리하여 식 2-1과 같이 여기 스펙트럼을 스케일링함으로써 피치 주기를 변경시키는 방법이다.

$$E'(K) = E(K \times \rho^{-1}) \qquad (2\text{-}1)$$

여기서 ρ^{-1}은 주파수축 스케일링 율이다. 기본 주파수를 높이기 위해서는 고조파의 간격을 ρ^{-1}만큼 늘이고 기본 주파수를 낮추기 위해서는 고조파의 간격을 ρ^{-1}만큼 줄이게 된다. 이렇게 여기 스펙트럼의 스케일링에 의해 고조파의 간격을 줄이거나 늘이게 되면 높은 주파수밴드에 스펙트럼 복사나 삭제에 의한 왜곡이 발생하게 된다. 이러한 스펙트럼 왜곡은 합성음에서 버즈음으로 나타나 합성음질의 열화를 초래하게 된다.

3. 평탄화된 여기 스펙트럼에서 켑스트럼 피치변경법

제안한 방법은 켑스트럼 피치변경법에서 피치를 변경하기 위해 큐퍼런시 상에서 영값을 삽입하거나 삭제하는 위치를 선정하기가 어렵고 잘못된 위치선정에 의해 스펙트럼 왜곡이 발생하는 단점을 보완하기 위해 스펙트럼상에서 음성 스펙트럼을 여파기 스펙트럼과 여기 스펙트럼으로 분리하여 처리하는 방법이다.

먼저 음성신호를 푸리에 변환하여 진폭 스펙트럼과 위상 스펙트럼으로 분리하여야 한다. 음성신호의 푸리에 변환은 식 3-1에 의해 수행된다.

$$S(K) = \int_{-\infty}^{\infty} s(n) e^{-j\frac{n}{2\pi N}k} dn \qquad (3\text{-}1)$$

푸리에 변환에 의해 얻어진 음성 스펙트럼은 식 3-2와 식 3-3과 같이 진폭 스펙트럼과 위상 스펙트럼으로 나타낼 수 있다.

$$M(K) = 10 \log S^2(K) \qquad (3\text{-}2)$$

$$\varphi(K) = \tan^{-1} \frac{Im[S(K)]}{Re[S(K)]} \qquad (3\text{-}3)$$

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 3-4와 같은 리프터함수를 적용하여 근사적인 포만트 스펙트럼을 구한다.

$$H(K) = \frac{1}{K_0} \sum_{L=0}^{K_0} M(K-L) \qquad (3\text{-}4)$$

이때 여기 스펙트럼 $E(K)$는 다음 식 3-5와 같이 구할 수 있다.

$$E(K) = S(K) - H(K) \qquad (3\text{-}5)$$

이렇게 구해진 여기 스펙트럼을 IFFT를 수행하여 켑스트럼 영역으로 변환한다. 이것은 여기 성분만의 켑스트럼이므로 피치 펄스 외에는 거의 영값을 갖게 된다. 피치 주기를 변경하기 위하여 영-큐퍼린시와 피치펄스 사이에 변경하려는 주기만큼의 영값을 삽입하거나 삭제한다. 이때 영값을 삽입하거나 삭제하는 위치는 합성음질에 거의 영향을 미치지 않게 된다. 피치를 변경하고 다시 FFT를 수행하여 피치가 변경된 여기 스펙트럼 $E'(K)$를 구하여 식 3-6과 같이 피치가 변경된 진폭 스펙트럼을 구성한다.

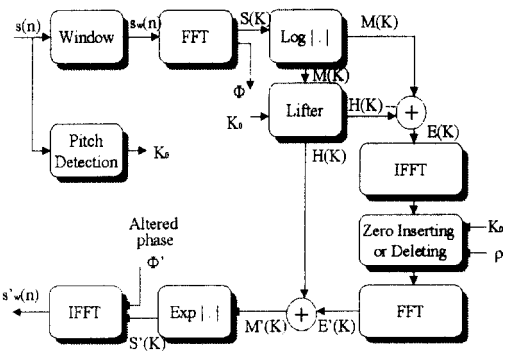

그림 3-1. 제안한 피치 변경법의 처리과정

$$M'(K) = H(K) + E'(K) \qquad (3\text{-}6)$$

이 대수 진폭 스펙트럼에 지수함수를 적용하여 진폭스펙트럼을 구하고 이와 동시에 피치가 변경된 위상 스펙트럼을 사용하여 IFFT를 수행하면 피치가 변경된 음성신호가 얻어진다. 이러한 과정을 그림 3-1에 나타내었다.

4. 실험 및 결과

제안한 피치 변경법의 성능을 평가하기 위해 IBM-PC/pentium(200㎒)에 음성 입·출력용 16bit AD/DA 변환기를 인터페이스하여 사용하였고 알고리즘은 C언어로 구현하였다. 성능 평가를 위한 음성 자료는 24세의 남성, 30세의 남성, 27세의 여성, 29세의 여성에게 아래와 같은 문장들을 각각 5회씩 발성하게 하여 11㎑로 표본화하고 16bit로 양자화하여 저장하였다.

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

알고리즘의 처리를 위한 분석 프레임의 길이는 256표본을 사용하였다. 먼저 한 프레임의 음성 신호에 대해 면적 비교법을 사용하여 피치 주기를 결정하였다. 동시에 음성신호에 대해 해밍윈도우를 취하고 푸리에 변환하여 주파수 영역으로 변환하였다. 여기에 로그 연산을 적용하여 로그 스펙트럼을 구성하고 근사적인 포만트 스펙트럼을 구하기 위해 기본 주파수를 차단 주파수로 갖는 리프터(lifter) 함수를 로그 스펙트럼에 적용하였다. 다음으로 원래의 로그 스펙트럼으로 부터 근사적인 포만트 스펙트럼을 빼어내어 평탄화된 여기 스펙트럼을 구하였다. 이 평탄화된 여기 스펙트럼을 역푸리에 변환하여 켑스트럼 영역으로 변환하여 영값 삽입이나 삭제를 통하여 피치주기를 변경하였다. 피치가 변경된 여기 켑스트럼은 푸리에 변환하여 피치가 변경된 여기 스펙트럼을 구성하여 근사적인 포만트 스펙트럼을 다시 더하여 줌으로써 피치가 변경된 진폭 스펙트럼을 구성하였다. 여기에 지수함수와 IFFT를 적용하여 피치가 변경된 음성 신호를 재구성하였다.

제안된 피치 변경법의 성능을 평가하기 위해 스펙트럼 왜곡율 측정하였다. 음성 신호의 피치 주기를 120%에서 200%까지 변경시키면서 원래음성 신호의 스펙트럼에 비해 나타나는 스펙트럼의 왜곡율을 측

표 4-1. 피치 변경에 따른 스펙트럼 왜곡율 비교

	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

정하여 백분율로 환산하여 표 4-1에 제시하였고 그림 4-1에는 피치 주기를 120%로 신장한 경우의 처리 결과를 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 4-1에 제시된 바와 같이 평균 스펙트럼 왜곡율은 기존의 성분분리형 피치변경법의 2.47%에서 제안한 방법이 2.29%로 0.18%가 개선되었다.

5. 결 론

파형 부호화법은 인간의 개성과 감정을 대별해주는 여기정보와 메시지전달을 나타내는 성도의 여파기정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다. 이러한 문제의 해결을 위해서는 파형부호화법을 규칙에 의한 합성에 적용되도록 음원피치를 변경시킬 수 있어야 한다.

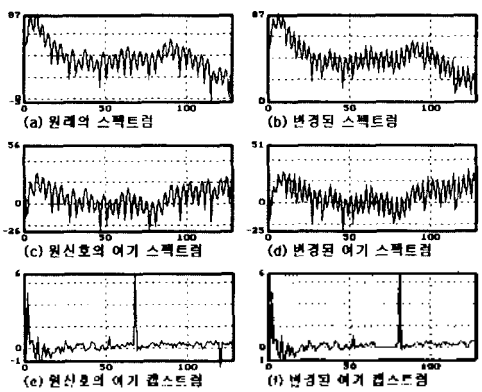

그림 4-1. 처리결과의 예(120% 신장)

따라서 본 논문에서는 스펙트럼 왜곡을 최소화하기 위한 위해 주파수 영역에서 여기 스펙트럼을 분리하여 캡스트럼 영역에서 피치를 변경하는 피치변경법을 새로이 제안하였다. 제안한 방법은 스펙트럼상에서 여기 스펙트럼을 분리하여 캡스트럼 영역에서 영삽입이나 영삭제에 의해 피치를 변경하는 기법을 적용하였다.

제안한 방법에 의해 피치가 변경된 음성의 스펙트럼 왜곡을 측정하여 본 결과 평균 스펙트럼 왜곡율은 기존의 성분분리형 피치변경법의 2.47%에서 제안한 방법이 2.29%로 개선되었다. 본 논문에서 제안한 방법은 스펙트럼 왜곡특성 및 위상 왜곡특성이 우수하나 계산량이 많다는 단점이 있기 때문에 복잡한 계산을 간략화 하는 방법에 대해서 더욱 연구해야 하겠다.

6. 참고 문헌

[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] 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.

[5] 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.

[6] 강동규, 김을제, 배명진, 안수길, "음성 파형의 halving 기법에 의한 파형코딩의 피치변경에 관한 연구," 한국음향학회 추계발표회(국제음향학회)논문집, pp.107-111, 1990.

[7] 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.

[8] 배명진, "위상보상된 고조파 스케일링에 의한 음성합성용 피치변경법," 한국음향학회, 한국음향학회지, Vol.13, No.6, pp.91-97, 1994.

[9] 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.