서브밴드 선형근사에 의한 피치변경법에 관한 연구

김 영 규, 김 봉 영, 배 명 진
숭실대학교 정보통신공학과

A Study on the Pitch Alteration Technique by Sub-band Linear Approximation in Spectrum

YoungKyu Kim, BongYoung Kim , MyungJin Bae
Dept. of Information & telecommunication Engr. Soongsil Univ.
mjbae@saint.ssu.ac.kr

Abstract

음성합성은 합성방식에 따라 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다. 특히 고음질 합성을 위해서는 파형부호화를 이용한 합성방식이 적합하다. 하지만 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 서브 선형근사에 의하여 스펙트럼 평탄화 시킨 후 스펙트럼 스케일링을 이용하여 피치를 변경하는 방법에 대하여 제안하였다. 기존 방법인 LPC법, Cepstrum법과 비교하여 어느정도의 우수성을 보이는지 평가하였고 평가방법은 각각의 평탄화된 신호의 분산을 구하여 평탄화의 정도를 측정하였다. 이때 평탄화된 신호는 최고점이 영이 되도록 정규화 시키고 평균이 영인 분산을 계산하였다. 제안한 방법의 성능을 평가하기 위해 스펙트럼 왜곡율을 측정하여 본 결과 평균 스펙트럼 왜곡율은 평균 2.12%이하로 유지되었으며 실험결과 제안한 방법이 기존의 방법보다 우수함을 보여주었다.

Ⅰ. 서 론

음성합성 시스템은 분류기준에 따라 분석된 데이터를 그대로 합성에 이용하는 분석에 의한 합성법과 규칙에 의해 음성을 발생시켜 합성하는 규칙에 의한 합성법으로 구분할 수 있다. 음성합성은 합성단위에 따라서 문장단위, 음절단위, 음소단위 등으로 나눌 수 있다. 부호화 방식에 따라서는 파형부호화법, 신호원 부호화법, 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 사용하면 데이터베이스에 저장해야 할 메모리 규모가 방대하고 음원피치의 변경이 어렵다는 문제점이 발생한다. 부호화에 필요한 메모리 문제는 현재의 기술수준으로 충분히 극복이 가능하며, 나머지 문제를 해결하기 위해서는 파형부호화법을 규칙에 의한 합성에 적용하기 위해 음원피치를 변경시킬 수 있어야 한다. 여기서 피치검출이나 포만트검출은 매우 중요하다. 하지만 음성신호에서는 여파기성분과 여기성분이 상호 작용하기 때문에 피치검출이나 포만트검출이 매우 어렵다. 특히 음성신호에 잡음이 부가될 경우에는 더욱 어려워진다. 따라서 낮은 SNR 조건에서도 피치정보나 포만트 정보를 유지하는 스펙트럼 신호는 음성처리 분야에서 매우 중요하다고 할 수 있다. 그런데 스펙트럼 신호에는 고조파 성분과 포만트 성분이 함께 나타난다. 따라서 이를 잘 분리하는 것이 피치검출이나 포만트검출의 관건이라 할 수 있다.

본 논문에서는 피치 변경시에 발생하는 스펙트럼 왜곡

을 최소화하기 위해 스펙트럼 신호를 최대한 평탄화 시킴으로써 포만트의 영향을 제거하고 고조파 성분을 분리하여 주파수 스케일링에 의하여 피치를 변경하는 방법을 새로이 제안하였다. 제안한 방법은 기존의 스케일링 방법의 정확히 여기스펙트럼을 분리해 내지 못해 발생하는 스펙트럼 왜곡을 보완하기 위하여 스펙트럼 상에서 서브 선형 근사에 의해 고조파 성분을 얻고 스케일링을 이용하여 피치를 변경하는 기법을 적용하였다.

Ⅱ. 기존의 피치 변경법

곡은 합성음에서 버즈음으로 나타나 합성음질의 열하를 초래하게 된다.성분분리형 피치 변경법은 음성신호의 스펙트럼을 여파기 스펙트럼과 여기 스펙트럼으로 분리하여 여기 스펙트럼을 스케일링함으로써 피치 주기를 변경시키는 방법이다. 근사적인 여파기 스펙트럼은 식 2.1에 의해 구할 수 있다.

$$H(K)=\frac{1}{K_0}\sum_{i=0}^{K_0-1}M(K-i) \tag{2.1}$$

여기서 K_0는 기본 주파수에 해당하며 $M(K)$는 음성신호의 진폭 스펙트럼이다. 음성신호의 진폭 스펙트럼에서 근사적인 여파기 스펙트럼을 빼면 평탄화된 여기 스펙트럼을 구할 수 있다.

$$E(K)=M(K)-H(K) \tag{2.2}$$

이 신호에 대하여 주파수영역에서 스케일링함으로 써 기본주파수를 변경한다. 이때 스케일링 율은 시간축 스케일링 율의 역수를 사용하여야 한다.

$$E'(K)=E(K\times\rho^{-1}) \tag{2.3}$$

여기서 ρ^{-1}은 주파수축 스케일링 율이다. 기본 주파수를 높이기 위해서는 고조파의 간격을 ρ^{-1}만큼 늘이고 기본 주파수를 낮추기 위해서는 고조파의 간격은 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 열하를 초래하게 된다

Ⅲ. 새로운 평탄화 기법을 이용한 피치변경법

음성신호는 FFT변환을 통해 주파수 영역에서 스펙트럼 분석이 이루어진다. 그림1은 본 논문에서 사용한 알고리즘의 블럭도이다. 스펙트럼 신호로부터 포만트의 영향과 천이진폭의 영향을 제거하기 위한 첫 단계로서 주파수 대역을 몇 개의 서브밴드로 나눈다. 이때 서브밴드의 대역폭은 스펙트럼 평탄화에 많은 영향을 준다. 본 논문에서는 피치의 범위가 보통 2.5- 25ms인 것을 감안하여 300Hz와 400Hz를 서브밴드의 대역폭으로 사용하였다. 이는 입력음성에 따라 적응적으로 대처하기 위한 것이다. 다음 단계로 각각의 서브밴드에서 최대값을 취하여 프레임의 파라미터로 저장한다. 이 파라미터의 값은 8KHz 샘플링을 했을 경우 10-13개가 된다. 이 값들은 직접 포만트 성분들을 반영하기 때문에 포만트 포락선을 잘 모델링한다고 할 수 있다. 다음은 구해진 파라미터들로 선형보간을 하여 대략적인 포만트 포락선을 얻은 후 스펙트럼 신호로부터 이를 빼주면 제 1차 스펙트럼 평탄화가 되는 것이다. 가장 이상적인 결과는 입력음성의 피치단위로 서브밴드의 대역폭이 결정된 경우에 나타난다. 따라서 제 1차 스펙트럼 평탄화의 결과를 보상하기 위해 평탄화된 신호를 가지고 다시 한번 위의 알고리즘을 거쳐 제 2차 스펙트럼 평탄화를 시킨다. 이때 서브밴드의 대역폭은 각각 3가지 경우의 대역폭을 사용했다. 제 1차 평탄화의 대역폭이 300Hz였을 경우 200Hz, 300Hz, 400Hz를 사용하고 400Hz였을 경우에는

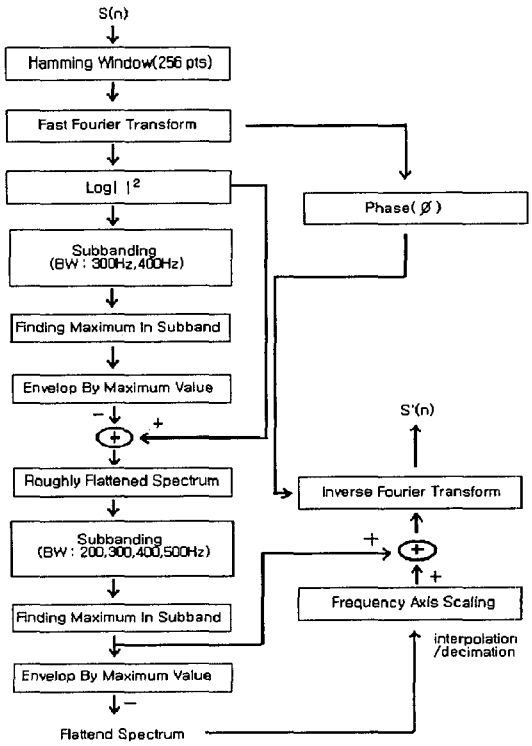

그림1. 제안한 알고리즘

300Hz, 400Hz, 500Hz를 사용했다. 각각의 결과에 대한 비교 평가 방법은 분산을 이용하였다. 분산을 계산하기 전에 각 결과신호들은 최대값이 영이 되도록 정규화 시키고 평균이 영인 분산을 계산하여 분산값이 작은 것을 최종적인 결과로 사용하였다. 본 논문에서 사용한 분산은 다음과 같다.

$$Variance = \frac{2}{N}\sum_{k=1}^{N/2}(x(k)-m)^2 \qquad (3.1)$$

여기서 N은 FFT포인터 수이고 스펙트럼 신호가 Y축으로 대칭이기 때문에 분산은 N/2까지만 이루어진다. 또한 k는 주파수 영역에서의 샘플인덱스이고 m은 평균을 의미한다. 이때 m값은 0을 사용하여 0을 기준으로 평탄화의 정도를 평가하였다. 그림 2는 음성신호의 스펙트럼 신호를 평탄화한 결과이다. 그림에서 알 수 있듯이 LPC법과 Cepstrum법보다는 제안한 알고리즘이 훨씬 우수한 결과를 보이고 있다. 특히 주파수 영역의 양끝부분에서 좋은 성능을 보인다. 이 평탄화된 스펙트럼을 스케일링하여 피치가 변경된 여기 스펙트럼을 구성하여 근사적인 포만트 스펙트럼을 다시 더하여 줌으로써 피치가 변경된 진폭 스펙트럼을 구성하였다. 피치가 변경된 진폭 스펙트럼에 지수함수와 IFFT를 적용하여 피치가 변경된 음성 신호를 재구성하였다. 그림 3은 평탄화된 여기 스펙트럼을 주파수영역에서 Decimation/Interpolation을 이용하여 피치주기를 변경한 결과를 나타내고 있다. 그림에서는 원음성과 기본주파수를 150%로 높인 스펙트럼을 보여주고 있다. 제안된 피치 변경법의 성능평가는 객관적인 평가로서 식3.2를 이용하여 스펙트럼 왜곡율을 측정하였다.

$$SD = \frac{1}{2\pi}\int_{-\pi}^{\pi}[10\log|H(e^{iw})|^2 - 10\log|\hat{H}(e^{iw})|^2]dw \qquad (3.2)$$

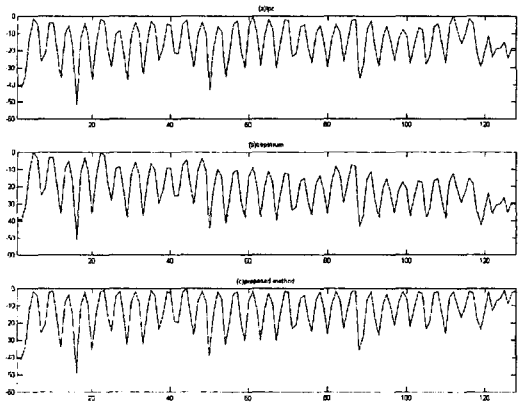

그림2. 평탄화된 스펙트럼 신호
(a) LPC method (b) Cepstrum method
(c) Proposed method

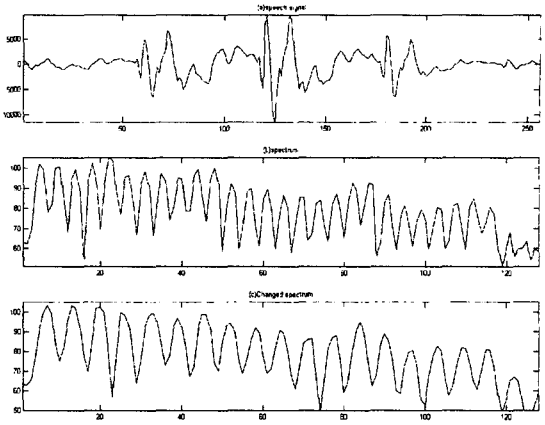

그림3. 기본주파수를 150%로 높힌 스펙트럼
(a) speech signal (b) speech spectrum
(c) Spectrum using Proposed method

Ⅳ. 실험 및 결과

이상의 과정을 컴퓨터 시뮬레이션하기 위하여 IBM 펜티엄에 마이크가 부착된 16-비트 A/D변환기를 인터페이스시키고, 아래의 문장들을 남녀 각 3명에게 발성시키면서 8kHz의 표본화 주파수로 표본화하여 저장한 다음에 시뮬레이션의 시료로 사용하였다:

발성1) "인수네 꼬마는 천재소년을 좋아한다."
발성2) "예수님께서 천지창조의 교훈을 말씀하셨다."
발성3) "창공을 날으는 인간의 도전은 끝이없다."
발성4) "숭실대학교 정보통신과 음성통신 연구팀이다"

표1은 남성화자의 발성별 분산값을 보여주고 있다. 결과값에서 보여지듯이 Cepstrum법이 가장 큰 분산값을 나타내고 LPC법은 양호한 특성을 보이지만 제안한 방법보다 약 1.5배 큰 분산값을 보이고 있다. 표 2는 여성화자의 발성별 분산값 비교인데 역시 제안한 방법이 가장 우수한 결과를 보여주고 있다.

표1. 남성화자의 분산값[dB]

	LPC	Cepstrum	New method
발성1	178.98	759.56	117.08
발성2	157.84	703.28	104.41
발성3	177.56	700.41	114.65
발성4	146.23	694.37	93.44
Average	165.15	714.405	107.39

표2. 여성화자의 분산값[dB]

	LPC	Cepstrum	New method
발성1	303.10	842.59	210.97
발성2	266.82	756.29	172.93
발성3	269.69	718.35	182.25
발성4	237.41	688.99	151.00
Average	269.25	751.55	179.28

피치 변경법의 성능평가는 음성신호의 피치주기를 120%에서 180%까지 변경시키면서 원래음성 신호의 스펙트럼에 비해 나타나는 스펙트럼의 왜곡율을 측정하여 백분율로 환산하여 표 3에 나타내었으며, 스펙트럼의 비교기준은 피치가 변경되기 이전의 원래 음성의 스펙트럼을 사용하였다. 피치를 변경시키면 원래의 스펙트럼과 직접 비교할 수 없기 때문에 피치주기를 120%, 150%, 180%로 각각 신장시킨 다음에 83%, 66%, 55%로 각각 압축하여 원래의 음성 스펙트럼과 고조파를 일치시킨 다음에 에너지 왜곡율을 측정하였다. 표 3과 같이 평균 스펙트럼 왜곡율은 기존의 성분분리형 피치 변경법의 2.31%에서 제안한 방법이 2.12%로 0.19%가 개선되었다.

표3. 피치변경율에 따른 스펙트럼 왜곡율 비교

변경법 / 변경율	기존의 방법			제안한 방법		
	남자	여성	평균(%)	남성	여성	평균(%)
120%	1.67	2.03	1.85	1.51	1.83	1.67
150%	2.12	2.45	2.28	1.84	2.25	2.04
180%	2.68	2.95	2.81	2.52	2.79	2.65
Average	2.15	2.47	2.31	1.95	2.29	2.12

V. 결론

음성합성은 합성단위에 따라서 문장단위, 음절단위, 음소단위 등으로 나눌 수 있다. 또한 합성방식에 따라서는 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다. 파형부호화법은 파형 자체의 잉여성분을 제거한 후에 부호화하는 방법으로 음질이 우수하다. 그러나 파형부호화법은 음성신호를 인간의 개성과 감정을 대별해 주는 여기 정보와 메시지전달을 나타내는 여파기 정보로 분리하지 않고 처리하기 때문에 음원을 변경시켜야하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다. 고음질 음성합성을 위해서는 파형부호화법이 바람직하다. 그렇지만 파형부호화법을 사용하면 음원피치의 변경이 어렵다는 문제점이 발생한다. 이러한 문제의 해결을 위해서는 파형부호화법을 규칙에 의한 합성에 적용하기 위해 음원피치를 변경시킬 수 있어야 한다. 여기서 피치검출이나 포만트검출은 매우 중요하다. 하지만 음성신호에서는 여파기성분과 여기성분이 상호작용하기 때문에 피치검출이나 포만트검출이 매우 어렵다. 특히 음성신호에 잡음이 부가될 경우에는 더욱 어려워진다. 따라서 낮은 SNR 조건에서도 피치정보나 포만트 정보를 유지하는 스펙트럼 신호는 음성처리 분야에서 매우 중요하다고 할 수 있다. 그런데 스펙트럼 신호에는 고조파 성분과 포만트 성분이 함께 나타난다. 따라서 이를 잘 분리하는 것이 피치검출이나 포만트검출의 관건이라 할 수 있다.

본 논문에서는 피치 변경시에 발생하는 스펙트럼 왜곡을 최소화하기 위해 스펙트럼 신호를 최대한 평탄화 시킴으로써 포만트의 영향을 제거하고 고조파 성분을 분리하여 주파수 스케일링에 의하여 피치를 변경하는 방법을 새로이 제안하였다. 제안한 방법의 성능을 평가하기 위해 스펙트럼 왜곡율을 측정하여 본 결과 평균 스펙트럼 왜곡율은 평균 2.12%이하로 유지되었으며 실험결과 제안한 방법이 기존의 방법보다 우수함을 보여주었다.

본 연구는 한국과학재단의 특정기초연구(과제번호 R01-2002-000-00278-0)의 지원에 의하여 이루어졌음.

참고 문헌

[1] Thomas W.Parsons, Voice and Speech Processing, McGraw-Hill, 1986.

[2] A.M. Kondoz, Digital Speech Coding for Low Bit Rate Communications Systems, John Wiley & Sons, 1994.

[3] L. R. Rabiner and R. W. Schafer, Digital Processing of Speech signals, Englewood Cliffs, Prentice-Hall, New Jersey, 1978.

[4] Panos E. Papamichalis, Practical Approaches to Speech Coding, Prentice-Hall, 1987.

[5] S. Seneff, "Real Time Harmonic Pitch Detection," IEEE Trans. Acoust. Speech, and Signal Processing, Vol. ASSP-26, pp. 358-365, Aug. 1978.

[6] S. D. Stearns & R.A. David, Signal Processing Algorithms, Prentice-Hall, Inc, Englewood Cliffs, New-Jersey, 1988.

[7] M. Bae, and S. Ann, "Fundamental Frequency Estimation of Noise Corrupted Speech Signals Using the Spectrum Comparison," J., Acoust., Soc., Korea, Vol. 8, No. 3, June 1989.

[8] M. Lee, C. Park, M. Bae, and S. Ann "The High Speed Pitch Extraction of Speech Signals Using the Area Comparison Method," KIEE, Korea, Vol. 22, No. 2, pp.13-17, March 1985.