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

시간 - 주파수 변환에 의한 파형 대칭 피치변경법

박형빈, 김종득, 하태성, 배명진

숭실대학교 정보통신공학과

On A Pitch Alteration using the Waveform Symmetry with Time - Frequency Conversion

HyungBin Park, JongDeuk Kim, TaeSeong Ha, MyungJin Bae

Dept. of Telecom. Engr. Soongsil Univ.

hbpark@assp.soongsil.ac.kr, jdkim@assp.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

Abstract

In the case of speech synthesis, the waveform coding method with high quality is mainly used to the synthesis by analysis. Because the parameters of this coding method are not classified as both excitation and vocal tract parameters, it is difficult to apply the waveform coding method to the synthesis by rule. Thus, in order to apply the waveform coding method to the synthesis by rule, a pitch alteration is required for the prosody control. In the speech synthesis method by the conventional PSOLA technique, applying symmetric window function to asymmetric speech waveform, it occurs the unbalance phenomenon of energy according to the overlapped degree of pitch interval adjustment.

In this paper to overcome the unbalance phenomenon of energy, we proposed a new method that can convert asymmetric waveform to symmetric one by time-frequency conversion. As a result, we can obtain an average spectrum distortion ratio with 6.38% according to the pitch alteration ratio.

Ⅰ. 서론

음성 합성분야에서 합성단위에 따라서는 문장단위, 음절단위, 음소단위 등의 합성법으로 나눌 수 있다. 또한 합성방식에 따라서는 파형부호화법, 신호원부호화법, 혼성부호화법으로 분류할 수 있다[1]. 최근 다양해진 음성서비스 분야에서는 고음질의 합성음을 요구하고 있다. 이러한 고음질 합성방식으로는 파형부호화법이 바람직하다. 그렇지만 파형부호화법을 사용하면 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.

지금까지 제안된 피치변경법은 그 처리영역에 따라 시간, 주파수, 시간-주파수 혼성처리법이 있다.

시간영역법에는 Multi-Pulse법, 피치반분법 등이 있다. Caspers와 Atal은 MPLPC에서 펄스사이에 영을 삽입하거나 삭제하는 방법을 제안했으나, MPLPC상의 펄스 열은 피치와 포먼트에 대한 상호 연관을 갖고 있으므로 스펙트럼의 왜곡이 심하다[3]. Varga와 Fallside는 LPC계수를 이용한 피치연장법을 제안했으나, 이 방법은 피치주기를 줄이는 경우에는 단지 파형의 일부분을 소거하고 평활화하는 방법을 사용하고 있기 때문에 스펙트럼의 왜곡이 많이 나타난다[4]. 피치반분법은 임의로 변경하려는 피치주기의 2배 파형을 만든 후에 그 파형의 주기를 반분하는 피치 변경법이다[6].

Quatieri와 McAulay는 주파수영역에서 위상을 보존하는 피치변경법을 제안하였는데 이것은 입력된 음성에 대해 진폭 및 위상스펙트럼을 추출하여 별도로 처리하는 방법이다. 진폭스펙트럼에 대해서는 두드러진 스펙트럼 봉우리들을 추출한 다음에 이것을 피치변경율(ρ) 만큼 인터폴레이션하여 진폭스펙트럼의 피치를 변경시킨다. 위상스펙트럼에 대해서는 시간영역에서 구한 피

치 개시시간(pitch onset time)에 해당하는 위상을 제거하고 나서 피치가 변경되었을 때의 새로운 피치 개시시간의 위상을 더해줌으로서 새로운 위상을 구성하게 된다. 그렇지만 피치 변경 시에 피치주기와는 별도로 피치의 개시시간을 공급해 주어야 하고, 또한 진폭 스펙트럼 상에서 두드러진 봉우리 위주로 고조파의 인터폴레이션을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

시간-주파수 혼성법으로는 켑스트럼의 특징을 이용하여 켑스트럼값이 거의 영이 되는 부분에서 영값을 삽입하거나 삭제하므로써 피치를 변경하는 방법이 있다[5]. 그러나 이 방법 역시 위상의 보존이 어렵다는 문제점을 가지고 있다. Takagi와 Miyasaka가 제안한 시간-주파수혼성법[7]은 시간영역에서 피치변경을 하였을 때에 나타나는 스펙트럼왜곡을 스펙트럼영역상에서 LPC 포락을 통해 수정하는 방법이다. 이 방법은 LPC스펙트럼 포락이 갖는 극점에 치중된 시스템 전달 특성 때문에 모든 유성음에 적합하지 못하다는 한계성을 갖는다. 피치변경시에 포만트 스펙트럼이 왜곡되면 성도의 여파기 정보가 변경되므로 의사정보를 제대로 보존할 수 없고, 위상이 왜곡되면 인근 프레임간 진폭레벨의 변동이 커져서 음소간의 연결이 부자연스럽게 된다. 이러한 현상은 원래 음성의 피치주기를 중심으로하여 피치 변경되는 율이 증가함에 따라 더 큰 왜곡이 발생하게 된다.

II. PSOLA 합성 기법

기존의 PSOLA 합성방식은 먼저 원래의 음성 파형을 피치주기 단위로 분해한 다음 분해된 피치 단위에 윈도우 함수를 곱해서 ST(Short-Term)신호의 열로 만든다. 분해된 단위의 운율조절을 하고 이렇게 조절된 단위로부터 음성을 합성한다.

1) 분 석

원래 음성 파형이 유성음인 경우에는 피치단위로 분해한 다음 윈도우 함수를 곱하여 ST(Short-Term)신호의 열로 만든다. 무성음인 경우에는 10ms의 주기로 일정하게 분석한다. 분석 윈도우 함수에는 다음과 같은 Hanning, Hamming, Blackman 등의 형이 쓰인다.

Hanning Window :

$$W(n)=\frac{1}{2}\left\{1-COS(2\pi\frac{n}{N-1})\right\} \quad ,0\le n\le N-1$$

Hamming Window :

$$W(n)=0.54-0.46COS(2\pi\frac{n}{N-1}) \quad ,0\le n\le N-1$$

Blackman Window :

$$W(n)=0.42-0.5COS(2\pi\frac{n}{N-1}) +0.08COS(2\pi\frac{2n}{N-1}) \quad ,0\le n\le N-1$$

[N : 윈도우 함수에 곱해지는 음성 샘플수]

이런 윈도우 함수를 원래의 음성 샘플에 곱하므로써 다음 (1)식과 같은 피치단위로 분해된 샘플열들을 얻는다.

$$S_{analysis}(n)= W_{analysis}(m - n)S(n) \quad (1)$$

$S_{analysis}(n)$: 피치주기 단위의 ST(Short-Term)신호

$W_{analysis}(n)$: 분석 윈도우 함수

m : m번째 피치

$S(n)$: 원 음성 파형

2) 운율 조절 및 합성

분석과정에서의 ST(Short-Term)신호의 열은 원래의 음성 샘플의 피치 단위로 배열되어 있다. 따라서 피치를 변경하기 위해서는 이 간격들을 변경할 피치 간격들로 재배열하면 된다. 다음 (2)식은 피치가 변경된 신호를 나타낸 것이다.

$$S_{synthesis}(n)= S_{analysis}(n - m_a) \quad (2)$$

$S_{synthesis}(n)$: 피치가 변경된 ST(Short-Term) 신호

m_a : 변경할 피치 간격

따라서 피치를 높일때는 ST(Short-Term)신호의 간격을 작게 배열하고 피치를 낮출때는 ST(Short-Term) 신호의 간격을 크게 배열하면 된다. 하지만 이런 순차적인 배열사이에서 정확한 피치 동기화를 유지하는 것이 중요하다. 이렇게 재배열된 ST(Short-Term)신호에서 겹쳐지는 부분은 더해주면 된다.

다음 그림1은 기존의 PSOLA방법을 사용해서 피치를 150% 신장시킨 경우와 본 논문에서 제안한 방법으로 피치를 150% 신장시킨 경우를 나타낸 것이다.

하지만 이러한 기존의 방법은 음성 파형이 비대칭적인 경우에도 대칭적인 윈도우 함수를 적용한다. 따라서 피치 간격을 조절할 때 겹쳐지는 정도에 따라 에너지의 불균형 현상이 생긴다. 그러므로 에너지를 일정하게 하기 위한 정규화 과정이 필요하다.

본 논문에서는 피치를 변경할 때 이러한 에너지 정규화 방법으로 시간영역에서의 비대칭적인 음성 파형을 시간-주파수 변환과정을 통해서 기존의 PSOLA방법에 적합한 대칭적인 파형으로 바꾸어서 적용하였다.

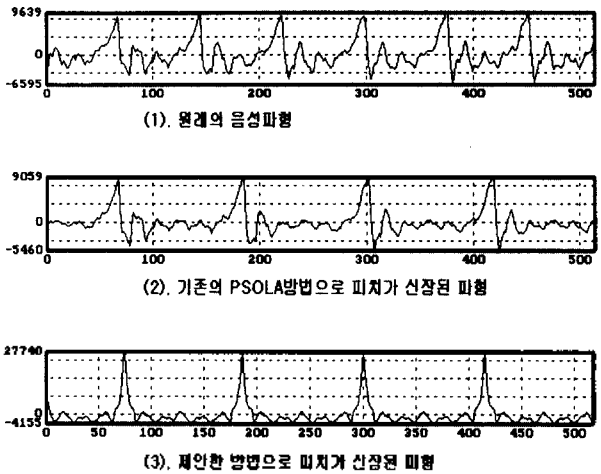

그림1. 피치주기를 신장한 경우의 파형

III. 피치동기된 파형 대칭에 의한 피치변경법

앞 절에서 언급했듯이 기존의 PSOLA기법에서 피치주기단위로 곱해지는 윈도우 함수타입은 보통 대칭적인 Hanning이나 Hamming 형이다. 하지만 기존의 PSOLA 방법에서는 음성 파형이 비대칭적인 경우에도 이러한 윈도우 타입을 그대로 사용한다. 그로인해 피치를 변경할 때 즉 피치 간격의 조절로 인해서 겹쳐지는 정도에 따라 에너지 불균형 현상이 생긴다. 이러한 단점을 극복하기 위해 본 논문에서는 시간영역에서의 비대칭적인 음성 파형을 기존의 PSOLA방법에 적합한 대칭적인 파형으로 바꾸었다. 다음 그림2는 본 논문에서 제안한 원래의 비대칭적인 파형을 시간-주파수 변환에 의해서 대칭적인 파형으로 바꾼 경우를 나타낸 것이다.

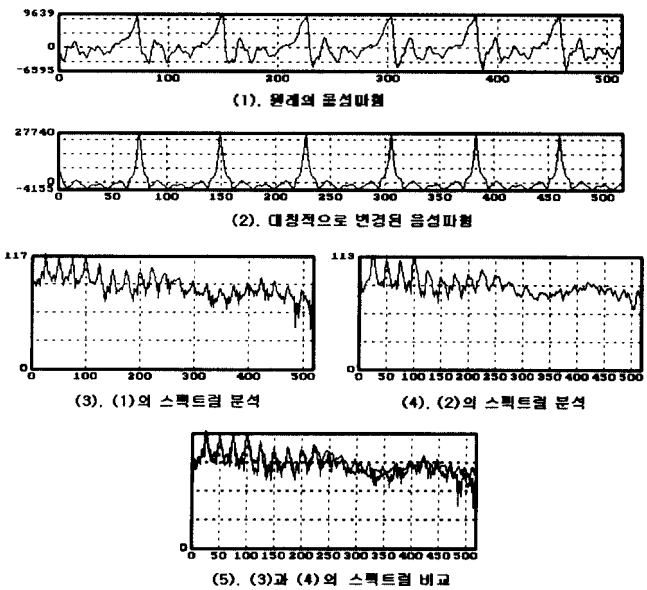

그림 2. 피치동기된 대칭 파형과 스펙트럼

본 논문에서 제안한 피치동기된 대칭 파형은 원래 음성 파형의 위상성분을 0으로 놓고 처리한 것이다. 시간영역상에서 보면 원래의 음성 파형과 유사도가 적어 보인다. 하지만 원래 음성 파형의 진폭성분만을 가지고 복원하였기 때문에 그림2에서도 알 수 있듯이 원래 음성 파형의 스펙트럼과 크게 유사하다는 것을 알 수 있다. 또한 성도의 여파기 성분이 거의 일치하기 때문에 명료성의 열하가 적다.

IV. 피치검출법

신호원 부호화법과는 달리 파형 부호화법에서 피치를 변경하려면 사전에 그 발성자의 피치변화를 알고 있어야 한다. 이것은 발성자의 억양이나 감정의 변화가 중심된 피치를 기준으로 하여 피치의 상대적인 변화로 나타나기 때문이다. 특히 파형 부호화에서는 발성자의 개성과 메시지 정보를 보존하여 음질의 명료성이 우수하다. 이 때문에 피치변경시에는 발성자가 주로 사용하는 피치주기를 기준으로 피치를 변경시킬 필요가 있다. 따라서 정확한 피치검출이 선행되어야 한다.

음성신호의 피치는 음성 파형의 반복되는 봉우리에서 봉우리까지 또는 골에서 골까지로 정의된다. 눈으로 파형을 보고 피치를 찾을 때는 두드러진 파형봉우리의 반복구조에 주로 관심을 가지게 된다. 음성 파형에서 피치 주기구간의 첫 봉우리인 G-Peak를 찾을 수 있다면 다음 G-Peak까지의 간격이 피치가 된다.

지금까지 제안된 피치검출법은 크게 시간영역법, 주파수영역법, 그리고 시간-주파수 혼성영역법으로 나눌 수 있다[1]. 본 논문에서는 피치검출법으로 시간영역에서의 면적비교법을 적용하였다. 그렇지만 합성을 위해 파형을 편집하는 경우에는 피치의 추출이 반드시 자동화될 필요는 없으며, 면적비교법[9]과 함께 눈으로 피치를 추출하는 반자동법이나 눈으로 찾는 수동법으로 처리하여도 된다.

V. 실험 및 결과

본 논문에서 제안한 방법을 시뮬레이션 하기 위해 IBM-PC/586에 마이크 입력이 가능한 16비트 A/D변환기를 인터페이스하여 3명의 남성과 2명의 여성화자를 통해 다음 음성시료를 발성하게 하고 이를 11kHz의 표본화율로 16비트 양자화하여 저장하였다.

발성 1: /인수네 꼬마는 천재소년을 좋아한다./
발성 2: /숭실대 정보통신과 음성통신연구팀이다./
발성 3: /공일이삼사오육칠팔구/

다음 그림 3은 본 논문에서 제안한 방법을 블록도로 나타낸 것이다.

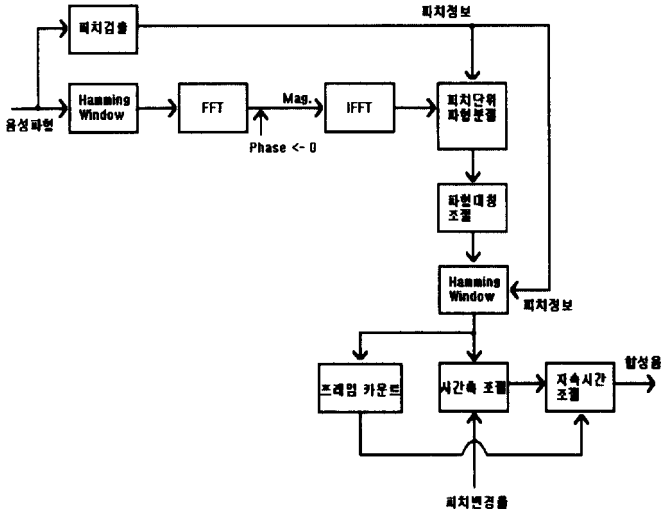

그림3. 본 논문에서 제안한 방법의 블록도

시뮬레이션은 한 프레임의 길이를 256표본으로 사용하였다. 먼저 면적비교법[9]을 사용하여 한 프레임 구간에서 피치를 검출하였다. 입력된 음성 파형을 주파수영역으로 변환한 다음 위상성분을 0으로 하고 진폭성분만을 가지고 다시 시간영역으로 변환한다. 그런 다음 피치주기 단위로 파형을 분절한 다음 2배의 피치주기를 가지도록 파형을 대칭적으로 만들고 시간축을 조절하여 피치주기에 일치하도록 한다. 피치 변경율에 따라서 지속시간을 조절하여 피치를 변경한다.

피치신장시에 나타나는 스펙트럼의 왜곡을 측정하기 위해서 피치주기를 100%에서 200%까지 변경시켰을 때 원래 스펙트럼에 비해 나타나는 스펙트럼의 왜곡율을 측정하였다. 스펙트럼의 기준은 피치변경되기 이전의 원래 음성의 스펙트럼이다. 본 논문에서 제안한 방법을 통해 피치주기를 신장시켰을 때 평균 스펙트럼 왜곡율은 6.38%로 측정되었다.

VI. 결 론

본 논문에서는 기존의 PSOLA방법에서 비대칭적인 음성 파형에 대칭적인 윈도우 함수를 적용함으로써 야기되는 에너지의 불균형 현상을 해결하는 방법을 제안하였다. 비대칭적인 파형을 시간-주파수 변환에 의해서 대칭적인 파형으로 바꿈으로써 문제점을 해결하였다. 또한 본 논문에서 제안한 대칭적인 파형은 스펙트럼 왜곡율이 피치변경율에 비례적으로 나타나지 않도록 2배의 피치주기를 갖는다.

결과적으로 제안한 방법에서 피치주기를 100%에서 200%까지 변경하였을 때 평균 스펙트럼 왜곡율이 6.38%정도로 비교적 우수한 결과를 얻을 수 있었다.

본 연구는 정보통신부의 1998년도 대학기초과제 연구지원비에 의해 이루어졌습니다.

VII. 참고문헌

[1] L.R Rabiner & R.W. Schafer, Digital Processing of speech Signals, Prentice-Hall, 1978.

[2] M.R. Portnoff, "Time-Scale Modification of speech Based on Short-Time Fourier Analysis," IEEE, Trans, Acoust. Speech, Signal Processing, Vol.ASSP-29, No.3, pp.374-390, June 1981.

[3] M.G. Stalla and F.J. Charpentier, "Diphon Synthesis using Multipulse Coding and a Phase Vocoder",Proc. IEEE ICASSP'85, PP.740~744, 1985.

[4] 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, APRIL 1987.

[5] 배명진, 이미숙, 이해군, 안수길, "켑스트럼 분석에 의한 음성 파형코딩의 피치변경에 관한 연구", 제 4 회 신호처리합동 학술대회 논문집, 제 4 권 1호, pp.304~309, 1991년 9월.

[6] 강동규, 김율제, 배명진, 안수길, "음성합성의 halving기법에 의한 파형코딩의 피치 변경에 관한 연구", 한국 음향학회 추계발표회(국제 음향학회 논문집), pp.107~111, 1990년 11월 10일.

[6] 강동규, 김율제, 배명진, 안수길, "음성합성의 halving 기법에 의한 파형코딩의 피치 변경에 관한 연구", 한국 음향학회 추계발표회(국제 음향학회 논문집), pp.107~111, 1990년 11월

[7] T. Takigi and E. Miyasska, "A Speech Prosody Conversion System with a High Quality Speech Analysis-Synthesis Method",Proc.EUROSPEECH '93, pp.995 ~ 998, September 1993.

[8] 배명진, "위상보상된 고조파 스케일링에 의한 음향합성용 피치변경법", 한국음향학회, 한국음향학회지, 제13권 6호, pp.995 ~ 998, September 1993.

[9] 배명진, 안수길, "면적 비교법을 이용한 음성신호의 고속 피치 추출", 전자공학회지, 제22권, 2호, pp.13 ~ 17, 1985년 3월.

[10] 손상목, 배재옥, 김용, 배명진, 기석철, 김상용, "진폭대칭법에의한 배주기 피치변경법", 한국통신학회, 제 8 회 신호처리합동학술대회 논문집, 제 8 권, 1호, pp.708~711, 1995년 9월.