피치 변경법의 성능평가

(*) 금 홍 (**) 배성근 (*) 조왕래 (*) 배명진
(*) 숭실대학교 정보통신공학과 (**) 건국대학교 전자공학과

On a Performance Evaluation of the Pitch Alteration Techniques of speech waveform coding

(*) Hong Keum (**) Seonggyun Bae (*) Wangrae Jo (*) Myungjin Bae
(*) Soongsil University (**) Kon-KuK University

ABSTRACT

Generally we are used to apply waveform coding method obtaining the high quality synthesized speech. But we have to solve the problems, memory capacity and pitch alteration, for applying the waveform coding method to speech synthesis by rule. The former problem is conquered by improving the integrated semiconductor technology, but the latter problem remains. In this paper, we compare the methods that have proposed for pitch alteration in our laboratory until now. These methods are not change properties of vocal tract formants and only altered the pitch. The reference of the comparison is spectrum distortion. We obtain the spectrum distortion, 13.94% for pitch halving method, 1.14% for cepstrum analysis method, and 2.36% for harmonics compensated with the phase method.

I. 피치변경의 필요성

음성합성시스템은 그 기준에 따라 여러가지로 분류할 수 있는데 먼저 분석된 데이타를 그대로 합성에 이용하는 분석에 의한 합성(synthesis by analysis)법과 규칙에 의해 음성을 발생시켜 합성하는 규칙에 의한 합성(synthesis by rule)법으로 구분할 수 있다[1].

합성을 위해 출력되는 데이타의 처리방식에 따라서는 메모리형 합성법과 전송형 합성법으로 구분할 수 있다. 메모리형 합성법은 분석한 데이타를 메모리에 저장시켜 두고 필요에 따라 다시 합성시키는 방법으로, 모저(Mozer)법, 피치단위분절법 등이 있다[2][4].

한편 전송형 합성법은 전송채널의 효율을 높이기 위해 분석후 바로 전송을 목표로하여 처리하고 수신측에서 다시 조합하여 합성하는 방법이다. 이러한 합성법은 공중전화망의 보코더(voice coder decoder)시스템에 오랫동안 응용되어 왔다[5-6]. 지금까지 전송형 합성법으로 연구된 방법들은 크게 파형부호화, 신호원부호화, 혼성부호화법이 있으며[2][7], 저장메모리나 전송채널의 대역폭절약을 위해서는 신호원부호화법을, 음질을 높이기 위해서는 파형부호화법을 주로 사용하고 있다.

고음질의 합성음성을 얻기 위해 사용하는 파형부호화법은 음성정보를 발성모델에 따라 분리하지 않고 파형자체의 잔여성분을 제거한 후에 부호화하는 방법이며 PCM, ADPCM, ADM 등이 제안되어 있다[8]. 그렇지만 파형부호화법은 인간의 개성과 감정을 나타내는 여기(excitation)정보와 의사잡담을 나타내는 성도특성과의 포먼트(formants)정보를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 규칙에 의한 합성기법으로는 바람직하지 못하다.

대부분의 음성합성 응용분야에서는 합성단위를 한가지로 고정해서 사용하지 않고 몇가지 합성기법들을 혼성하여 적용하고 있다. 이러한 경우, 보통 문장단위의 합성에는 명료성을 유지하기 위해 파형부호화법이나 혼성부호화법을, 그리고 단어나 음절단위의 합성에는 규칙합성이 가능하도록 신호원부호화법을 사용해야 한다[1]. 이 때문에 간단한 응용에 대해서도 동일한 데이타베이스를 사용할 수 없게 된다. 또한 이러한 경우에 파형부호화법과 신호원부호화법이 연결되는 부분에서는 파형부호화의 음원피치가 자연스럽게 연결되지 못하고, 신호원부호화에서는 합성시에 수렴속도가 요구되기 때문에, 자연성과 명료성이 저하될 수도 있다. 이러한 문제점을 해결하려면 파형부호화법에 의해 규칙합성이 가능하도록 파형부호화법에서 음원피치를 변경시킬 수 있어야 한다.

따라서 본 논문에서는 고음질의 합성음을 얻기 위하여 본 연구실에서 제안하였던 세가지 피치변경법들을 비교평가하였다. 이들 피치변경법들은 파형부호화법에 기초하여 규칙합성이 가능하도록 음원의 피치를 변경하는 기법들이다. 비교의 대상이 되는 방법 및 기술 순서는 다음과 같다. 먼저, 본 연구실에서 피치주기 변경법을 제안하게 된 배경을 설명하고, 파형의 주기를 간단히 판분하는 주기판분법을 적용하여 시간영역상에서 피치주기를 변경하는 피치주기 반복법과 시간-주파수 영역의 변환특성을 이용한 켑스트럼 피치변경법 및 시간 영역에서 위상을 보상하고 주파수영역에서 피치를 변경하는 위상 보상된 고조파를 이용한 피치변경법을 비교할 것이다. 끝으로, 실험결과를 기술하고 각 방법의 장단점을 이야기하고 결론을 맺고자 한다.

II. 기존의 피치주기 변경법

일반적인 선형예측시스템에서는 여기 임펄스들의 시간간격을 변경함으로써 음성음의 피치가 간단히 변경되고, 지속시간(duration)은 예측계수의 갱신율을 변경시킴으로써 변경된다. Caspers와 Atal은 이러한 선형예측기법의 관점에서, 멀티펄스 여기원의 피치주기를 변경하기 위

해 영-주기나 삭제 기법을 적용하였다. 또한 저속시간은 피치주기들을 추가 또는 제거하는 방법으로 변경하였다[10]. 그렇지만 혼성부호화법인 멀티펄스-LPC법에서는 각 임펄스들이 이전에 찾아진 모든 임펄스들의 효과를 고려하여 최소의 오류가 되도록 계산되기 때문에 임펄스들 사이의 간격조절에 의해 피치가 간단히 변경될 수 없다. 임펄스 위치의 변경은 합성파형의 왜곡을 초래할 뿐이다.

Stella와 Charpentier는 멀티펄스 여기원을 정도모델과 여기원의 한 결합으로 보는 관점에서, 이러한 문제를 해결하였다. 그들은 우선 멀티펄스 분석법으로 다이폰(diphone) 단위의 데이타 압축을 하였고, 합성시에는 원래형태로 다이폰 합성을 한 다음, 위상(phase) 보코더를 적용하여 운율을 변경하였다[11].

Varga와 Fallside는 파형 편집부분과 선형예측에 의한 합성부분으로 구성된 혼성기법을 적용하여 파형부호화의 피치변경을 시도하였다[12]. 피치주기를 연장하기 위해 음성표본의 일부분과 선형예측이파기 계수를 사용하여 일반적인 합성여파기에 의해 합성하고 연결부분을 평활화 하였다. 그러나 피치주기를 줄이는 경우에는 단순이 한 피치구간 파형의 일부분을 제거하고 평활화하는 것으로 피치변경을 시도하였기 때문에 피치가 짧은 여성이나 어린 화자의 경우에는 스펙트럼 왜곡이 많아진다.

Quatier와 McAulay는 주파수 영역에서 위상을 보존하는 피치변경법을 제안하였는데[13] 이것은 입력된 음성에 대해 진폭 및 위상 스펙트럼을 추출하여 별도로 처리하는 방법이다. 이러한 방법은 파형의 특을 그대로 유지하기 때문에 프레임단위로 분석처리하는 통상의 처리법에서 인근 프레임간의 연결이 아주 용이해진다는 장점이 있다. 그렇지만 피치변경시에 피치주기와는 별도로 피치의 재시시간을 공급해 주어야 하고, 또한 진폭 스펙트럼 상에서 부드러운 봉우리 윤곽을 하모닉스의 인터폴레이션을 수행하기 때문에 스펙트럼의 왜곡이 높아진다는 단점이 있다.

III. 본 연구실에서 제안한 피치변경법들

첫째, 피치반분(pitch halving)법은 시간영역상에서 음성신호의 피치를 반분하는 방법으로 음성의 발성모델에 근거하여 인위적으로 변경하려는 피치주기의 2배파형을 선형예측합성법에 의해 만든 다음에 그 파형의 주기를 반분하는 기법이다[14-15]. 주어진 음성음에 대해 기본주파수를 2배로 늘리는 것은 시간영역에서 음성음의 피치를 반분(halving)하는 것이 된다. 시간-주파수관계에 따라 음성음의 파형 s(n)에서 피치 p-를 반분해 보면 다음과 같다.

$$s'(n) = s(n) \times \sum_{i=0}^{P} \delta(n-ip/2) \qquad (3\text{-}1)$$

여기서 유성음의 피치는 p=1/Fo이고 다른 피치검출법들을 통해 이미 피치에 대해 알고 있다고 가정한다. 또한 음성음 s(n)은 피치단위로 주기함수이므로 식(3-1)을 다시 쓰면 다음과 같이 간략화될 수 있다

$$s'(n) = s(n)+s(n-p/2)+s(n-p)+s(n-1.5p)+ \quad + = P[s(n)+s(n-p/2)] \qquad (3\text{-}2)$$

여기서 P-는 현재의 음성프레임에서 피치주기 반복회수이다.

둘째, 켑스트럼 분석에 의한 피치변경법은 quefrency의 낮은쪽 부분이 성도의 특성을 나타내면서 켑스트럼의 진폭이 빨리 감소하고, 또한 피치주기 가까이에서는 켑스트럼진폭이 서의 영이 되는 성질을 이용하여 피치를 변경하는 것이다[16-17]. 따라서 변경하려는 피치주기만큼의 영값을 quefrency상의 서의 영이 되는 부분에 삽입 또는 삭제하는 기법을 적용하였다.

켑스트럼 신호원부호화법에 의해 음성을 합성하고자 할 때는 quefrency상의 낮은쪽 시간부분을 여과시켜보고 취한 다음 역켑스트럼을 통해 임펄스응답을 구한다. 그리고는 음원정보에 해당하는 피치펄스와 잔별부신을 통해 음성신호를 합성하게 된다. 이렇게 하면 부호화 효율을 높일 수 있게 되어 낮은 전송율에서도 양질의 합성이 가능하고, 임펄스의 실제 컨볼루션되는 피치주기를 변경하여 운율을 조절할 수 있기 때문에, 규칙에 의한 합성법에 적용할 수 있다.

quefrency상에 영값을 삽입 또는 삭제하여 피치주기를 변경시키고자 할 때, 현재 분석중인 음성의 피치를 사전에 알고 있다면 영값대치를 피치주기 가까이에서 하는 것이 바람직하다. 그러나 분석중인 음성의 창함수내에서 시간에 따른 음원피치의 변화가 선형중인 경우에는 피치주기 근방의 켑스트럼 봉우리가 일정폭을 유지하는 것이 보통이다. 따라서 피치주기변경을 위해 영값을 대치하는 위치는 영-quefrency에서는 멀리 떨어지게 하면서도 피치펄스에 너무 근접시키지 않는 것이 바람직하다.

켑스트럼에 의한 피치주기 변경법은 시간과 주파수영역을 왕복하면서 처리하기 때문에 계산이 복잡하고, 피치변경시에 켑스트럼상에서 피치펄스의 위치를 시프트시키기 때문에 이것을 스펙트럼영역으로 변환했을 때 그 위상정보를 잃어버리게 된다. 이것은 단순히 복소켑스트럼을 사용한다고 하여도 해결되지 않는다. 왜냐하면 피치주기의 변경은 스펙트럼의 고조파의 위치를 변경시키기 때문이다. 따라서 위상을 복원하는 문제가 남게된다[18].

셋째, 시간영역에서 상위 포만트의 영향을 제거하고 여기원의 위상을 추출하기 위해, 주어진 음성신호를 한 피치주기를 통과대역폭로 갖는 시여통과여파기에 통과시킨다. 이 신호를 시간축 스케일링하고 FFT를 취하여 위상성분을 얻는다.

한편, 주어진 음성신호를 주파수 영역으로 변환하여 진폭성분을 분리한 후, 대역폭이 F0인 리프터에 통과시켜 포만트 정보를 얻는다. 앞서 분리해 둔 진폭성분으로부터 포만트 성분을 제거하여 여기원에 해당하는 잔여신호를 스케일링율을 적용하여 기본주파수를 변경시킨다. 이 신호와 포만트 정보를 합하면 기본주파수가 변경된 진폭스펙트럼이 얻어진다.

시간축 스케일링된 위상정보와 기본주파수가 변경된 진폭스펙트럼을 IFFT하면 원하는 합성음성파형을 얻을 수 있다.

이와 같이 시간영역에서 피치변경된 신호를 주파수영역으로 변환시킨 후에, 이에 대한 위상을 검출하여, 켑스트럼 피치변경법에 의해 변경된 진폭스펙트럼과 결합시킴으로써 피치가 변경된 합성음을 얻는 방법을 위상 보상된 고조파를 이용한 피치변경법이라 한다[19].

IV. 실험과정 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC/486DX II(50)이며, 여기에 음성신호를 입출력하기 위한 AD/DA 변환기를 인터페이스하여 8 kHz의 표본률로 표본당 16비트의 데이타를 입력하였다. 이상에서 설명한 방법들의 성능을 평가하기 위해 아래와 같은 대표적인 문장을 20세의 여성, 22세의 남성, 26세의 남성 화자가 각각 5 번씩 발성하여 시료로 사용하였다.

발성 1 : / 인수네 포마는 천재소년을 좋아한다./

발성 2 : / 예수님께서 천지창조의 교훈을 말씀하셨다./

발성 3 : / 지금 서신 전화는 /

전체 실험과정은 시간영역에서 면적비교법을 이용하여 피치를 구한 후 한 프레임을 256샘플 단위로 처리하였다. 아래의 그림 4-1, 4-2, 4-3은 각 피치변경법의 처리 블럭도이다.

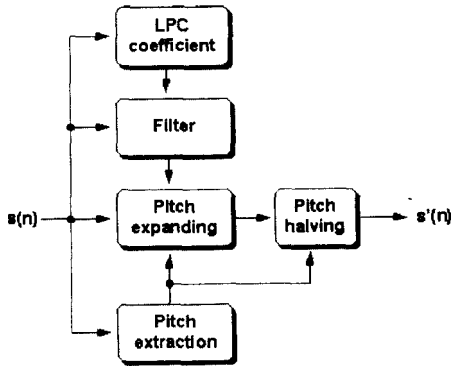

그림 4-1. 피치주기 반분법의 블럭도.

Fig. 4-1. Block diagram of pitch halving method.

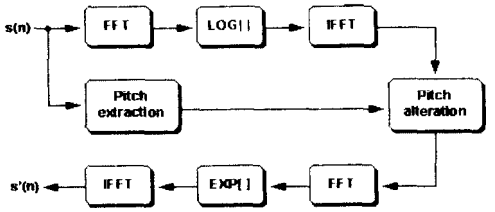

그림 4-2. 캡스트럼 분석에 의한 피치변경법.

Fig. 4-2. Pitch alteration method by using the cepstrum analysis.

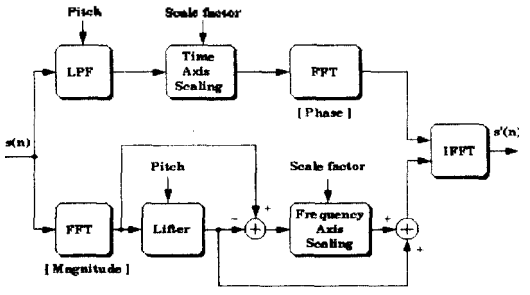

그림 4-3. 위상보상된 고조파를 이용한 피치 변경법의 블럭도.

Fig. 4-3. Block diagram for pitch alteration method by using the harmonics compunsated with phase.

그림 4-1에 나타낸 것처럼 피치주기 반분법을 사용하는 경우는 음성파형으로부터 먼저 피치와 LPC 계수를 구한다. 다음으로 가역필터를 거친 합성음성을 원하는 피치의 2배 피치를 갖도록 늘린 후 늘어난 피치를 반분하는 방법으로 피치를 변경하는 방법이다.

그림 4-2의 캡스트럼 분석에 의한 피치변경법은 주어진 음성신호의 캡스트럼에 대하여 검출해 놓은 피치에 근거한 캡스트럼변경을 수행하고 변경된 캡스트럼을 다시 시간영역으로 변환하여 피치가 변경된 합성음성을 얻는다. 이 과정에서 음성의 위상성분은 영으로 가정하였다.

그림 4-3은 위에서 설명한 두 방법의 특성들로부터 도출한 방법으로 먼저, 시간영역에서 여기원과 유사한 파형으로 고려되는 파형을 얻기 위하여 여러 검출해 놓은 피치를 차단대역을로 갖는 저역통과필터에 통과시킨다. 이 파형에 대하여 시간축을 원하는 변경율만큼 스케일링하고 FFT를 취하여 위상 정보를 보존한다. 한편 동일한 음성신호를 주파수 영역으로 변환하여 진폭스펙트럼을 얻고, 이 스펙트럼을 리프터에 통과시켜 성도의 특성을 갖는 성분과 성대의 특성을 나타내는 성분으로 분리한다. 여기원의 특성을 나타내는 성분에 대하여 주파수축을 스케일링하고 이렇게 주파수축이 스케일링된 여기원 스펙트럼과 리프터를 통과한 후의 성도성분을 합쳐 스펙트럼 간격이 변화된 진폭스펙트럼을 얻는다. 끝으로 첫번째 처리과정에서 얻은 위상정보와 두번째 처리과정에서 얻은 진폭성분을 IFFT하여 시간영역으로 변환시키면 피치가 변경된 합성파형이 얻어진다.

표 4-1은 앞서 설명한 세가지 피치변경법들을 음성시료에 적용하여 원래 음성시료의 스펙트럼과 피치를 변경한 후의 스펙트럼 사이의 왜곡을 조사하여 성별별 최대 왜곡율 및 평균값을 나타내었다.

V. 결 론

VLSI 설비가 자동화되고 고도로 기계화됨에 따라 음성을 통하여 직접 설비를 제어 및 운용하려는 음성신호처리에 관한 연구가 활발히 진행되고 있다. 특히, 음성신호의 특성만을 사상하여 상황에 맞는 음성응답을 사용자에 할 수 있는 합성기법에 관한 연구가 많이 제안되고 있다.

본 논문은 기존에 제안된 파형부호화법에서의 피치변경법들을 간단히 소개하고 그 한계성을 다룬 다음, 본 연구실에서 파형부호화와 피치검출법들의 장점을 이용하여 제안했던 세가지 피치변경법들을 비교 대상으로 하였다. 이 방법들은 데이타량의 규모에 한계를 갖지 않으면서 자동 음남사스템에서 고음질을 제공해 줄 수 있는 규칙합성용 파형보당 합성기법들이다.

피치주기반분법은 시간영역에서 변경하고자 하는 피치주기의 두 배의 피치를 다시 반분하는 방법이므로 피치가 변경되어도 파형의 위상

표 4-1 스펙트럼 왜곡율(%).

Table 4-1. Spectrum distortion rate(%).

성별 / 방법	남	여	평균
방법 1	12.57	15.31	13.94
방법 2	0.8	1.47	1.135
방법 3	1.784	2.94	2.36

정보가 유지된다. 따라서 한 피치주기단위로 합성시에 파형편집이 용이하다. 또한 피치검출이 이루어졌을 경우에 기존의 선형예측합성법을 파형의 연장된 부분에만 적용하므로 처리시간이 N번의 덧셈과 N번의 시프트연산만 요구된다.

두번째, 시간-주파수영역의 변환특성을 이용한 켑스트럼 분석에 의한 피치변경법을 사용하는 경우 스펙트럼의 변경을 최소화시키는 역컨볼루션 기법을 적용하기 때문에 스펙트럼의 왜곡율이 낮게 나타난다.

마지막으로 위상보상된 고조파 스케일링에 의한 피치변경법은 시간영역에서 위상정보를 보존하고 주파수 영역에서 피치를 변경하는 방법이기 때문에 한 피치주기 단위로 합성시에 파형 편집이 용이하고 래퍼런싱이 작아서 자연스러운 음소연결이 가능하다.

실험결과에서 피치주기기반분법은 성문파 성도의 특성을 원래스펙트럼과 같게 유지하지 못하고 스펙트럼의 왜곡을 초래하게되는 단점이 있다. 원래의 스펙트럼을 기준으로 하였을 때 남녀화자의 모든 경우에 백분율로 평균 90.5%를 추정하였으며, 피치를 늘린 경우 보다는 줄인 경우가 스펙트럼왜곡이 9% 정도 높게 나타났다. 켑스트럼에 의한 피치변경법은 시간-주파수 혼성 영역에서 처리되어 계산이 복잡하지만 디지탈 신호처리 전용칩의 도움을 받아 실시간으로 처리할 수도 있다. 최대 스펙트럼왜곡은 남성화자가 0.8% 정도이고, 여성화자는 1.47% 정도로 아주 우수하게 나타났다. 위상보상된 고조파를 이용한 피치변경법의 경우 포만트 성분이 변경되지 않으므로 의사정보가 그대로 보존된다는 장점이 있고 평균 피치왜곡율이 2.36%로 우수하게 나타났다.

REFERENCE

[1] G. Bristow, *Electronic Speech Synthesis*, McGraw-Hill, 1984.

[2] T.W. Parsons, *Voice and Speech Processing*, McGraw-Hill, 1986.

[3] E.J. Yannakoudakis and P.J. Hutton, *Speech Synthesis and Recognition Systems*, Ellis Horwood Ltd., 1987.

[4] J.L. Flanagan, 2nd ed., *Speech Analysis, Synthesis and Perception*, NY:Springer-Verlag, 1972.

[5] A.N. Ince, *Digital Speech Processing - Speech Coding, Synthesis, and Recognition* -, Kluwer Academic Publishers, 1992.

[6] J.R. Deller, J.G. Proakis, J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Co., 1993.

[7] P.E. Papamichalis, *Practical Approaches To Speech Coding*, Prentice-Hall, 1987.

[8] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, 1978.

[9] S. Saito and K. Nakata, *Fundamentals of Speech Signal Processing*, Academic Press, 1985.

[10] B. E . Caspers and B. S. Atal, "Changing pitch and duration in LPC synthesised speech using multiplse excitation," J. Acoust. Soc. Amer., suppl., vol.73, No.1, pp.S5, Spring, 1983.

[11] M. G. Stella and F. J. Charpentier, "Diphone synthesis using multiplse coding and a phase vocoder," in Proc. IEEE ICASSP'85, pp.740-744, 1985.

[12] A. Varga and F. Fallside," A technique for using Multipulse Linear Predictive Speech Synthesis in Text-to-Speech Type Systems", IEEE Trans. Acoust., Speech, Signal processing, Vol.ASSP-35, No.4, PP.586-587, April 1987.

[13] 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, March 1992.

[14] 강동규, 김율채, 배명진, 안수길, "음성 파형의 halving 기법에 의한 파형코딩의 피치변경에 관한 연구", 한국음향학회 추계발표회(국제음향학회)논문집, pp.107-111, 1990. 11. 10.

[15] M. Bae, H. Yoon, S. Ann," On Altering the Pitch of Speech Signals in Waveform Coding - Alteration Method by the LPC and the Pitch Halving -," J., Acoustics of Korea, Vol.10, No.5, pp.11-19, October 1991.

[16] 배명진, 이미숙, 이혜군, 안수길, "켑스트럼 분석에 의한 음성 파형코딩의 피치변경에 관한 연구", 제 4회 신호처리합동학술대회 논문집, Vol.4, No.1, pp.304-309, 1991. 9.

[17] 민경중, 배명진, 윤희상, 안수길, "음성 파형코딩의 음원피치 변경에 관한 연구", 한국음향학회 학술발표회 논문집, Vol.10, No.1(s), pp.45-49, Nov.9, 1991.

[18] U. Kim, J. Lee, M. Bae, S. Ann, "On a Pitch Change of the Waveform Coding by the Cepstrum Analysis for Speech Waveforms on the Phase Compensation for the Pitch Change -", J., Acoustics of Korea, Vol.11, No.5, pp.51-58, October 1992.

[19] 금홍, 김대식, 배명진, "위상 보상된 고조파를 이용한 파형코딩의 피치변경법", 한국통신학회 하계종합학술발표회 논문집, Vol.17, No.1, pp.151-154, 1992. 7.