음성 파형코딩의 음원피치 변경에 관한 연구

- LPC와 주기반분법에 의한 피치변경법 -

(*)인 경중 (*)배 명진 (*)윤 희상 (**)안 수길
(*) 호 서 대 학 교 (**) 서 울 대 학 교

On Altering the Pitch of Speech Signals in Waveform Coding

-(Altering Method by the LPC and the Pitch Halving)-

(*)Kyungchoong YIN (*)Myungjin BAE (*) Heesang YOON (**) Souguil ANN
(*) Hoseo University (**) Seoul National University

본 논문은 91-통신학술단체 육성지원금에 의하여 이루어 졌음.

ABSTRACT

In area of the speech synthesis, the waveform codings with high quality are mainly used to the synthesis by analysis. However, it is difficult to applying the waveform coding to the synthesis by rule, because the parameters of this coding are not classified as either excitation parameters and vocal tract parameters.

In this paper, we proposed a new pitch change method that can alter the pitch periods in the waveform coding. The proposed method expands the pitch period by the LPC synthesis method, and then the period is compressed by the waveform halving technique. Thus, it is possible that the waveform coding is carried out the synthesis by rule in speech processing.

I. 서론

음성신호의 데이타량에 따른 합성단위로는 문장단위의 합성법, 음절단위의 합성법, 음소단위의 합성법 등으로 나눌 수 있다. 한편, 음성합성을 하드웨어로 실현하기 위한 코딩기법으로는 파형코딩, 소우스코딩, 혼성코딩법이 있으며, 메모리 절약을 위해서는 소우스코딩법을, 음질을 높이기 위해서는 파형코딩법을 주로 사용하고 있다.

파형코딩법은 음성의 정보를 합성모델에 따라 분리하지 않고 파형 자체의 잉여성분을 제거한 후에 코딩하는 방법이며 PCM, ADPCM, ADM등이 제안되어져 있다. 최근에는 디지탈 신호처리 전용칩의 제조기술과 파형코딩법의 분석 및 합성알고리즘이 잘 개발되어 32Kbps전송율을 갖는 ADPCM의 표준화가 실현되어 졌다. 그렇지만 파형코딩법은 인간의 개성과 감정을 대변해 주는 성문의 여기정보(excitation)와 의사전달을 나타내는 성도의 필터정보(formants)를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로는 바람직하지 못하다.

소우스코딩법은 여기정보와 필터정보를 분석시에 분리시켜서 독립적으로 코딩하는 방법으로서 LPC, PARCOR, LSP등의 알고리즘이 제안되어 있다. 이들 알고리즘은 10Kbps 이내로 전송율을 낮출 수 있기 때문에 메모리효율적인 코딩법이 된다. 또한 분석시에 추출된 여기정보나 필터정보를 합성시에 인위적으로 변경시킬 수 있기 때문에 음절단위나 음소단위의 합성기법으로 적용이 용이하다. 그렇지만 분석시에 성문을 분리하고, 다시 그 정보를 이용해서 합성하기 때문에 분석시의 오차와 합성시의 오차가 합해져서 합성음질은 자연성이나 명료성이 크게 떨어지게 된다.

소우스코딩의 메모리효율성과 파형코딩의 명료성과 자연성을 적당히 유지하기 위해 이 두가지 코딩기법을 결합시킨 혼성코딩법이 있으며, MLPC, RELP, VELP등이 제안되어져 있다. 그렇지만 혼성코딩법에서는 성도의 필터정보를 코딩하는데 소우스코딩법을 적용하고, 여기정보의 코딩에는 파형코딩법을 주로 적용하고 있다. 이 때문에 여기정보를 변경시켜야 하는 음절단위나 음소단위의 합성알고리즘으로 적용하기에는 바람직하지 못하다.

최근 메모리 제조업체에서는 칩당 16M-bit를 집적화하여 시판하고 있고, 이것을 바이트단위로 쓰기위해 8개 사용

하면 32kbps의 ADPCM 파형코딩법으로 합성하여도 $(16*10^6) * 8 / (32 * 10^3) = 4000$초의 음성 데이타를 수록할 수 있는 많은 양이 된다. 따라서 메모리효율을 살리기 위해 소우스코딩 합성법으로 제대하는 것은 현실적이지 못하며, 음질을 보장받기 위해서도 상용화될 음성합성 코딩법은 파형코딩이나 혼성코딩법이 바람직하게 된다.

그렇지만, 파형코딩법이나 혼성코딩법은 분석후 합성을 하는 문장단위의 합성법으로는 오랫동안 적용되었으나, 음원의 변경이 용이하지 못하기 때문에 단어나 음절 및 음소단위의 합성기법으로 사용되지 못하는 실정이다. 가끔, 단어나 반음절이 단위로 파형코딩법이나 혼성코딩법을 적용하고 있지만 같은 단어라도 연결되는 유형에 따라 다른 데이타 베이스를 적용하고 있는 실정이다.

Ⅱ. 유성음의 피치반분법

음성신호는 그 발성음원에 따라 유성음, 무성음, 묵음으로 구분지을 수 있다. 무성음은 불규칙한 잡음이 성도를 자극하는 입력으로 되어 성도를 통과하는 동안 성도의 협착점에서 공명이 발생한다. 따라서 무성음의 스펙트럼에서는 2500Hz 근방에서 주된 공명봉우리를 갖는 준색잡음의 형태가 된다.

유성음은 준주기적인 성문(glottal)펄스가 성도를 통해 감으로써 발생되기 때문에 유성음 각 음소마다 성도에서 고유한 공명이 일어난다. 이러한 공명봉우리를 포먼트들이라 하고 낮은쪽 주파수에서 부터 두드러진 포먼트들을 차례로 제일, 제이, 등으로 순번을 붙인다. 유성음의 스펙트럼에서는 보통 제일포먼트가 250-750Hz 사이에 존재한다. 또한 유성음은 공명현상 때문에 무성음에 비해 에너지가 크고, 성대의 진동에 의해 준주기성을 띠게 된다. 성대의 진동주기는 남녀노소 및 발성환경에 따라 다르지만 2.5-25ms 정도가 된다.

유성음의 진폭스펙트럼 S(K)는 기본주파수 Fo의 하모닉스 마다 값이 존재하는 라인스펙트럼의 형태를 갖는다. 발성모델에 따라 성문, 성도의 특성 H(K)와 이것을 자극하는 성대의 진동 특성을 E(K)라 하면, 유성음의 스펙트럼 S(K)는

$$S(k) = E(K)\,H(K) = \sum_{l=0}^{N} \delta(K-lFo) \cdot H(k) \qquad \text{---(1)}$$

과 같이 유추할 수 있다. 여기에서 유성음의 기본주파수 Fo를 알고있다면 기본주파수가 두배로 늘어난 유성음의 스펙트럼 S'(k)를 다음과 같이 구할 수 있다. 즉,

$$S'(K) = S(K) \cdot \sum_{l=0}^{N/2} \delta(k-2lFo) = \sum_{l=0}^{N} \delta(K-2lFo) \cdot H(k) \qquad \text{---(2)}$$

스펙트럼 S'(K)는 원래의 유성음 스펙트럼 S(K)에서 기본주파수를 두 배로 늘린 것이 된다. 주어진 유성음에 대해 기본주파수를 두 배로 늘리는 것은 시간영역에서 유성음의 피치를 반분(halving)하는 것이 된다. 시간-주파수관계에 따라 유성음의 파형 s(n)에서 피치 p를 반분해 보면,

$$s'(n) = s(n) * \sum_{l=0}^{P} \delta(n-lp/2) \qquad \text{---(3)}$$

이 된다. 여기서 유성음의 피치는 p=1/Fo이고 이것을 알고 있다고 가정한다. 또한 유성음 s(n)은 피치단위로 주기함수이므로 (3)식을 다시 쓰면,

$$s'(n) = s(n)+s(n-p/2)+s(n-p)+s(n-1.5p)+\cdots+ = P\,[\,s(n) + s(n-p/2)\,] \qquad \text{---(4)}$$

로 간략화될 수 있다. 이렇게 시간영역상에서 음성신호의 피치를 반분하는 방법을 지금부터는 피치반분법(pitch halving)이라고 규정한다.

Ⅲ. 파형코딩의 피치조절

피치반분법은 유성음의 피치를 2의 지수함수로 줄일 수 있지만, 그 사이에 변화를 가하기는 어렵다. 이 때문에 피치를 선형적으로 줄이려면 피치를 늘리고 나서 피치반분법으로 줄이면 가능하게 된다. 유성음의 피치를 p, 늘리는 길이를 L이라 하면 피치반분법을 함께 사용하여 줄일 수 있는 피치구간 p'은,

$$p' = (p+L)/2 = p/2 + L/2 \qquad \text{---(5)}$$

이된다. 따라서 늘리는 샘플길이 L을 조절하면 현재의 피치와 반분된 피치사이의 값으로 변경시킬 수 있게된다.

이제는 피치주기를 L-샘플 만큼 늘리는 방법에 대해 고

려한다. 유성음 스펙트럼에서 각 포먼트의 봉우리는 얼마간의 대역폭을 갖게된다. 이것을 시간영역에서 살펴보면 한 피치 주기만에서 포먼트주파수로 발진하면서, 포먼트 대역폭에 의해 시간에 따라 제동이 발생한다. 이것을 시스템적인 측면에서 고려하면 성도의 조음메카니즘은 안정된 시스템이기 때문에 성대의 진동으로 여기되어진 다음에 일정 시간이 경과하면 점차 감쇠되고 더 이상의 여기가 없으면 음성파형이 영에 도달하게 된다.

이 때문에 유성음의 파형은 피치 구간 사이에 성도의 공명현상이 나타나며 이것은 안정한 성도시스템의 특성을 나타내기 때문에 다음 피치가 나타날 때 까지는 그 파형의 진폭이 점차 감쇠하는 모양이 된다. 역으로, 유성음의 다음 피치가 발생하기 전에 파형진폭은 거의 영에 근접하게 됨으로, 피치주기를 늘리려면 이 부분에 영을 삽입하면 되고, 이 경우에 스펙트럼의 왜곡을 최소화할 수 있게 된다.

Ⅳ. 선형예측 합성에 의한 피치주기의 신장

피치를 늘리기 위해 주기의 끝 부분에 단순히 영값을 삽입하게 되면, 안정된 성도의 특성을 충분히 나타내기 전에 성문의 새로운 여기가 시작되는 발성의 경우에는 명료성이 크게 저하될 수 있다. 이러한 경우의 예로는 짧은 피치를 갖는 여성 및 어린이 발성이나 성도의 길이가 다른 발성에 비해 길게 모델링되는 비음 또는 유성음중에서 /이/의 파형을 들 수 있다. 이때 스펙트럼의 왜곡을 최소화하는 한 방법으로는 피치를 늘리는 부분에 영값을 넣지않고 성도의 특성을 연장시켜주면 된다.

이제 피치주기를 늘리는 방법을 고려한다. 음성신호는 생성모델에 근거하여 선형 예측에 의해 다음과 같이 all-pole모델로 합성될 수 있다.

$$s\hat{}(n) = e(n) + \sum_{i=1}^{4} a_i\, s(n-i) \qquad \text{---(6)}$$

여기서 e(n)은 과거치들의 선형조합에 의해 현재의 표본값이 예측될 때 나타나는 예측에러이지만, 합성시스템에서는 성문의 특성을 나타내는 여기원으로 분류할 수 있다. 여기원은 유성음일 경우 피치주기의 임펄스열로 근사되며, 계수 a_i는 성도 여파기의 특성을 나타낸다. 그러므로 유성음을 발성하기 위해 성문이 열려 한 피치주기가 시작되는 위치에서 예측에러가 최대값이 되고, 다음 피치가 시작되기 전의 부분에서 이 값은 거의 영이 된다.

따라서 유성음 한 피치구간의 끝 부분은 성도의 특성을 지배적으로 나타냄으로써 여기원을 의미하는 에러값은 영으로 근사될 수 있고, 또한 이 부분의 음성 표본값을 식 6의 과거값 s(n-i)으로 하면 피치주기가 연장된 새로운 음성표본값이 예측될 수 있게 된다.

Ⅴ. 피치검출법

피치조절을 수행하려면 우선 유성음에 대해 피치를 정확히 검출하는 것이 중요하며, 지금까지 제안된 피치검출법[8]은 크게 시간영역법, 주파수영역법, 그리고 시간-주파수 혼성법으로 구분지을 수 있다.

시간영역법으로는 parallel processing법, Autocorrelation법, AMDF법, ACM법, 등[1,10,12]이 있으며, 이들은 보통 음성파형의 주기성을 강조시킨 후에 결정논리에 의해 주기성을 판정하는 explicit처리법이다. 시간영역에서 처리하기 때문에 합, 차, 비교, 등의 연산만 보통 필요하다. 그렇지만 음소(phoneme)나 음소결합에 따라 진폭의 크기도 변화되어 피치검출이 어렵게 되며, 특히 잡음이 섞인 경우 에는 분리하기 위한 결정논리가 복잡해져 검출에러가 커지게 된다.

주파수영역법으로는 하모닉스분석법, Lifter법, Comb filtering법 등[1,5-6]이 있으며, 음성 스펙트럼상의 하모닉스간격을 측정하여 그 기본주파수를 보통 측정하게 된다. 일반적으로 스펙트럼은 한 프레임(20-40msec) 단위로 구해지므로 이 구간에서 음소의 천이나 변동이 일어나거나 배경잡음이 발생되어도, 평균화되므로 그 영향을 적게 받게 된다.그러나 처리과정상 주파수영역으로의 변환과정이 필요해서 복잡해지며 기본주파수의 정밀성을 높이는 것은 FFT의 포인트수가 증가되어 처리시간이 길어 진다.

시간-주파수 혼성법에서는 시간영역법의 계산시간 절감과 피치의 정밀성, 그리고 주파수 영역법의 배경잡음이나 음소변화에서도 정확한 피치를 구하는 장점을 취할 수 있다. 이러한 혼성법으로는 캡스트럼법, 스펙트럼비교법[11] 등이 있으나, 시간과 주파수 영역이 동시에 적용되어 계산과정이 복잡하고, 시간과 주파수영역을 왕복할 때 윈도우의 적용에 따른 오차가 피치추출에 영향을 줄 수 있다[1].

피치조절은 시간영역에서 바로 수행되어야 하며, 동시에 피치의 싯점이 검출되어야 한다. 따라서 본 논문에서는 시간영역법의 면적비교법을 적용하였다. 그렇지만 합성을 위해서 파형을 편집하는 경우에는 피치추출이 반드시 자동화될 필요는 없으며, 면적비교법과 함께 반자동법이나 눈으로 찾는 수동법으로 처리하여도 된다.

II. 실험및 결과

시뮬레이션을 위해 IBM PC/AT를 사용하여 여기에 마이크입력이 가능하도록 12-비트 analog to digital converter를 인터페이스시켰다. 화자는 남성 화자와 여성화자를 통해 다음 음성을 발성케하고 8KHz의 샘플링으로 표본화하면서 저장시켰다.

발성1) 23세 남성화자: "인수네 꼬마는 천재소년을 좋아한다."

발성2) 25세 여성화자: "감사합니다."

각 음성시료에 대해 그림 1와 같이 처리 하였다. 먼저 피치 p-를 구한 다음에 변경할 피치 p'-를 얻기 위해서는 L=2(p'-p/2)개의 표본값을 LPC합성법에 의해 만들어 한 피치가 끝나는 곳에 삽입해야 한다. 이렇게 한 다음에 피치변경법에 통과시키면 피치주기가 조정된 파형이 얻어지게 된다.

발성1)에 대해 피치를 50%로 줄인 경우의 결과를 그림 2에 제시하였다. 또한 발성2)의 음성에 대해 한 유성음 부분의 피치를 200%로 연장한 것을 그림 3에 제시하였다. 각 결과 그림에는 비교의 목적으로 유성음의 원래 스펙트럼과 피치를 변경시킨 스펙트럼을 함께 제시하였다.

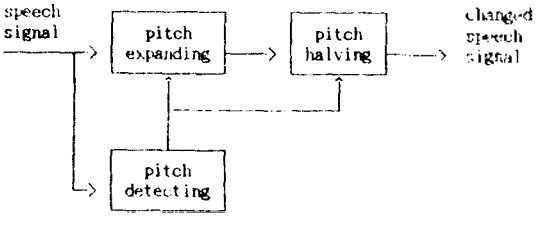

그림 1. 제안한 피치조절 처리과정의 블럭도.

Fig. 1 Processing block diagram for the pitch control of voiced speech.

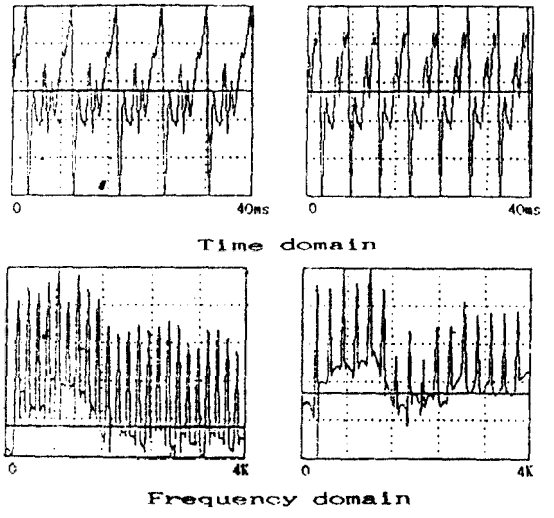

그림 2. 남성화자에 대해 50%로 피치를 줄인 결과.

Fig. 2 Results with compressing the pitch of 50% for male speaker.

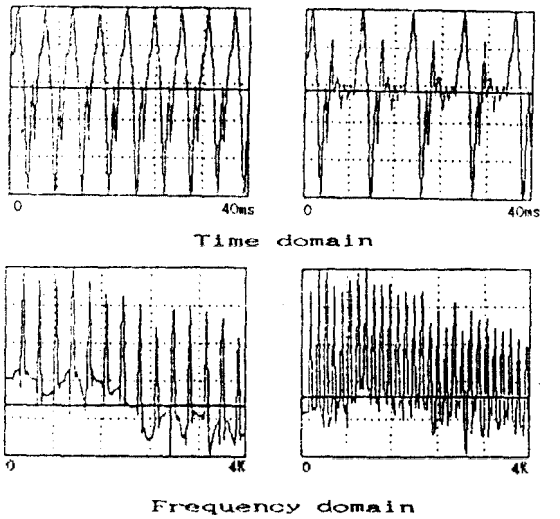

그림 3. 여성화자에 대해 200%로 피치를 늘린 결과.

Fig. 3 Result with expanding the pitch of 200% for female speaker.

Ⅵ. 결 론

음성합성을 하드웨어로 실현하기 위한 코딩기법으로는 파형코딩, 소우스코딩, 혼성코딩법이 있다. 파형코딩법이나 혼성코딩법은 분석후 합성하는 문장단위의 합성법으로는 오랫동안 적용되었으나, 음원의 변경이 용이하지 못하기 때문에 단어나 음절 및 음소단위의 합성기법으로 사용되지 못하고 있다. 가끔, 단어나 반음절어 단위로 파형코딩법이나 혼성코딩법을 적용하고 있지만 같은 단어라도 연결되는 유형에 따라 다른 데이타를 적용하고 있는 실정이다.

따라서 본 논문에서는 파형코딩법들 중에서 선형 PCM법에 대한 음성음의 피치를 제어하는 새로운 방법을 제안하였다. 제안한 방법은 음성의 발성모델에 기인하여 인위적으로 변경시키려는 피치주기의 2배 파형을 선형예측 합성법으로 만든 다음에 그 파형의 주기를 반분하는 기법을 적용하였다. 여기서 제안한 방법은 시간영역에서 처리되며 파형코딩의 다른 변환을 수행하지 않는다.

[REFERENCES]

[1] L.R. Rabiner & R.W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1978.

[2] E.O. Brigham, *The Fast Fourier Transform*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

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

[4] P.E. Papamichalis, *Practical Speech Processing*, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1987.

[5] S.Seneff," Real time harmonic pitch detector," IEEE Trans. Acoust., Speech, Signal Processing, Vol.ASSP-26, pp.358-365, Aug. 1978.

[6] T.V. Screenivas and P.V.S. Rao," pitch extraction from corrupted harmonics of the power spectrum," J. Acoust. Soc. Amer., vol.65, pp.223-228, Jan. 1979.

[7] C.K. On and S.C. Yang," A pitch extraction algorithem based on LPC inverse filtering and AMDF," IEEE Trans. Acoust., Speech, Signal processing, vol. ASSP-25, pp. 565-572, Dec. 1977.

[8] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, and C.A. McGonegal," A comparative perfomance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-24, pp.399- 417, Oct. 1976.

[9] M. Lahat, R.J. Niederjohn, and D.A. Krubsack," A Spectral Autocorrelation Method for Measurement of the Fundamental Frequency of Noise-Courrupted Speech", IEEE Trans., Acoust., Speech, Signal processing, Vol.ASSP-35, No.6, June 1987.

[10] M.BAE, S.SHIN, and S.ANN," The Pitch Extraction of Voiced Speech by the Comparison Between the Original and the Repeated Partial Waveform", J., Acoust., Soc., Korea, Vol.7, No.5, 1988.

[11] 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, 1989.

[12] M.BAE, and S.ANN," Inverse Rate Type Filtering for the Pitch Extraction", J., Acoust., Soc., Korea, Vol.5, No.3, 1986.

[13] ANDREW VARGA & FRANK FALLSIDE," A Technique for Using Multipulse Linear Predictive Speech Synthes- is in Text-to-Speexh Type System", ASSP-35., NO.4, APPIL 1987.

[14] 강동규,김을제,배명진,안수길," On a pitch change of the waveform coding by the halving method for for speech waveform", 국제음향학술발표회논문집pp.107-111, 1990.