

## 켈스트럼 분석에 의한 파형부호화의 피치변경에 관한 연구

# On a Pitch Change of the Waveform Coding by the Cepstrum Analysis of Speech Waveforms

배 명 진\*, 이 미 숙\*  
(Myungjin Bae\*, Misuk Lee\*)

※본 연구는 92년도 한국과학재단 일반목적기초연구비 지원에 의한 결과임.

### 요 약

음성신호의 합성기법들 중에서 파형부호화법은 음질이 우수하기 때문에 분석에 의한 합성법으로 많이 사용되고 있다. 그렇지만 음원과 성도의 특성을 분리하지 않고 파형의 잉여분만을 제거한 후에 파형 자체를 저장하기 때문에 규칙에 의한 합성기법으로 사용하기에는 어려움이 많다. 본 논문에서는 파형부호화법 중에서 선형 PCM 부호화법으로 저장된 음성파형에 대해 피치주기를 조절할 수 있는 켈스트럼 분석법을 제안하여 파형자체의 음원을 분리하지 않고 피치주기를 변경시킬 수 있는 새로운 피치 변경법을 제안하였다. 따라서 음질이 우수한 파형부호화 합성법으로 규칙에 의한 합성을 수행할 수 있다.

### ABSTRACT

The waveform coding is concerned with simply preserving the wave shape of speech signal through a redundancy reduction process. In area of the speech synthesis, the waveform codings with high quality are mainly used to the synthesis by analysis. However, because the parameters of this coding are not classified as either excitation parameters and vocal tract parameters, it is difficult to applying the waveform coding to the synthesis by rule. In this paper, we proposed a new pitch alteration method that can change the pitch periods in the waveform coding by using the cepstrum analysis. Thus, it is possible that the waveform coding is carried out the synthesis by rule in speech processing.

### I. 서 론

음성신호의 데이터량에 따른 합성단위로는 문장단위의 합성법, 음절단위의 합성법, 음소단위의 합성법

등으로 나눌 수 있다. 한편, 음성합성을 하드웨어로 실현하기 위한 부호화기법으로는 파형부호화법, 선형부호화법, 혼성부호화법이 있으며, 메모리 절약을 위해서는 선형부호화법을, 음질을 높이기 위해서는 파형부호화법을 주로 사용하고 있다.

파형 부호화법은 음성의 정보를 발생모델에 따라 분리하지 않고 파형 자체의 잉여성분을 제거한 후에

\*호서대학교 전자공학과  
Hoseo University  
접수일자: 1992. 3. 15

부호화하는 방법이며 PCM, ADPCM, ADM 등이 제안되어져 있다. 최근에는 디지털 신호처리 전용칩의 제조기술과 과형부호화법의 분석 및 합성알고리즘이 잘 개발되어 32Kbps 전송율을 갖는 ADPCM의 표준화가 실현되었다. 그렇지만 과형부호화법은 인간의 개성과 감정을 대별해 주는 여기정보(excitation)와 의사전달을 나타내는 성도의 여파기정보(formants)를 분리하지 않고 처리하기 때문에 음원을 변경시켜야 하는 음절단위나 음소단위의 합성기법으로 바람직하지 못하다.

신호원부호화법은 분석시에 여기정보와 여파기정보를 분리시켜서 독립적으로 부호화하는 방법으로서 LPC, PARCOR, LSP 등의 알고리즘이 제안되어져 있다. 이들 알고리즘은 10Kbps 이내로 전송율을 낮출 수 있기 때문에 메모리효율적인 부호화법이다. 또한 분석시에 추출된 여기정보나 여파기정보를 합성시에 인위적으로 변경시킬 수 있기 때문에 음절단위나 음소단위의 합성기법으로 적용하기에 용이하다. 그렇지만 분석시에 성분을 분리하고, 다시 그 정보를 이용해서 합성하기 때문에 분석시의 오차와 합성시의 오차가 합해져서 합성음질은 자연성이나 명료성이 크게 떨어진다.

신호원부호화의 메모리효율성과 과형부호화의 명료성 및 자연성을 적당히 유지하기 위해 이 두가지 부호화기법을 결합시킨 혼성부호화법으로는 MPLPC, RELP, VELP 등이 제안되어져 있다. 그렇지만 혼성부호화법에서는 성도 여파기정보를 부호화하는 때는 신호원부호화법을 적용하고, 여기정보의 부호화에는 과형부호화법을 주로 적용하고 있다. 이 때문에 여기정보를 변경시켜야 하는 음절단위나 음소단위의 합성알고리즘으로 적용하기에는 어려운 점이 있다.

최근에는 메모리 재조업체에서 칩당 16M-bit를 집적화하여 시판하고 있고, 이것을 바이트단위로 쓰기 위해 8개를 사용하면 32kbps의 ADPCM과형부호화법으로 합성하여도  $(16 \times 10^6) \times 8 / (32 \times 10^3) = 4000$  초의 음성 데이터를 수록할 수 있다. 따라서 합성시에 메모리 양을 줄이기 위해 신호원부호화 합성법을 채택하는 것은 현실적이지 못하며, 음질을 보장받기 위해서도 상용화될 음성 합성부호화법으로는 과형부호화법이나 혼성부호화법이 바람직하다.

그렇지만, 과형부호화법이나 혼성부호화법은 분석 후 합성을 하는 문장단위의 합성법으로는 오랫동안 적용되었으나, 음원의 변경이 용이하지 못하기 때문에 단어나 음절 및 음소단위의 합성법으로는 사용되지 못하고 있는 실정이다. 가끔, 단어나 반음절어 단

위로 과형부호화법이나 혼성부호화법이 적용되고 있지만 같은 단어도 연결되는 유형에 따라 다른 데이터를 적용하고 있는 실정이다.

따라서 본 논문에서는 과형부호화법들 중에서 선형 PCM법에 대한 유성음의 피치를 제어하는 새로운 방법을 제안하고자 한다. 제안할 방법은 음성의 발생 모델에 기인하여 캡스트림의 낮은쪽 시간부분이 성도의 특성을 나타내면서 빠르게 감소하여 피치주기가 가까이에서는 거의 영이되는 성질을 이용하였다. 따라서 변경하려는 피치주기 만큼의 영값을 거의 영이 되는 큐프런시상에 삽입 또는 삭제함으로써 피치주기를 변경하였다. 먼저 음성신호의 발생모델에 따른 컨벌루션의 성질에 대해 알아보고, 분석을 위해 호모머픽 처리를 통하여 역-컨벌루션을 취하는 캡스트림 분석기법과 그 특성에 대해 알아본다. 그다음 음성의 기본주기를 변경하는 피치제어법에 대해 설명하면서 실제의 음성에 대해 처리한 결과볼 검토하였다.

## II. 캡스트림 분석법

짧은 구간의 음성신호를 고려할 때, 유성음은 준주기적인 임펄스열에 의해 그리고 무성음은 불규칙한 잡음에 의해 특성이 시간에 따라 느리게 변화하는 선형시변 시스템을 여기하여 얻어지는 출력으로 모델링할 수 있다. 즉 유성음  $s(n)$ 은

$$s(n) = p(n) * g(n) * v(n) * r(n) = p(n) + h_v(n) \quad (2-1)$$

여기서  $p(n)$ 은 주기  $N_p$ 인 샘플의 주기적인 임펄스열이고,  $h_v(n)$ 은 성분파형의 모양  $g(n)$ , 성도의 임펄스 응답  $v(n)$ , 그리고 방사되는 임펄스 응답  $r(n)$ 의 영향을 조합하는 선형시스템의 임펄스 응답이다.

바탕가져도 무성음의 짧은 구간의 파형은

$$s(n) = u(n) + v(n) * r(n) = u(n) + h_u(n) \quad (2-2)$$

으로 생각 할 수 있다. 여기서  $u(n)$ 은 불규칙한 잡음 여기원이고,  $h_u(n)$ 은 성도의 조합된 영향과 방사효과로 표현하는 시스템의 임펄스 응답이다.

선형 사분면시스템의 여기와 임펄스 응답이 컨벌루션에 의해 조합되기 때문에, 음성분석의 문제는 컨벌루션의 성분을 분류하는 역-컨벌루션의 문제로 귀착된다. 컨벌루션이 덧셈과 같은 대수적 특성을 가졌다

고 볼 수 있기 때문에, 컨벌루션된 선형시스템은 일반화된 중첩의 원리를 적용할 수 있다. 즉,

$$H[x(n)] = H[x_1(n) * x_2(n)] = H[x_1(n)] * H[x_2(n)] \\ = y_1(n) * y_2(n) = y(n) \quad (2-3)$$

이것은 선형 벡터공간의 개념에서 호모머픽 변환으로 취급할 수 있기 때문에, 이것을 만족하는 시스템을 컨벌루션에 대한 호모머픽(homomorphic)시스템이라 한다.

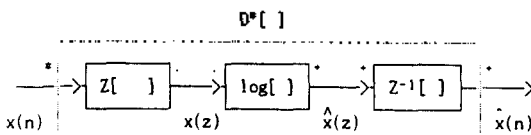
호모머픽 역-컨벌루션의 특성시스템  $D^*[n]$ 은 다음과 같이 입력동작이 컨벌루션의 일반화된 중첩원리에 따르고, 그 출력 동작은 보통의 덧셈이 된다.

$$D^*[x(n)] = D^*[x_1(n) * x_2(n)] = D^*[x_1(n)] * D^*[x_2(n)] \\ = \hat{x}_1(n) + \hat{x}_2(n) = \hat{x}(n) \quad (2-4)$$

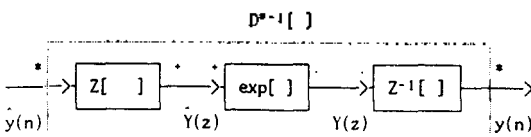
마찬가지로, 역 특성시스템  $D^{*-1}[n]$ 은 아래와 같이 정의된다.

$$D^{*-1}[\hat{y}(n)] = D^{*-1}[\hat{y}_1(n) + \hat{y}_2(n)] \\ = D^{*-1}[\hat{y}_1(n)] + D^{*-1}[\hat{y}_2(n)] \\ = y_1(n) * y_2(n) = y(n) \quad (2-5)$$

한편 입력이 컨벌루션이면, 그 입력의 Z-변환은 대응하는 Z-변환의 곱이 된다. 식 (2-4)로부터 특성 시스템의 출력에 대한 Z-변환은 Z-변환의 더한 조합이어야만 된다. 따라서 이러한 시스템을 만들기 위한 한 가지 방법은 곱에 대한 로그는 개별적인 항의 로그



a) 호모머픽 역-컨벌루션을 위한 특성 시스템



b) 호모머픽 역-컨벌루션을 위한 역 특성 시스템

그림 2-1. 신호의 켈스트럼 분석에 대한 처리 블록도  
Fig 2-1. Block diagram for the cepstrum analysis of signals.

합과 같다는 사실을 다음과 같이 적용하면 된다.

$$\hat{x}(z) = \log[X(z)] = \log[X_1(z) \cdot X_2(z)] \\ = \log[X_1(z)] + \log[X_2(z)] \quad (2-6)$$

만일 신호를 주파수영역이 아닌 시퀀스영역으로 표현하기를 원한다면 특성 시스템은 그림 2-1(a)와 같이 처리할 수 있고, 특성시스템의 역은 그림 2-1(b)로 나타낼 수 있다.

### III. 켈스트럼의 성질

음성신호의 신호원부호화법으로 켈스트럼분석을 사용하는 것은 켈스트럼상에서 음원정보와 여파기정보를 쉽게 분리할 수 있기 때문이다. 그림 3-1처럼 켈스트럼의 낮은쪽 시간부분은 성도, 성분펄스와 방사정보를 나타내며, 높은쪽 시간부분은 여기정보를 나타낸다. 켈스트럼에서 필요한 성분의 추출은 주파수 불변 선형여파기법이라고 하는 다음의 켈스트럼 창함수  $L(n)$ 에 의해 선택할 수 있다.

$$L(n) = 1 \quad |n| < n_0 \\ = 0 \quad |n| \geq n_0 \quad (3-1)$$

여기서  $n_0$ 는 피치주기  $N_p$ 보다 작게 선택된다. 스무딩된 대수함수는 켈스트럼에 창함수를 취하여 구할 수 있다. 이 스무딩된 대수스펙트럼은 입력 음성세그먼트의 광명구조를 나타내고, 포먼트 주파수와 일치한다. 또한 성도, 성분펄스와 방사효과가 조합된 특성을 나타내는 켈스트럼의 낮은쪽 시간 성분은 큐프런시의 증가에 따라 빠르게 감소된다.

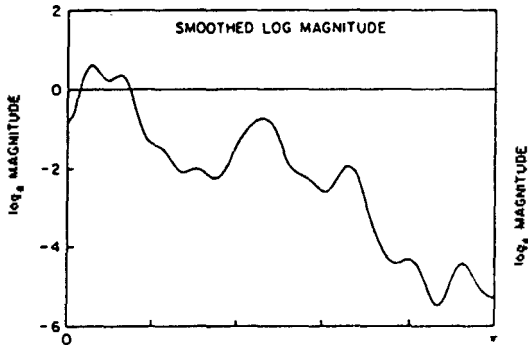
유성음인 경우에 입력 음성세그먼트의 기본주기 위치에서 켈스트럼상에 피크가 나타나지만, 무성음의 켈스트럼에는 이러한 피크가 나타나지 않는다. 이러한 특성은 켈스트럼을 통해 지금의 음성세그먼트가 유성 또는 무성음인지를 결정할 수 있고, 유성음의 기본 주기검출을 가능하게 해 준다.

그렇지만 유성음이라도 켈스트럼상에 뚜렷한 피크가 없거나 낮은 레벨이 될 수 있다. 이것은 켈스트럼 계산과정이 음성파형의 일부 구간만을 취해서 처리해야 하기 때문에 창함수를 적용해야 하는데, 이때 창함수의 길이나 형태에 따른 영향을 받기 때문이다. 즉, 창함수길이가 음성피치의 두배 길이보다 짧다면, 켈스트럼상에 주기성이 나타나지 않는다. 한편, 창함수가 피치주기에 비해 길어지면 길어질수록 켈스트

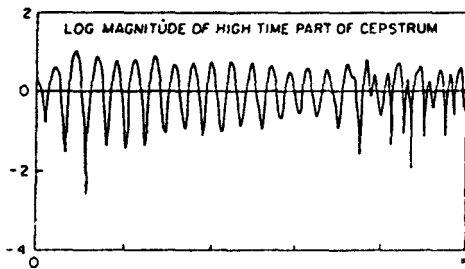
럼의 시작에서 끝까지의 변화가 더욱 커지고 분석도 델로부터 편차가 커지게 된다. 창함수를 적당히 유지하기 위한 한가지 방법은 이전의 피치 추정에 근거하여 추정된 피치값이의 두배로 창함수 길이를 가변하여 적용하는 것이다.

유성음일지라도 켈스트럼상에 두드러진 피크가 나타나지 않는 경우는 대역이 극단적으로 제한된 경우이다. 음성신호에서 유성폐쇄(voiced stop)음은 수백 Hz이상의 주파수에서 고조파 구조가 별로 나타나지 않는 대역이 극단적으로 제한된 경우이다. 일례로 음성신호가 순수한 정현파 형태이면 그의 대수스펙트럼상에는 하나의 피크만 존재한다. 이때 대수스펙트럼에 주기적인 발진이 나타나지 않기 때문에 켈스트럼상에는 피크가 나타나지 않는다.

이처럼 켈스트럼이 피치와 유성음을 잘 표현하지 못하는 경우에는 신호원부호화에 오류로 나타나서 합성음의 음질이 저하된다. 이것을 보상하려면 영-교차율이나 에너지와 같은 다른 정보를 추가해야 하는 등 검출과정이 복잡해진다.



(A) 켈스트럼의 낮은쪽 부분



(B) 켈스트럼의 높은쪽 부분

그림 3-1. 켈스트럼에 대한 신호원부호화  
Fig 3-1. Source coding for the cepstrum.

#### IV. 큐프런시상에서의 피치변경

신호원부호화법으로 켈스트럼을 합성하고자 할 때는 여파기정보를 큐프런시상의 낮은쪽 시간부분에 배치한 다음 역 컨벌루션을 통해 임펄스특성을 구한다. 그런다음 음원정보에 해당하는 피치펄스와 컨벌루션을 통해 유성신호를 합성한다. 이렇게 하면 부호화 효율을 높일 수 있게 되어 낮은 전송율에서도 양질의 합성이 가능하고, 임펄스특성과 컨벌루션되는 피치주기를 변경하여 음율을 조절할 수 있기 때문에, 규칙에 의한 합성법에 적용할 수 있는 부호화기법이 될 수 있다.

그렇지만, 앞장에서 설명한 것처럼 스펙트럼상에서 피치주기 위치에 두드러진 피크가 나타나지 않는 경우에는 신호원부호화시 오류가 발생하게 되고, 이에 의해 합성된 음성은 음질의 저하를 초래한다. 만일, 여파기정보와 음원의 여기정보를 별도로 분리하지 않고 켈스트럼 자체를 부호화하는 과형부호화법을 적용하면 자연음의 음질을 최대한 유지할 수 있다. 그렇지만 이 때 부호화에 필요한 메모리 규모가 방대하고, 음원피치의 변경이 어려워진다. 부호화에 필요한 메모리의 문제는 서론에서도 언급하였듯이 현재의 기술수준으로 극복이 가능하다. 나머지 문제는 규칙에 의한 합성이 이루어지도록 음원의 피치주기를 변경시켜야 한다.

켈스트럼을 살펴보면 영-큐프런시 근방에 대부분의 켈스트럼 값이 모여있고, 큐프런시가 증가함에 따라 그 값이 빠르게 감소된다. 특히, 피치주기 가까이에서는 거의 영이 된다. 이 때문에 변경하려는 주기만큼 켈스트럼값이 거의 영이되는 부분에 삽입하거나 삭제시킴으로써 켈스트럼의 성질에는 거의 영향을 주지 않으면서 음원의 피치주기를 쉽게 변경시킬 수 있다.

켈스트럼상에 영값을 삽입 또는 삭제하여 피치주기를 변경시키고자 할 때, 현재 분석 중인 음성의 피치를 사전에 알고 있다면 영값 대체를 피치주기 가까이에서 하는 것이 바람직하다. 그러나 피치주기 근방의 켈스트럼 피크는 일정한 폭을 유지하는 것이 보통인데, 이것은 분석중인 창함수내에서 시간에 따른 피치의 변화가 진행중인 경우이다. 따라서 피치주기 변경을 위해 영값을 대체하는 위치는 영-큐프런시에서는 멀리 떨어지게 하면서 피치주기에 너무 근접시키지 않는 것이 좋다.

성도의 주파수특성 중에서 제일포먼트의 주파수는 250Hz이상으로 알려져 있고, 음원의 기본주파수는

V. 피치검출법

최대 400Hz까지로 알려져 있다. 따라서 큐프런시상에서 4msec까지는 성도의 특성을 나타내기 때문에 피치를 변경하려면 4msec이상에서 부터 피치주기까지 영값을 삽입하거나 삭제하는 것이 바람직하다.

그림 4-1에는 캡스트럼 분석을 통해 성도의 특성을 그대로 유지하면서 피치를 변경하는 과정을 그림으로 예시하였다. 그림 4-1(a)의 파형은 여성화자가 발성한 /아/부분의 파형이고, 대수스펙트럼은 그림 4-1(b)에 나타내었다. 이에 대한 캡스트럼을 구하여 피치주기를 20샘플 늘리기 위해 4msec에서 부터 영값을 삽입한 캡스트럼과 그의 스펙트럼을 그림 4-1(c)와 (d)에 나타내었다. 그림 4-1(b)와 (d)를 비교해 보면 스펙트럼의 포락선은 그대로 유지되면서 기본주파수가 원하는 값으로 낮아졌음을 알 수 있다.

그러나 4msec이상에서 무조건 피치를 변경하면 스펙트럼왜곡이 크게 나타날 수도 있다. 따라서 이러한 스펙트럼 왜곡을 최소화 하기 위해서는 4msec부터 피치주기 위치까지 피치 삭제구간 길이의 윈도우를 시프트시켜가면서 영향이 최소인 큐프런시 구간을 찾아야 한다. 큐프런시의 대소는 그 윈도우 구간내에 포함되는 에너지값으로 판정하였다. 우선 에너지를 계산한 후, 이 에너지값이 최소가 되는 구간에서 부터 피치주기 변경을 수행한다.

캡스트럼은 그림 4-1(c)에서처럼 피치주기 간격으로 봉우리들이 반복적으로 나타나고 있다. 이러한 캡스트럼의 특징 때문에 피치주기를 변경하기 위해서는 피치주기 구간마다 반복적으로 큐프런시상에 영값을 삽입 또는 삭제하여야 올바른 피치 변경이 된다.

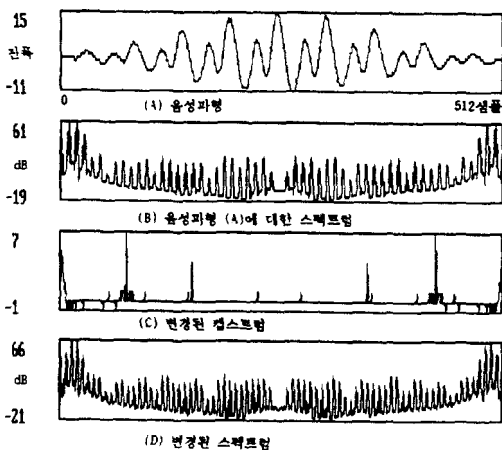


그림 4-1. 캡스트럼 분석에 의한 피치 변경의 일예  
Fig 4-1. One example to alter the pitch by the cepstrum analysis.

신호원부호화법과는 달리 파형부호화법에서 피치를 변경하려면 사전에 그 발성자의 피치변화를 알고 있어야 한다. 이것은 발성자의 억양이나 감정의 변화가 중심된 피치를 기준으로 하여 상대적인 피치주기 변화로 나타나기 때문이다. 따라서 피치 변경시에는 고음질 합성음이 되려면 발성자가 주로 사용하는 피치주기를 기준으로 피치를 변경시켜야 한다. 이때문에 정확한 피치검출이 선행되어야 한다.

지금까지 제안된 피치검출법<sup>(10)</sup>은 크게 시간영역법, 주파수영역법, 그리고 시간-주파수 혼성영역법으로 구분할 수 있다. 시간영역법으로는 parallel processing법, Autocorrelation법, AMDF법, ACM법 등<sup>(11, 10, 12, 13)</sup>이 있으며, 이들은 보통 음성파형의 주기성을 강조시킨 후에 길정논리에 의해 주기성을 판정하는 독립(explicit)처리법이다. 이 방법들은 시간영역에서 처리되기 때문에 보통 합, 차, 비교논리 등의 연산만 필요로 한다. 그렇지만 음소(phoneme)나 음소결합에 따라 진폭의 크기가 변화되어 피치검출이 어렵게 되며, 특히 잡음이 섞인 경우에는 분리하기 위한 결정논리가 복잡해져서 검출오류가 커진다.

주파수영역법으로는 하모닉스 분석법, Lifter법, Comb filtering법 등<sup>(14, 15)</sup>이 있으며, 음성스펙트럼상에서 고조파들의 간격을 측정하여 그 기본주파수를 결정한다. 일반적으로 스펙트럼은 한 프레임(20-40 msec) 단위로 구해지므로 이 구간에서 음소의 전이나 변동이 일어나거나 배경잡음이 발생되어도 평균화되어 나타나므로 그에 따른 영향을 직게 받는다. 그러나 처리과정상 주파수영역으로의 변환과정이 필요하여 복잡해지고, 기본주파수의 정밀성을 높이는 것은 FFT의 포인트 수가 증가되어 처리시간이 길어진다.

시간-주파수 혼성영역법에서는 시간영역법의 계산 시간 절감과 피치의 정밀성, 그리고 주파수영역법의 배경잡음이나 음소변화에서도 정확한 피치를 구하는 장점을 취할 수 있다. 이러한 혼성법으로는 앞에서 설명된 Cepstrum법, 스펙트럼 비교법<sup>(11)</sup> 등이 있으나, 시간과 주파수영역이 동시에 적용되어 계산과정이 복잡하고, 두 영역을 왕복할 때 항함수의 적용에 따른 오차가 피치추출에 영향을 줄 수 있다<sup>(16)</sup>.

피치검출은 캡스트럼을 수행하기 전이나 그 과정 에서 수행되어야 한다. 그리고 캡스트럼을 계산할 때는 항함수의 적용에 따른 영향이 크기 때문에, 그 영향을 최소화하기 위해서는 항함수가 적용되는 사실

을 성문펄스의 여기부분에 일치시키는 것이 좋다. 이렇게 하려면 과형의 위상정보를 잃지 않아야 하며, 이 경우에는 시간영역 피치검출법이 바람직하다. 따라서 본 논문에서는 시간영역에서 수행되는 면적비교법을 적용하였다<sup>13)</sup>.

그렇지만 합성을 위해 과형을 편집하는 경우에는 피치추출이 반드시 자동화된 필요는 없으며, 면적비교법과 함께 눈으로 피치를 검출하는 반자동법이나 눈으로 찾는 수동법으로 처리하여도 된다.

**VI. 실험 및 결과**

이상의 과정을 시뮬레이션 하기 위해 IBM PC / HT(80486)시스템에 마이크 입력이 가능하도록 12-비트 A/D 변환기를 인터페이스시켰다. 화자는 남성 화자와 여성화자를 통해 다음 음성을 발성케 하고 8KHz의 표본화율로 표본화하면서 저장시켰다.

- 발성1) 24세 남성화자: "인수내 꼬마가 천재소년을 좋아한다."
- 발성2) 27세 여성화자: "감사합니다."
- 발성3) 28세 남성화자: "호서대 전자공학과 음성 신호처리 연구"
- 발성4) 32세 남성화자: "예수님은 천지창조의 교훈을 말씀하셨다."
- 발성5) 23세 여성화자: "지금가신 전화는"

전체 실험과정을 통해 시간영역에서 면적비교법을 이용하여 피치를 구한 후에 한 피치구간을 512샘플동안 반복하였다. 그다음 FFT를 수행하여 주파수영역으로 변환하고 대수합수를 취하여 진폭 스펙트럼을 구한 다음 다시 IFFT를 하였다. 이렇게 구해진 캡스트럼상에서 피치주기를 변경하기 위해 영-큐브링 시에서는 멀리 떨어지게 하면서 피치주기에 너무 근접하지 않도록 변경하려는 피치주기 만큼의 최소 에너지구간을 찾아서 그 구간에 영값을 삽입하고 반복되는 나머지 피치구간에서도 같은 위치에 영값을 삽입하거나 삭제하였다.

발성 1)에 대해 피치를 80%로 줄인 경우의 결과를 그림 6-1에 제시하였다. 또한 발성 3)의 음성에 피치를 120%로 늘린 것을 그림 6-2에 제시하였다. 그림 6-1과 6-2에서 각각 (A)는 원래의 음성파형을 나타내고 (B)는 (A)의 음성파형에 대한 대수스펙트럼이며, (C)는 원하는 만큼의 피치주기를 변경시킨 후의 캡스트럼을 나타내고 (D)는 (C)의 캡스트럼에 대한 스

펙트럼이다. 각 결과 그림에서는 비교의 목적으로 음성유의 원래 스펙트럼과 피치를 변경시킨 스펙트럼을 함께 제시하였다. 또한 표 1에는 남성화자와 여성화자에 대해 피치주기를 줄였을 때의 스펙트럼 왜곡율을 나타내었다. 표에서도 알 수 있듯이 제안한 피치변경법은 피치주기를 50%까지 변경하여도 스펙트럼 왜곡율은 1.5%미만으로 우수하게 얻어졌다.

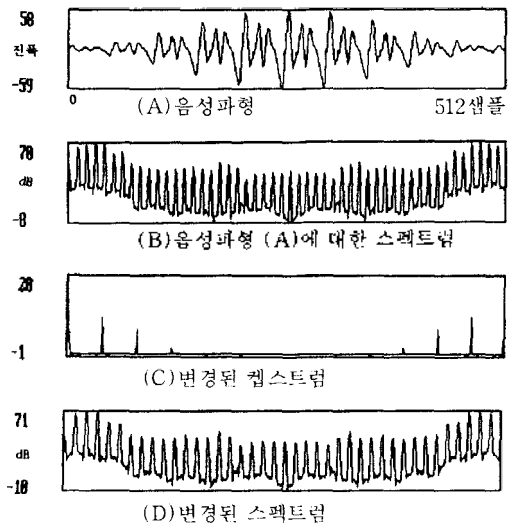


그림 6-1. 남성화자에 대해 80%로 피치를 줄인 결과.  
Fig 6-1. Result with compressing the pitch of 80% for male speaker.

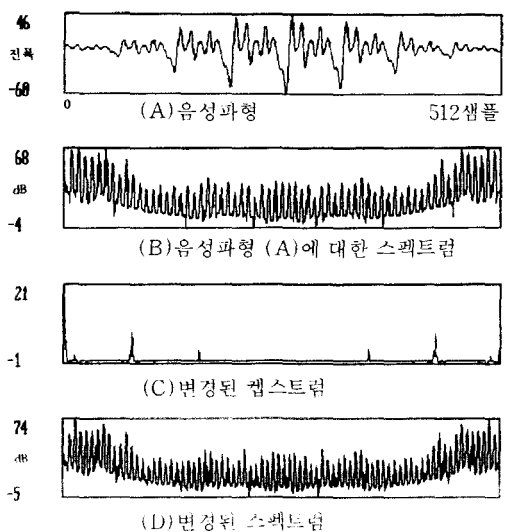


그림 6-2. 남성화자에 대해 120%로 피치를 늘린 결과  
Fig 6-2. Result with expanding the pitch of 120% for female speaker

표 1. 스펙트럼 왜곡율

피치 압축율	남 성	여 성
10%	0.17	0.21
20%	0.31	0.41
30%	0.47	0.74
40%	0.63	1.08
50%	0.80	1.47

## Ⅵ. 결 론

음성합성을 하드웨어로 실현하기 위한 부호화기법으로는 파형부호화법, 신호원부호화법, 혼성부호화법이 있다. 파형부호화법이나 혼성부호화법은 분석 후 합성하는 문장단위의 합성법으로는 오래동안 적용되었으나, 음원의 변경이 용이하지 못하기 때문에 단어나 음절 및 음소단위의 합성법으로 사용되지 못하고 있다. 가끔, 단어나 반음절어 단위로 파형부호화법이나 혼성부호화법을 적용하기도 하지만 같은 단어라도 연결되는 유형에 따라 다른 데이터를 적용하고 있는 실정이다.

따라서 본 논문에서는 파형부호화법들 중에서 선형 PCM법에 대한 음성음의 피치를 제어하는 새로운 방법을 제안하였다. 제안한 방법은 음성의 발생모델에 기인하여 캡스트럼의 낮은쪽이 성도의 특성을 나타내면서 빠르게 감소하고, 피치주기 가까이에서는 거의 영이되는 성질을 이용하였다. 따라서 변경하려는 피치주기 만큼의 영값을 그 부분에 삽입 또는 삭제함으로써 피치를 변경하였다. 이렇게 하면 피치주기를 50% 변경하여도 스펙트럼왜곡이 1.47%로 적게 나타났다.

본 논문에서 제안한 방법은 시간-주파수 혼성영역에서 처리되어 계산이 복잡하지만 디지털 신호처리 전용칩의 도움을 받아 실시간으로 처리할 수도 있다. 그렇지만 합성법에 적용할 경우에는 분석과정이 복잡하고 시간이 많이 소요되어도 합성과정만 간단하면 되기 때문에 계산량의 문제는 큰 부담이 되지 않는다. 또한 이 방법은 스펙트럼의 변경을 최소화시키는 역컨벌루션 기법을 적용하기 때문에 스펙트럼 왜곡율도 매우 낮게 나타났다.

## 참 고 문 헌

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 Approaches to Speech Coding*, 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.Screenivasans 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.Un and S.C.Yang, "A Pitch Extraction Algorithm 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 Performance 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 Auto-correlation Method for Measurement of the Fundamental Frequency of Noise-Corrupted 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. M.BAE and S.ANN, "The High Speed Pitch Extration of Speech Signals Using The Area Comparison Method," *KITE*, Vol.2, No.2, pp.101-105, Feb., 1985.

▲배 명 진

현 숭실대학교 정보통신공학과(11권 1호참고)

▲이 미 숙

호서대학교 전자공학과 대학원생(11권 1호참고)