켑스트럼 피치검색시간 단축에 관한 연구

정회원 조 왕 래*, 최 성 영**, 배 명 진*

A Study on the Pitch Search Time Reduction of Cepstrum Method

Wang-Rae Jo*, Seong-Young Choi**, Myung-Jin Bae* ***Regular Members***

요 약

음성 신호 처리 분야에서 정확한 피치 추출은 매우 중요하다. 음성 신호의 피치 주기를 정확하게 검출할 수 있다면 화자의 영향이 배제된 음성인식이 가능해져 인식의 정확도를 높일 수 있으며 피치 변경도 용이해져 합성음의 개성을 쉽게 변경할 수 있다. 켑스트럼 피치 검색 방법은 천이구간 등에서도 피치를 매우 잘 찾는 우수한 방법이지만 처리 영역변경에 많은 시간이 소요된다는 단점이 있었다.

본 논문에서는 켑스트럼 피치검출법의 피치 검색시간을 단축할 수 있는 새로운 방법을 제안하였다. 기존의 켑스트럼법이 시간영역과 주파수영역간의 변경에 많은 시간이 소요되는 문제점을 개선하기 위하여 영역변경과정에 사용되는 FFT와 IFFT의 비트-재정렬과정을 생략하고 피치 피크 검출시에도 피치가 존재하는 18~142샘플 구간에서만 검색하는 방법을 사용하였다. 제안된 방법을 적용하였을때 기존의 켑스트럼 피치검색법에 비하여 87.94%로 처리시간이 개선되는 결과를 얻을 수 있었다.

Key Words : Pitch Search, Cepstrum Domain, Bit-Reversing, Homomorphic

ABSTRACT

The accurate pitch extraction is very important in speech signal processing. If we measure the pitch period accurately, the accuracy of speech recognition can be higher due to the decrement of speaker dependent effect and we can change the characteristic of synthetic voice easily. Cepstrum pitch search method is a good pitch search method for transition region of speech but it has a drawback of excessive processing time for transform of processing domain.

In this paper, we proposed a new method that can reduce the pitch search time of Cepstrum method. To improve the excessive processing time of conventional method for transformation of processing domain, we use the method that omits the bit-reversing process of FFT and IFFT and searches pitch peak only between 18 and 142 quefrency. As a result of applying the proposed method, we can reduce the pitch searching time by 87.94% over the conventional method.

Ⅰ. 서 론

정보통신의 비약적 발전에 따라 무선 이동통신을 통한 서비스의 범위도 과거의 음성 통신에서 영상, 데이터, 멀티미디어 등으로 점차 확대되고 있다. 하지만 휴대전화의 기본 기능은 여전히 음성통신에 있다고 볼 수 있으며, 이를 위해 디지털 이동통신 분야와 인터넷 기반의 멀티미디어 전송에 적용하기

* 숭실대학교 정보통신공학과 소리공학연구실(wrjo@naver.com), ** 한국폴리텍Ⅱ대학 전자과
논문번호 : 09057-1002, 접수일자 : 2009년 10월 2일

위한 음성신호의 디지털 변환과 전송 데이터량을 줄이기 위한 음성 신호처리 기술에 대한 연구가 진행되고 있다.

음성 신호 처리 분야에 있어 음성 신호의 피치 주기 검출은 매우 중요하다. 음성 신호의 피치 주기를 정확하게 검출할 수 있다면 화자의 영향이 배제된 음성인식이 가능해져 인식의 정확도를 높일 수 있으며, 피치 변경도 용이해져 합성음의 개성을 쉽게 변경할 수 있게 된다[1].

이러한 중요성 때문에 피치검출에 대한 다양한 방법들이 제안되었으며, 이들은 처리영역에 따라 시간 영역법, 주파수 영역법, 시간-주파수 혼성 영역법으로 나눌 수 있다. 시간 영역법은 ACM법, AMDF법, 병렬처리법 등이 있으며 처리법이 매우 간단하지만 천이구간에서의 피치검출이 어렵다는 단점이 있다[2]. 주파수 영역법은 고조파 분석법, Lifter법, Comb-filtering법 등이 있다. 주파수 영역법은 음소의 천이나 변동에 영향을 적게 받는 장점이 있지만 주파수 정확도를 높이기 위해서는 프레임 사이즈가 커지므로 처리시간이 길어지고 변화특성에 둔감해지는 단점이 있다. 시간-주파수 혼성 영역법은 시간영역법과 주파수영역법의 장점을 취한 방법이지만 영역변환에 필요한 계산과정이 복잡하다는 단점이 있다[3].

본 논문에서는 시간-주파수 혼성영역 피치검출법인 켑스트럼 피치검출법의 처리시간을 단축하기 위한 새로운 알고리즘을 제안하였다. 시간영역에서의 음성신호를 켑스트럼영역으로 변환하기 위해 사용하는 FFT와 IFFT의 비트-재정렬 과정을 생략함으로써 피치 검출시간을 단축하고 추가적으로 피치 피크를 찾는 과정에서도 낮은 퀴퍼런시의 값들에 대해서는 생략함으로써 피치검출 시간을 단축할 수 있었다.

Ⅱ. 호모몰픽 디컨벌루션

음성신호 분석의 기본적인 가정중의 하나는 음성신호는 시간에 따라 느리게 변화하는 선형 시변 시스템의 출력으로 표현할 수 있다는 것이다. 이것은 음성신호의 짧은 구간만을 고려할 때 각 세그먼트는 준주기적인 임펄스나 불규칙 잡음에 의해 여기된 선형 시불변 시스템의 임펄스 응답으로 모델링된다는 것이다. 음성신호 분석은 컨벌루션된 여기성분과 성도성분을 분리하여 파라미터화하는 것을 말한다. 호모몰픽 디컨벌루션은 음성의 이러한 특성을 이용하여 여기성분과 성도성분을 분리하는 기법으로 호모몰픽 필터링이라고도 한다[4].

호모몰픽 필터는 시스템을 통과하는 동안 원하지 않는 성분은 제거하는 반면 원하는 성분에는 영향을 미치지 않는다. 컨벌루션된 신호를 분리하고 복원하기 위한 일반적인 호모몰픽 시스템은 그림 1에 나타낸 바와 같이 세 개의 호모몰픽 시스템의 직렬접속으로 표현할 수 있다[5].

그림 1에서 첫 번째 시스템은 특성 시스템이라 하며, 식 (1)과 같이 컨벌루션 입력을 취하여 각 입력에 대응하는 출력의 합으로 출력한다[4].

$$\begin{aligned} D_*[x(n)] &= D_*[x_1(n)*x_2(n)] \\ &= D_*[x_1(n)]*D_*[x_2(n)] \\ &= \hat{x_1}(n)+\hat{x_2}(n)=\hat{x}(n) \end{aligned} \tag{1}$$

이러한 특성 시스템은 컨벌루션을 곱의 형태로 변환하는 Z변환과 곱을 합의 형태로 변환하는 로그연산의 특성을 이용하여 그림 2와 같이 구현할 수 있다.

특성 시스템의 입력이 여기신호 $s(n)$과 성도성분$h(n)$의 컨벌루션이라 한다면 입력신호 $x(n)$은 식 (2)와 같이 표시할 수 있고 Z변환은 식 (3)과 같이 표현할 수 있다.

$$x(n)=s(n)*h(n) \tag{2}$$

$$X(z)=S(z)\ \cdot\ H(z) \tag{3}$$

이것은 로그연산에 의해 합의 형태로 변환되고 역 Z변환에 의해 시간영역으로 변환된다.

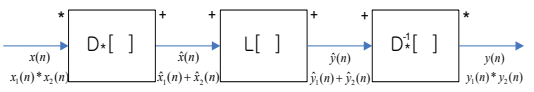


그림 1. 호모몰픽 디컨벌루션 시스템
Fig. 1. Homomorphic Deconvolution

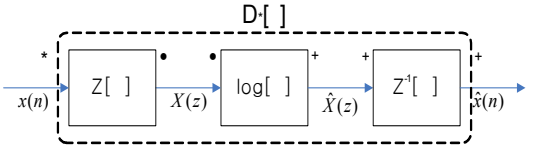


그림 2. 호모몰픽 디컨벌루션의 특성 시스템
Fig. 2. The characteristic system of homomorphic devonvolution

$$\begin{aligned}\hat{X}(z) &= \log[X(z)] \\ &= \log[S(z) \cdot H(z)] \\ &= \log[S(z)] + \log[H(z)] \\ &= \hat{S}(z) + \hat{H}(z)\end{aligned} \tag{4}$$

$$\hat{x}(n) = \hat{s}(n) + \hat{h}(n) \tag{5}$$

Ⅲ. 켑스트럼 피치 추정법

음성신호는 시간영역에서 식 (2)와 같이 여기성분과 여파기성분의 컨벌루션으로 나타낼 수 있으며 주파수 영역에서 음성 스펙트럼은 식 (3)과 같이 여기 스펙트럼과 여파기 스펙트럼의 곱으로 나타낼 수 있다[6].

이러한 스펙트럼을 로그형태로 나타내면 곱의 형태에서 합의 형태로 변환되기 때문에 여기성분과 여파기성분을 쉽게 분리할 수 있다. 이를 다시 시간영역으로 역변환하면 음성신호의 켑스트럼이 구해진다.

음성신호의 로그 스펙트럼과 켑스트럼을 그림 3에 나타내었다. 그림 3(b)와 같이 켑스트럼의 낮은 큐퍼런시 영역에는 여파기 모델에 관한 정보가 들어 있고, 높은 큐퍼런시 영역에는 여기 모델에 관한 정보가 들어있다. 따라서 식 (6)과 같은 리프터(lifter)를 이용하면 성도 여파기의 특성을 구할 수 있다[6].

$$\begin{aligned} l(n) &= 1, \qquad |n| < n_0 \\ &= 0, \qquad |n| \geq n_0 \end{aligned} \tag{6}$$

여기서 n_0는 피치주기 N_p보다 작게 선택된다. 이렇게 구해진 여파기 스펙트럼은 음성신호의 공명특성을 나타내며 포만트 스펙트럼과 같아진다. 또한 켑스트럼상의 성도 여파기 특성은 큐퍼런시가 증가함에 따라 급속히 감소하는 특성을 갖는다. 한편, 음성 켑스트럼의 '0' 큐퍼런시에서 피치 피크까지의 거리를 측정하면 해당 프레임의 피치 주기를 추정할 수 있다. 켑스트럼에 의한 음성 분석과정을 그림 4에 나타내었다.

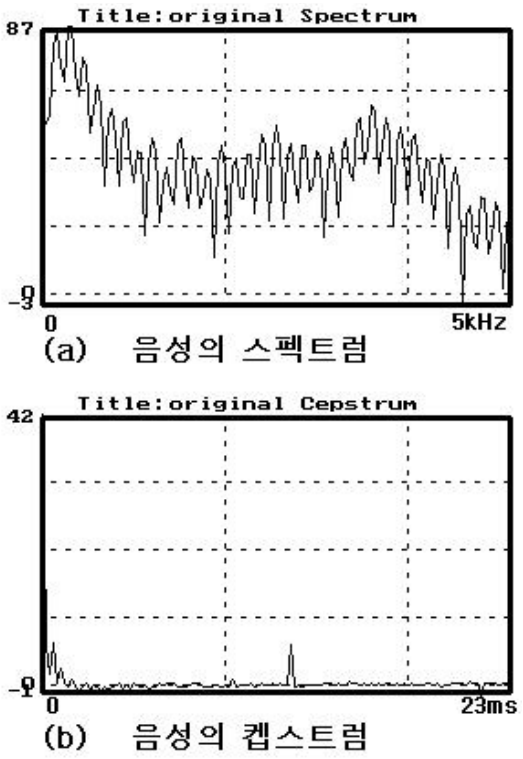


그림 3. 음성신호의 스펙트럼과 켑스트럼
Fig. 3. Spectrum and Cepstrum of Speech

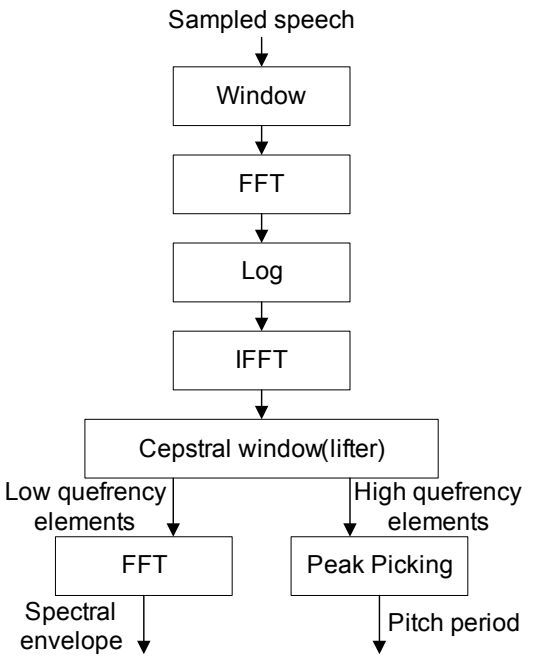


그림 4. FFT 켑스트럼 분석과정
Fig. 4. Process of FFT Cepstrum

Ⅳ. 피치추정 속도 개선

음성신호의 켑스트럼을 구하는 방법은 FFT(Fast Fourier Transform)를 이용하거나 LPC(Linear Prediction Coefficients) 분석을 이용할 수 있으며 전자를 FFT 켑스트럼이라 하고 후자를 LPC 켑스트럼이라 한다.

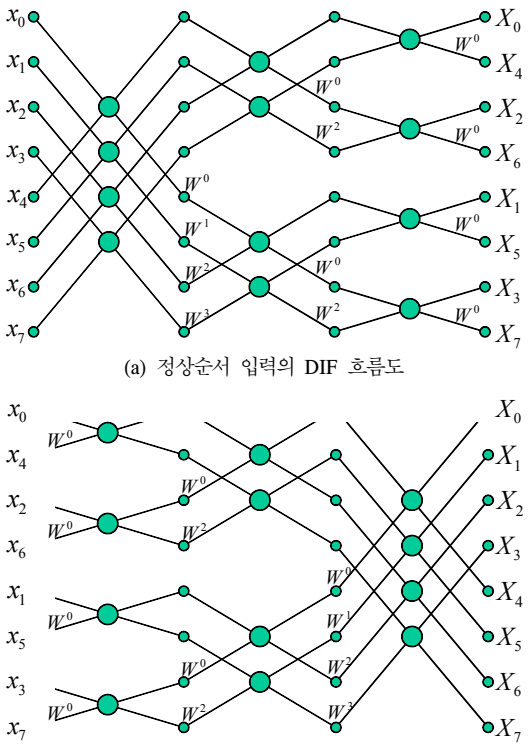
FFT 켑스트럼은 호모몰픽 디컨벌루션의 특성 시스템을 이용한 분석방법으로 그림 4에 나타낸 바와 같이 입력된 음성신호의 FFT를 구하고 로그 연산 후 다시 IFFT를 적용함으로써 구할 수 있다[7].

FFT는 DFT(Discrete Fourier Transform)를 계산하는데 있어 결과는 같으면서도 연산수를 줄여 계산속도를 높이는 방법이다. 계산량을 살펴보면 N개의 샘플을 DFT하는데 각 n에 대하여 N번의 복소수 곱셈이 필요하게 되어 결과적으로 N^2에 비례하는 계산량이 필요하게 된다. 그러나 N개의 샘플을 FFT하는 경우에는 같은 결과를 내면서 계산량은 $N\times\log_2 N$에 비례하도록 줄일 수 있다.

FFT는 DIT(decimation in time)와 DIF(decimation in frequency) 각각에 대해 정상 순서의 입력을 사용한 경우와 비트-재정렬된 입력을 사용하는 방법이 있다. FFT 알고리즘에 가장 많이 사용되는 Cooley-Tukey 알고리즘은 DIF 방법을 사용하며, IFFT의 경우에는 FFT와 같은 방법을 사용하면서 단지 계수들의 켤레 복소수(complex conjugate)를 사용하고 루틴의 끝에서 1/N 스케일링(scaling)을 수행하는 것만이 다르다[8].

그러나 FFT는 계산하고자하는 데이터 샘플수가 $N=2^v$ (v는 정수)가 되어야 한다는 것과 그림 5(a)에 나타낸 바와 같이 입력배열과 출력배열의 순서가 서로 일치하지 않는다는 단점이 있다. 따라서 FFT 수행 전이나 수행 후에 배열의 순서를 재정렬해 주어야만 한다. 이를 비트-재정렬(bit-reversing)이라 하며 계산량에 있어 큰 오버헤드로 작용하게 된다. 이러한 오버헤드는 적은 샘플수를 갖는 데이터에 대한 FFT 연산이 DFT에 비해 큰 이점이 없도록 하며 켑스트럼 분석과 같이 시간-주파수 영역 변환이 잦은 연산의 처리속도에 큰 영향을 미치게 된다[8].

본 논문에서는 켑스트럼 피치 추정 알고리즘의 처리속도를 개선하기 위한 새로운 방법을 제안하였다. 켑스트럼 피치 추정법은 그림 6에 나타낸 것처럼 입력된 음성을 프레임 단위로 나누어 FFT를 취한 후 FFT한 결과에 로그 함수를 적용하여 IFFT를 취하게 되면 그림 3에 나타낸 것처럼 음성 켑스트럼이 구해진다. '0'큐퍼런시 부근에는 성도 켑스트럼이 분포하고 높은 큐퍼런시 영역에는 피치펄스가 나타나게 되는데 큐퍼런시 축상에서 '0'큐퍼런시부터 피치펄스의 위치까지를 측정하여 그 프레임의 피치 주기로 결정하게 된다. 따라서 피크 검출기를 적용하여 피치를 구한다.



(a) 정상순서 입력의 DIF 흐름도

(b) 비트-재정렬된 입력의 DIF 흐름도

그림 5. 8-포인트 FFT의 흐름도
Fig. 5. Flow graphs for 8 point FFT

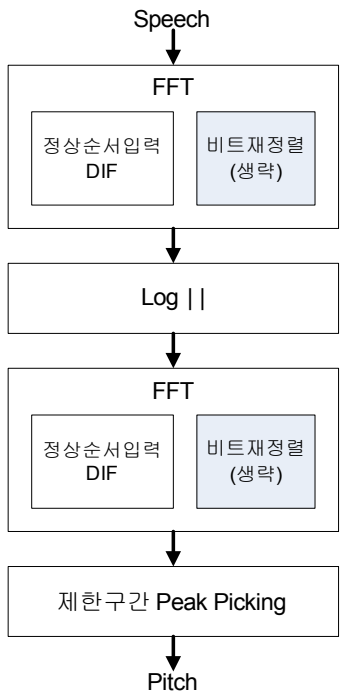


그림 6. 제안한 켑스트럼 피치 추정 알고리즘
Fig. 6. Proposed pitch estimation algorithm

본 논문에서는 FFT와 IFFT 함수에서 비트-재정렬 과정을 생략하고 피크를 찾는 구간을 제한함으로써 처리시간을 단축하는 방법을 제안하였다.

기존의 켑스트럼 피치변경법은 FFT와 IFFT에 동일한 알고리즘을 적용함으로써 필연적으로 비트-재정렬 과정을 수행하여야 하였으며 이러한 오버헤드는 처리시간에 큰 영향을 주게 된다[8]. 그러나 FFT에는 그림 5(a)와 같이 정상순서 입력의 DIF 방법을 사용하고 그 결과에 로그를 취한 후에 그림 5(b)와 같은 비트-재정렬된 입력의 DIF 방법을 사용하여 IFFT하면 정상순서의 켑스트럼을 얻을 수 있게 되어 FFT와 IFFT 과정에서 비트-재정렬 과정을 생략 할 수 있게 된다[9]. 또한 추가적인 처리속도의 개선을 위하여 성도성분이 몰려있는 켑스트럼의 낮은 큐퍼런시 영역에서에는 피크를 찾지 않고 18에서 142샘플 범위에서 찾는다. 제안한 피치 검색법의 처리과정을 그림 6에 나타내었다.

Ⅴ. 실험 및 결과

논문에서 제안한 방법의 성능을 측정하기 위해 기존의 켑스트럼 피치 검색법과 제안한 켑스트럼 피치 검색법을 Lenovo/T-61p(Intel Core2 Duo 2.6 ㎓) 노트북에서 C++로 구현하였다. 각 알고리즘의 함수는 부동소수점 연산을 사용하였고 FFT와 IFFT 프로그램은 참고문헌 10과 11에 소개된 프로그램을 기본으로 작성하였다[10],[11].

먼저 기존의 켑스트럼 피치 검색법의 FFT와 IFFT에서 비트-재정렬 과정이 차지하는 시간 비율을 알아보기 위하여 음성 신호 128샘플, 256샘플, 512샘플 단위로 FFT를 수행하면서 전체 처리시간과 비트-재정렬 시간을 측정하여 표 1에 나타내었다. 각각의 프레임 크기별로 1,000회씩 측정하여 평균시간을 나타내었다. 표 1에 나타낸 바와 같이 256샘플 FFT의 경우 전체 처리시간의 12.39%가 비트-재정렬에 소요됨을 알 수 있었다. 따라서 제안한 켑스트럼 검색법의 처리속도가 기존의 켑스트럼 검색법의 처리속도에 비해 약 12%~13% 정도가 개선되리라 예상할 수 있다.

다음으로 기존의 켑스트럼 피치 검색법과 제안한 켑스트럼 피치 검색법의 처리시간을 비교하여 표 2에 나타내었다. 128샘플, 256샘플, 512샘플을 한 프레임으로 처리하여 비교하였고 각각에 대하여 1/2프레임씩 오버랩하여 해밍윈도우를 사용하여 처리하였다. 기존의 켑스트럼 피치검색법의 FFT와 IFFT에는 정상순서 입력의 DIF방법을 사용한 후 비트-재정렬을 수행하였고 제안한 켑스트럼 피치검색법은 FFT를 할 때는 정상순서 입력의 DIF방법을 사용하고 IFFT에는 비트-재정렬된 입력의 DIF방법을 사용함으로써 비트-재정렬 과정을 생략하였다. 표 2의 결과에 나타낸 바와 같이 256샘플 단위로 처리하는 경우 기존의 켑스트럼 피치검색법의 1278.38㎲에 비하여 제안한 켑스트럼 피치검색법이 1124.26㎲로 87.94%로 단축됨을 알 수 있었다.

표 1. FFT 처리시간에서 비트-재정렬 시간의 비
Table 1. The ratio of FFT time vs. Bit-reversing time

	처리시간(㎲)		비트-재정렬 시간비(B/A)
	전체처리시간 (A)	비트-재정렬시간 (B)	
128샘플	399.90	44.91	13.21%
256샘플	617.69	75.56	12.39%
512샘플	1306.73	173.32	13.26%

표 2. 처리시간의 비교
Table 2. Comparison of processing time

	처리시간(㎲)		처리시간 단축율 (B/A)
	기존의 방법 (A)	제안한 방법 (B)	
128샘플	849.80	747.98	88.01%
256샘플	1278.38	1124.26	87.94%
512샘플	2660.46	2308.82	86.78%

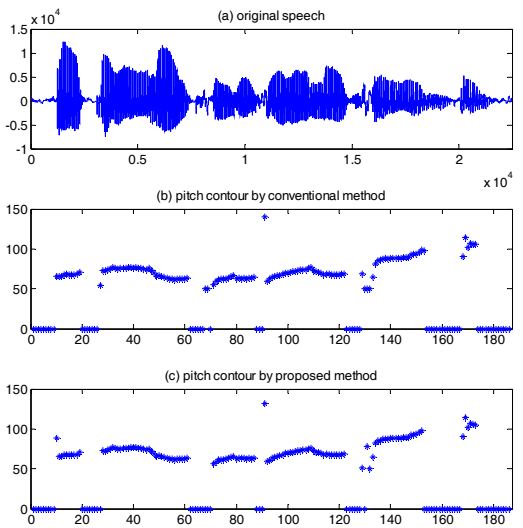


그림 7. 기존의 방법과 제안한 방법에 의한 피치 곡선
Fig. 7. Pitch contour by conventional and proposed method

그림 7에는 기존의 켑스트럼 피치검출법과 제안한 켑스트럼 피치 검출법에 의해 /인수네 꼬마는 천재소년을 좋아한다./라는 발성의 피치를 추정한 그래프를 나타내었다. 그림에 나타난 바와 같이 기존의 방법이나 제안한 방법 모두 음성신호의 피치를 양호하게 검출하고 있음을 알 수 있다.

Ⅵ. 결 론

본 논문에서는 시간-주파수 혼성 영역법인 켑스트럼 피치 검색법의 처리시간을 단축하기 위한 새로운 알고리즘을 제안하였다. 기존의 켑스트럼 피치 검색법이 시간영역과 주파수영역으로의 변환과정에서 많은 시간이 소요되는 문제점을 개선하기 위하여 영역 변경 과정에 사용되는 FFT와 IFFT의 비트-재정렬 과정을 생략하고 피치 피크를 찾는 과정에서도 전체 구간에서 피크를 찾지 않고 여기성분이 존재하는 제한된 큐퍼런시 영역에 대해서만 피크를 검색함으로써 처리시간을 단축하는 방법을 사용하였다. 제안한 방법을 적용한 결과 256샘플 단위 처리의 경우 기존의 켑스트럼 피치 검색법에 비하여 87.94%로 처리시간이 크게 개선되는 결과를 얻을 수 있었다.

참 고 문 헌

〔1〕 S. Narusawa, N. Minematsu, K. Hirose, and H. Fujisaki, "Automatic Extraction of Model Parameters From Fundamental Frequency Contours of English Utterances," *Proc. of ICSLP 2002*, Vol.3, pp.1725-1728, 2002.

〔2〕 Douglas O'Shaughnessy, *Speech Communications Human and Machine*, IEEE Press, 2000.

〔3〕 S. Seneff, "Real Time Harmonic Pitch Detection," *IEEE Trans. Acoust. Speech and signal Processing*, Vol.ASSP-26, pp.358-365, Aug. 1988.

〔4〕 P. E. Paparnichalis, *Practical Speech Processing*, Prentice-Hall, 1987.

〔5〕 L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Prentice-Hall. 1978.

〔6〕 Sadaoki Furui, *Digital Speech Processing, Synthesis, and Recognition*, Marcel Dekker Inc., 2001.

〔7〕 배명진, *디지탈 음성분석*, 동영출판사, 1998

〔8〕 Paul M. Embree, Bruce Kimble, *C Language Algorithm for Digital Signal Processing*, Prentice-Hall, 1991.

〔9〕 Wangrae Jo and Myungjin Bae, "On a Fast Pitch detection using the Cepstrum Analysis," *GESTS Int'l Trans. Acoustic Science and Engineering*, Vol.2, No.1, pp.1-8, December 2004.

〔10〕 미카미나오키(송봉길 譯), *C 언어에 의한 디지털 신호처리 입문 DSP*, 성안당, 2002.

〔11〕 정익주, *TMS320C5000 DSP를 이용한 실시간 디지털 신호처리*, 생능출판사, 2006.

조 왕 래 (Wang-rae Jo) 정회원

1996년 2월 숭실대학교 정보통신공학과(공학사)
1998년 2월 숭실대학교 전기공학과(공학석사)
2000년 3월 숭실대학교 정보통신공학과(박사과정)
1998년~2003년 벽성대학 전자과 전임강사
2003년~현재 디비정보통신 책임연구원
<관심분야> 음성합성, 음성부호화

최 성 영 (Seong-young Choi) 정회원

1980년 2월 울산공과대학 전자공학과(공학사)
1996년 2월 숭실대학교 전산공학과(공학석사)
2004년 8월 숭실대학교 전자공학과(공학박사)
1990년~2006년 서울정보 기능대학 통신전자과 부교수
2006년~현재 한국폴리텍Ⅱ대학 전자과 교수/ 산학협력단장
<관심분야> 음성신호처리, 음성합성, 음성코딩, 음성인식

배 명 진 (Myung-jin Bae) 정회원

1986년~1992년 호서대학교 전자공학과 조교수

1992년~현재 숭실대학교 정보통신공학과 교수

<관심분야> 음성신호처리, 소리공학, 음향처리