FFT 켑스트럼의 처리시간 단축에 관한 연구*

On a Reduction of Computation Time of FFT Cepstrum

조 왕 래** · 김 종 국** · 배 명 진**
Wang-rae Jo · Jong-kuk Kim · Myung-jin Bae

ABSTRACT

The cepstrum coefficients are the most popular feature for speech recognition or speaker recognition. The cepstrum coefficients are also used for speech synthesis and speech coding but has major drawback of long processing time.

In this paper, we proposed a new method that can reduce the processing time of FFT cepstrum analysis. We use the normal ordered inputs for FFT function and the bit-reversed inputs for IFFT function. Therefore we can omit the bit-reversing process and reduce the processing time of FFT ceptrum analysis.

Keywords: Homomorphic Deconvolution, Characteristic System, Cepstrum Analysis, Fourier Transform, Bit-reversing

1. 서 론

음성 신호처리분야는 크게 음성분석, 음성합성, 음성부호화 등으로 나눌 수 있다. 특히, 음성분석 분야는 다른 분야의 기반이 되는 분야로 매우 중요하다. 음성신호는 국가, 지역, 연령, 성별 등에 따라 특성이 다르고, 동일한 사람에 의해 발성된 음성이라 하더라도 시간적으로 또는 발성 당시의 주변환경 등에 따라 다른 특성을 나타낸다. 따라서 음성분석은 확률적이고 통계적인 방법으로 분석해야 하는 어려움이 있다.

현재 음성인식이나 화자인식에서 가장 많이 사용되는 음성의 특징값으로는 켑스트럼 계수(cepstrum coefficients)가 있다[1]. 켑스트럼 계수는 계산방법에 따라 LPC 켑스트럼과 FFT 켑스트럼으로 나뉜다. 두 방법의 인식률은 그다지 차이가 없으나 FFT 켑스트럼의 경우 잦은 영역변환으로 인한 처리시간이 길다는 단점으로 인해 일반적으로 잘 사용되지 않는다. 본 논문에서는 FFT 켑스트럼의 처리시간을 단축할 수 있는 새로운 방법을 제안하였다. 처리영역 변환에 사용되는 FFT 알고리즘과 IFFT 알고리즘에서 비트-재정렬(bit-reversing) 과정을 생략함으로써 FFT 켑스트럼의 계산과정을 단순화시키고 처리시간을 효과적으로 단축할 수 있게 된다.

---

* 본 연구는 한국과학재단 특정기초연구(과제번호 R01-2002-000-00278-0)의 지원에 의하여 이루어졌음.

** 숭실대학교 정보통신공학과.

논문은 다음과 같이 구성되었다. 2 장에서는 켑스트럼 분석의 기본이 되는 호모몰픽 분석법을 소개하고 3 장에서는 음성신호의 특징값으로 많이 사용되는 켑스트럼의 특징과 계산과정을 소개하였다. 4 장에서는 논문에서 제안한 FFT 켑스트럼의 계산시간 단축 방법에 대하여 설명하였으며 5 장에 실험결과를 제시하고 6 장에서 결론을 맺었다.

2. 호모몰픽 디컨벌루션

음성신호 분석의 기본적인 가정 중의 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[2].

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.

그림 1에서 첫 번째 시스템은 특성 시스템이라 하며, 식 (1)과 같이 컨벌루션 입력을 취하여 각 입력에 대응하는 출력의 합으로 출력한다[3].

$$\begin{aligned} D_*[x(n)] &= D_*[x_1(n)*x_2(n)] \\ &= D_*[x_1(n)]+D_*[x_2(n)] \\ &= \hat{x}_1(n)+\hat{x}_2(n)=\hat{x}(n) \end{aligned} \tag{1}$$

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

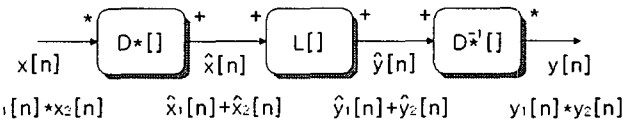


그림 1. 호모몰픽 디컨벌루션 시스템

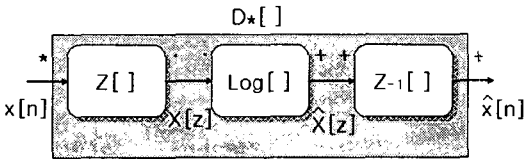


그림 2. 호모몰픽 디컨벌루션의 특성 시스템

특성 시스템의 입력이 여기신호과 성도성분의 컨벌루션이라 한다면 입력신호는 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.

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

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

이러한 특성 시스템의 출력을 복소 켑스트럼이라 하고 여기성분과 성도성분의 합으로 표현된다.

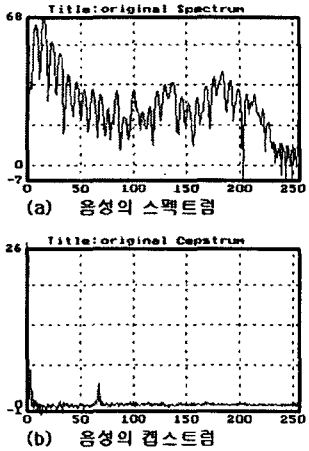
호모몰픽 디컨벌루션 시스템의 두 번째 시스템은 일반적인 선형시스템이고, 세 번째 시스템은 역특성 시스템이라 하며 식 (6)과 같이 정의된다.

$$\begin{aligned} D_*^{-1}[\hat{y}(n)] &= D_*^{-1}[\hat{y_1}(n) + \hat{y_2}(n)] \\ &= D_*^{-1}[\hat{y_1}(n)] * D_*^{-1}[\hat{y_2}(n)] \\ &= y_1(n) * y_2(n) = y(n) \end{aligned} \tag{6}$$

역특성 시스템은 특성시스템과 함께 사용되어 보코더 등의 시스템을 구성할 수 있으며 음성인식 또는 화자인식 등의 음성 특징을 이용하는 시스템에서는 생략되기도 한다.

3. 음성신호의 켑스트럼 분석

음성 신호는 시간영역에서 여기성분과 성도성분의 컨벌루션으로 나타낼 수 있으며 주파수 영역에서 음성 스펙트럼은 여기 스펙트럼과 성도 스펙트럼의 곱으로 나타내어진다. 이러한 스펙트럼을 로그형태로 나타내면 곱의 형태에서 합의 형태로 변환되기 때문에 여기성분과 성도성분을 쉽게 분리할 수 있다. 이를 다시 시간영역으로 역변환하면 음성신호의 켑스트럼이 구해진다. 음성 신호의 로그 스펙트럼과 켑스트럼을 그림 3에 나타내었다. 그림 3(b)와 같이 켑스트럼 계수의 낮은 영역에는 성도 모델에 관한 정보가 들어 있고, 높은 영역에는 여기 모델에 관한 것이 들어 있다. 따라서 가중함수를 이용한 리프터링(liftering)을 통해 성도성분과 여기성분을 분리할 수 있다[4][5].



(a) 음성신호의 스펙트럼 (b) 음성신호의 켑스트럼

그림 3. 음성신호의 스펙트럼과 켑스트럼

4. 켑스트럼 분석과정과 처리시간 단축

음성신호의 켑스트럼을 구하는 방법은 FFT(Fast Fourier Transform)를 이용하거나 LPC (Linear Predictive Coefficients) 분석을 이용할 수 있으며 전자를 FFT 켑스트럼이라 하고 후자를 LPC 켑스트럼이라 한다. 일반적으로 FFT보다 LPC분석 방법이 간편하고 실제 응용시에 처리 계산량이 적기 때문에 LPC계수를 이용하여 켑스트럼 계수를 구하는 것이 일반적이다.

그림 4. FFT 켑스트럼 분석과정

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[6].

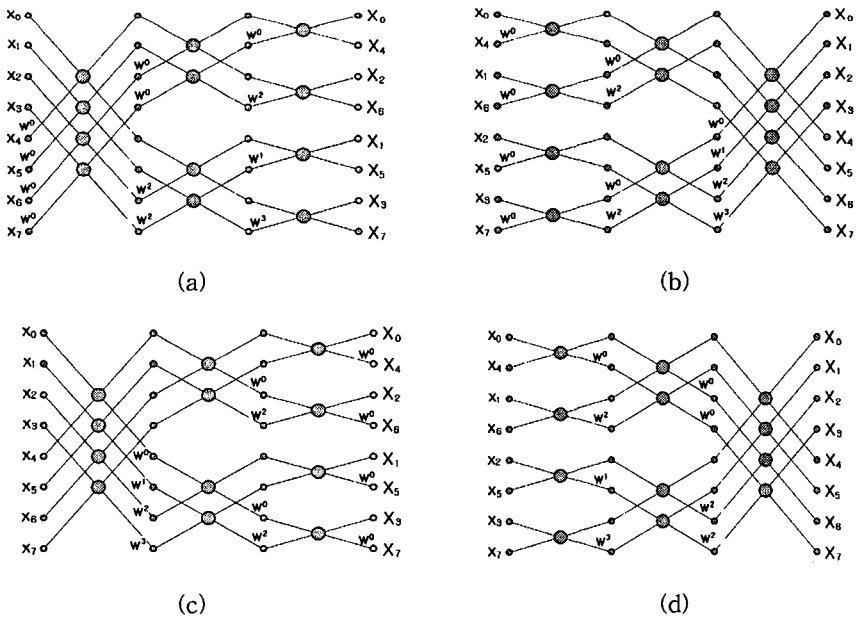
컨벌루션된 두 성분을 곱의 형태로 바꾸기 위해 사용한 FFT는 DFT(Discrete Fourier Transform)를 계산하는데 있어 결과는 같으면서도 연산수를 줄여 그 계산속도를 높이는 방법이다. 일반적인 DFT의 계산식은 식(7), 식(8)과 같다.

$$X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn} \tag{7}$$

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] W_N^{-kn} \tag{8}$$

계산량을 살펴보면 N 개의 샘플을 DFT하는데 각 n에 대하여 N 번의 복소수 곱셈이 필요하게 되어 결과적으로 N^2에 비례하는 계산량이 필요하게 된다. 그러나 $N=2^v$(v는 정수) 개의 샘플을 FFT하는 경우에는 같은 결과를 내면서 계산량은 $N\times \log_2 N$에 비례하도록 줄일 수 있다.

FFT는 그림 5와 같이 DIT (decimation in time)와 DIF (decimation in frequency) 각각에 대해 정상순서의 입력을 사용한 경우와 비트-재정렬된 입력을 사용하는 방법이 있다. FFT 알고리즘에 가장 많이 사용되는 Cooley-Tukey 알고리즘은 DIF 방법을 사용하며, IFFT의 경우에는 FFT와 같은 프로그램을 사용하면서 단지 계수들의 켤레 복소수(complex conjugate)를 사용하고 루틴의 끝에서 1/N 스케일링(scaling)을 수행하는 것만이 다르다[4]. 그러나 FFT는 계산하고자 하는 데이터 샘플수가 $N=2^v$(v는 정수)이 되어야 한다는 것과 그림 5에 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 FFT연산이 DFT에 비해 큰 이점이 없도록 하며 켑스트럼 분석과 같이 시간-주파수 영역 변환이 잦은 연산의 처리 속도에 큰 영향을 미치게 된다[7].

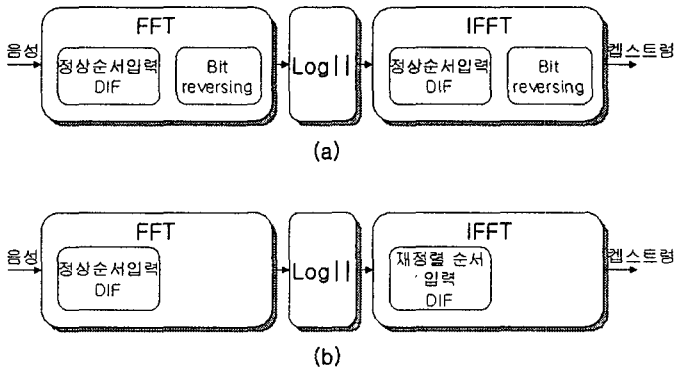


(a) 정상순서 입력의 DIT 흐름도 (b) 비트-재정렬 입력의 DIT 흐름도
(c) 정상순서 입력의 DIF 흐름도 (d) 비트-재정렬 입력의 DIF 흐름도

그림 5. 8-point FFT의 흐름도

본 논문에서는 음성신호의 FFT 켑스트럼을 구할 때 FFT와 IFFT의 비트-재정렬 과정을 생략함으로써 처리시간을 단축하는 새로운 방법을 제안하였다. 기존의 FFT 켑스트럼 분석에서는 FFT와 IFFT에 동일한 알고리즘을 사용함으로써 필연적으로 비트-재정렬을 수행하여야

하였으며, 이러한 오버헤드는 처리시간에 큰 영향을 주었다. 그러나 FFT와 IFFT에 다른 방법을 사용한다면 비트-재정렬 과정을 생략할 수 있게 된다. 즉, FFT에는 정상순서 입력의 DIF 방법을 사용하고 그 결과를 비트-재정렬된 입력의 DIF 방법을 사용하여 IFFT하면 정상순서의 켑스트럼을 얻을 수 있게 된다. 기존의 처리과정과 제안한 처리과정을 그림 6에 비교하여 나타내었다.



(a) 기존의 방법 (b) 제안한 방법

그림 6. FFT 켑스트럼 처리방법의 비교

5. 실험 및 결과

제안한 FFT 켑스트럼 계산법의 성능을 평가하기 위하여 기존의 FFT 켑스트럼 계산법과 제안한 FFT 켑스트럼 계산법을 IBM-PC/pentium III (866MHz)에서 C++로 구현하였다. 기존의 계산 방법은 그림 6(a)와 같이 정상순서 입력의 DIF 계산을 수행한 후 비트-재정렬하는 방법으로 FFT와 IFFT를 수행한 경우이고, 제안한 방법은 그림 6(b)와 같이 FFT에는 정상순서 입력의 DIF를 사용하고 IFFT에는 비트-재정렬된 입력의 DIF를 사용하여 비트-재정렬 과정을 생략하는 방법을 사용하였다. 기존의 방법이나 제안한 방법에 사용된 FFT 알고리즘과 IFFT 알고리즘은 일반적인 신호처리에 많이 사용되는 Cooley-Tukey 알고리즘을 사용하였다[7].

먼저 기존의 FFT 켑스트럼 계산법에 사용된 FFT 알고리즘의 전체 계산시간 중 비트-재정렬 과정이 차지하는 비율을 알아보기 위하여 처리시간을 측정하여 나타내었다. 표 1에 나타낸 바와 같이 128 샘플 FFT를 수행하는 경우 전체 처리시간의 13.1%에 해당하는 시간이 비트-재정렬을 위해 사용됨을 알 수 있다.

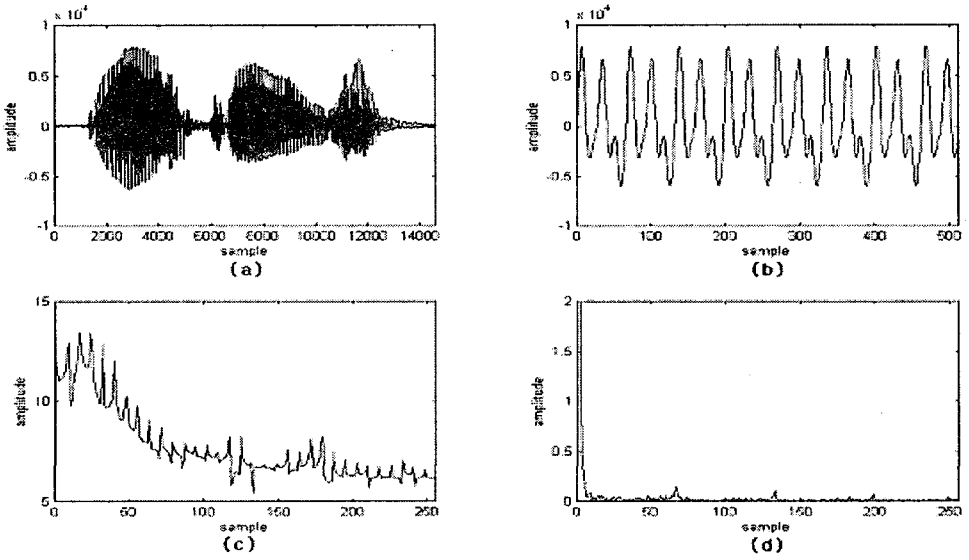
표 1. FFT처리시간 중 비트-재정렬 시간(μs)

	Total (A)	Bit-reversing (B)	Rate (B/A)
128-points	382	50	13.1%
256 points	855	100	11.7%
512 points	1,913	231	12.1%

또한, 기존의 FFT 켑스트럼 계산시간과 제안한 FFT 켑스트럼 계산시간을 비교하기 위하여 128 샘플, 256 샘플, 512 샘플 데이터에 대하여 기존의 방법과 제안한 방법을 적용하여 계산한 시간을 측정하였다. 표 2에 나타낸 바와 같이 512 샘플에 대한 켑스트럼 연산의 경우 제안한 방법이 3,550 usec 소요되어 기존 방법 4,188 usec의 84.8%로 단축됨을 알 수 있었다. 그림 7에는 기존의 방법과 제안한 방법으로 구한 FFT 켑스트럼을 그래프로 나타내었다.

표 2. 처리 시간의 비교

	Processing Time [μs]		Rate (B/A)
	Conventional method (A)	Proposed method (B)	
128 points	835	712	85.3%
256 points	1,863	1,638	87.9%
512 points	4,188	3,550	84.8%



(a) 전체 음성신호 (b) 프레임 음성신호
(c) 음성신호의 스펙트럼 (d) 음성신호의 켑스트럼

그림 7. FFT 켑스트럼 결과

6. 결 론

음성신호의 호모몰픽 분석법에 해당하는 켑스트럼 분석은 처리과정에 따라 FFT 켑스트럼과 LPC 켑스트럼으로 나눌 수 있다.

본 논문에서는 FFT 켑스트럼을 구할 때 FFT 처리시간의 약 11.7%를 차지하는 비트-재정렬(bit-reversing) 과정을 생략함으로써 FFT 켑스트럼의 처리시간을 단축할 수 있는 방법을 제안하였다. FFT 켑스트럼을 구하기 위해 수행하는 FFT와 IFFT에 동일한 알고리즘을

사용하지 않고 FFT는 정상순서 DIF 알고리즘을 사용하고 IFFT는 비트-재정렬 순서 입력 DIF 알고리즘을 사용하면 비트-재정렬 과정을 생략하더라도 기존의 FFT 켑스트럼과 동일한 결과를 얻으면서 처리시간은 기존의 방법의 84.8%로 단축할 수 있게 된다.

이러한 처리시간 단축은 음성신호의 분석시간을 줄임으로써 음성인식이나 음성합성 등의 처리시간 단축에 크게 기여할 수 있을 것으로 기대된다.

참 고 문 헌

[1] Furui, Sadaoki. 1981. "Cepstral analysis technique for automatic speaker verification." *IEEE Trans. on ASSP*, Vol. 29, No. 2, 254-272.

[2] 정혜경, 김유진, 정재호. 2002. "켑스트럼으로부터 변환된 로그 스펙트럼을 이용한 포먼트 평활화 켑스트럴 평균 차감법." *한국음향학회지*, 21(4), 361-373.

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

[4] Juang, B. H., L. R. Rabiner & J. G. Wilpon. 1987. "On the use of bandpass liftering in speech recognition." *IEEE Trans. on ASSP*, Vol. 35, 947-954.

[5] 조왕래, 함명규, 배명진. 1998. "평탄화된 여기 스펙트럼에서 켑스트럼 피치 변경법에 관한 연구." *한국음향학회지*, 17(8), 82-87.

[6] 전선도, 강철호. 1999. "잡음에 강한 음성 인식을 위한 성문 가중 켑스트럼에 관한 연구." *한국음향학회지*, 18(5).

[7] Embree, Paul M. & Bruce Kimble. 1991. *C Language Algorithms for Digital Signal Processing*. Prentice-Hall.

접수일자: 2003. 4. 30.
게재결정: 2003. 6. 5.

▲ 김종국
서울특별시 동작구 상도5동 1-1 (우: 156-743)
숭실대학교 정보통신공학과 음성통신연구실
Tel: +82-2-824-0906
E-mail: kokjk@hanmail.net

▲ 조왕래
서울특별시 동작구 상도5동 1-1 (우: 156-743)
숭실대학교 정보통신공학과 음성통신연구실
Tel: +82-2-824-0906
E-mail: wrjo@unitel.co.kr

▲ 배명진
서울특별시 동작구 상도5동 1-1 (우: 156-743)
숭실대학교 정보통신공학과 음성통신연구실
Tel: +82-2-820-0902
E-mail: mjbae@ssu.ac.kr