상관관계 대칭성을 이용한 CELP 보코더의 델타피치 검색에 관한 연구

정현욱[1], 민소연[2], 배명진[1]
숭실대학교 정보통신공학과[1], 숭실대학교 전자공학과[2]

A Study on Delta Pitch Searching of CELP Vocoder using the Symmetry of Correlation

Hyun Uk Jung[1], So Yeon Min[2], Myung Jin Bae[1]
Dep. of Information and Telecom. Engr.[1], Electronics[2], Soongsil Univ.
digital@ssu.ac.kr

요약

G.723.1은 저 전송률 환경에서 고 음질을 제공하여 주고 있으나 CELP형 부호화기가 갖는 합성에 의한 분석(Analysis by Synthesis)방식의 구조로 인해 많은 처리 시간과 계산량을 요구하게 된다. 본 논문에서는 G.723.1에 대해 NAMDF함수를 적용하여 델타 피치 검색과정의 계산량을 줄여 부호화기의 전체 계산량을 감소시키는 방법을 제안하였다. 기존의 피치 검출 알고리즘에서 피치 검출을 위해사용하고 있는 자기상관함수는 곱셈 연산에서 발생하는 bit의 dynamic range가 커서 나눗셈 연산에서도 과도한 연산량을 필요로 한다. 따라서, 이러한 계산량의 감소를 위해 기존의 자기상관함수 대신 계산량을 감소하기 위하여 NAMDF 방법을 적용하였고 추가된 skipping 기법을 사용하였다.

계산량 감소율 측면에서는 약 64%의 감소율을 보였고 기존의 방법과 제안한 방법간의 피치 pitch contour은 원음성의 피치 contour와 유사하였고, 음질 평가에서도 기존의 G.723.1 부호화기 합성음과 유사한 결과를 얻을 수 있었다.

1. 서론

음성 부호화법에는 크게 파형 부호화법, 신호원부호화법, 혼성부호화법등 세 가지로 나눌 수 있다. 이중에서 파형 부호화법은 음성 신호의 성분 분리 없이 파형 자체의 잉여 성분만을 제거한 후 부호화하여 전송하고 다시 합성하는 방식이다[1]. 또한 파형 부호화법은 고 음질과 화자의 개성이 유지되는 반면에 전송 파형을 유지하기 위한 데이터 량이 많기 때문에 전송률이 높으며 대용량의 메모리가 필요하게 된다. 이에 비해, 신호원 부호화법은 음성의 발성 모델에 근거하여 음성 신호의 여기 성분과 여파기 성분을 분석하여 각각을 독립적으로 분리시켜 부호화하는 방법을 사용하기 때문에 전송 대역폭이 작고 메모리 사용량이 작다. 그러나, 분석 시와 합성 시의 오차가 누적되는 문제점을 갖기 때문에 합성 음질은 자연성과 명료성이 크게 떨어지게 된다[2].

과거의 음성 정보를 전송하는 방법은 한정된 일부 사용자와 특정 분야에 적용되었기 때문에 전송률은 크게 고려되지 않았으며 고음질의 음성을 수신 단에 보낼 수 있으면 되었다. 그러나 과거에 비해 정보통신 문화의 발달에 따라 디지털 이동통신이나 멀티미디어, 음성우편 시스템 등 음성을 이용한 여러 가지 새로운 산업들이 급속히 성장하고 특정 집단이나 개인에게 제공되었던 서비스들이 다수의 대중에게도 제공되어지고 그 숫자 역시 기하급수적으로 늘어나게 되었다. 디지털 이동 통신 분야에 주로 사용되어지는 혼성 부호화법은 신호원 부호화법의 메모리 효율성 및 저 전송률과 파형 부호화법의 고 음질 유지의 장점을 결합시킨 것으로 포만트 정보는 선형 예측 부호화법을 사용하게 되며, 그 나머지 잔여 신호를 어떻게 부호화 하느냐에 따라 RELP, VELP, MELP, CELP법 등이 제안되어져 있다[1,2]. 이들 중에서 CELP계열의 보코더들은 16kbps에서 2.4kbps까지의 저 전송률에서 고 음질을 유지해주기 때문에 이동 통신 및 멀티미디어 환경의 응용 분야에서 주로 사용되어져있다. 이 계열의 부호화법들은 ITU-T에서 여러 기술들이 연구, 논의되어 왔으며 현재는 표준안으로 채택되어졌다. 이들 중에서 G.723.1은 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다.

G.723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷 폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 더불어 최적의 전송 환경을 위하여 두 개의 전송률을 사용하기 때문에

다른 보코더 표준안들에 비해서 더욱 융용성이 높다. 그러나 G.723.1 역시 음성신호를 성분 분리하여 합성하는 방식인 CELP 보코더 계열의 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리 시간의 소모를 피할 수 없다는 문제점을 갖고 있다[1,2].

본 논문에서는 음성부호화기를 실시간으로 구현하는 경우에 발생하는 문제점들을 극복하여 음질 향상과 저 전송률 환경 하에서 계산량을 감소시키는 기법을 제안하고자 하는 것이다. 이를 위하여 고 음질 저 전송 보코더인 G.723.1에 관한 연구를 수행하여 이후 저 전송 환경 하에서도 음질 열하 없이 처리가 가능한 음성 부호화기를 개발할 수 있는 기반 기술을 확보하고자 한다. 본 논문에서는 G.723.1의 피치 검색과정에서 사용하는 자기상관 법 대신에 NAMDF(Normalized Average Magnitude Difference Function)방식과 새로운 skipping 기법을 적용시킨 후 기존의 G.723.1에 비해 피치 검색 시 계산량을 감소하는 방법을 제안하였다[1-3].

본 논문의 구성은 다음과 같다. 2장에서는 자기상관방법과 AMDF(Average Magnitude Difference Function)방법, NAMDF(Normalized AMDF)의 특징에 대하여 설명하고 3장에서는 제안하는 방법에 대하여 설명한다. 4장과 5장에서는 각각 실험결과 및 결론을 맺는다.

2. 근사 분석기법

2-1. 자기 상관법

음성신호의 퓨리에 변환으로 음성스펙트럼의 진폭과 위상을 분리한다. 많은 음성응용분야에서 위상을 무시한다. 왜냐하면 인지적인 측면에서 상대적으로 진폭에 비해서 중요도가 작고, 진폭에 비해 적은 정보를 가지고 있기 때문이다. 에너지 스펙트럼의 역변환인 시간 신호를 원신호의 자기상관이라 한다. 자기상관함수는 신호의 고조파와 포만트 진폭, 주기 정보를 유지하고, 위상정보는 무시한다. 이 자기상관함수는 피치검출, 유/무성음 결정 그리고 선형예측에 사용된다. 자기상관함수는 다음과 같은 상호상관함수의 특수한 경우이다[1,2,4].

$$\phi(k) = \sum_{m=-\infty}^{\infty} x(m)y(m+k) \tag{1}$$

상호상관함수는 신호 x(n)과 y(n)사이에서 시간지연으로 유사도를 측정한다. 신호표본과 지연된 다른 신호표본과의 곱의 합에 의해서 두 신호의 유사도를 측정한다. 만약 두 신호가 비슷한 파형이라면 상호 상관 값은 크게 나타난다. 자기상관함수의 범위는 창 함수로 제한되고, 그 합은 총 표본수로 나누어 정규화한다. x(n)과 y(n)의 신호가 같을 때, 식(1)은 자기상관함수의 정의식인 식(2)로 쓸 수 있다. 이 때 자기상관함수는 even함수 ($\phi(k)=\phi(-k)$)이고, $k=0$일 때 최대 값을 갖으며, $\phi(0)$은 이 신호의 에너지이다.

$$\phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \tag{2}$$

단 구간 자기상관함수는 창함수를 통과한 음성신호를 식(3)으로 표현된다.

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m-k)w(n-m+k) \tag{3}$$

여기서, 신호 x(n)과 지연신호 x(n-k)의 곱은 w(n), w(n+k)의 필터를 통과된 것이다.

$$h_k(n) = w(n)w(n+k) \tag{4}$$

식(3)을 다시 쓰면 다음과 같이 표현가능하다.

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)x(m-k)h_k(n-m) \tag{5}$$

2-2. AMDF 방법

AMDF(Average Magnitude Difference Function)방법은 자기 상관 법에서의 x(m)과 x(m-k)곱 대신에 다음과 같이 절대 값으로 정의된다.

$$AMDF(k) = \sum_{m=-\infty}^{\infty} |x(m) - x(m-k)| \tag{6}$$

AMDF 방법은 자기 상관 법에 수행하는 곱 연산을 절대 값과 차분으로 대신하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용된다. 자기 상관 법에서는 피치주기에 최대 값을 이루지만 AMDF 법에서는 피치 주기 배수에 최소 값을 갖는다.

2-3. AMDF 방법

AMDF법에서는 피치주기 배수에 최저 값을 이루지만 NAMDF를 사용 시 피치주기 배수에 거의 영 값을 갖게 된다. 현재 프레임의 피치를 측정하는 방법으로는 다음과 같이 NAMDF를 정의하여 사용할 수 있다.

$$NAMDF(d) = \frac{\sum_{n=1}^{N-1} |s(n) - s(n-d)|}{\sum_{n=1}^{N-1} |s(n)| + |s(n-d)|} \tag{7}$$

여기서 s(n)은 음성신호이고 N은 NAMDF를 구하려는 윈도우 구간이다. 지연인자 d를 점차 증가시키면서

NAMDF를 구해보면, 지연인자가 프레임 내 음성피치에 정수배가 될 때마다 NAMDF는 거의 영이 된다. AMDF는 곱셈을 사용하지 않는 장점이 있다. 단, 규준화 시 한 번의 나눗셈은 전체 계산량에 커다란 영향을 주지 않기 때문에 NAMDF의 장점을 유지할 수 있다[1,2,5].

3. 제안한 방법

그림 1은 본 논문에서 제안한 피치 검색 방법을 나타낸 블록도이다. 즉, 제안한 방법에서는 G.723.1 부호화기에서 피치 검색을 위해 사용하는 자기상관법을 대신하여 NAMDF 방식과 skipping 기법을 적용하여 계산량 감소를 도모한다. 본 논문에서 제안한 고속 피치 검색 방법에서 NAMDF 알고리즘은 2-3절에 설명되어졌다. 또한 계산량을 감소하기 위하여 NAMDF 방법에 추가된 skipping 기법의 특징은 다음과 같다. 우선 여러 종류의 음성시료의 각 프레임별의 특성을 통계 내어 문턱 값을 구하였다. 즉, 인덱스를 하나씩 증가하면서 NAMDF 방법을 사용하여 피치 검색을 하는 과정에 있어서 문턱 값보다 작은 경우의 봉우리에 해당하는 인덱스의 값을 카운트 한 다음 그 수만큼 양의 봉우리에 해당하는 영역에 대해서는 피치 검색을 수행하지 않는다. 또한 두 번째 문턱 값보다 큰 영역에 대해서도 피치 검색을 수행하지 않고 skipping하게 된다. 이는 식(7)에 나타났듯이 NAMDF 알고리즘을 사용하면 피치 주기에 해당된 경우에 있어서 최소 값이 얻어지기 때문에, 이 특성과 일반적 음성신호 특성에 맞추어 새로운 피치 검색방법을 제안하였다.

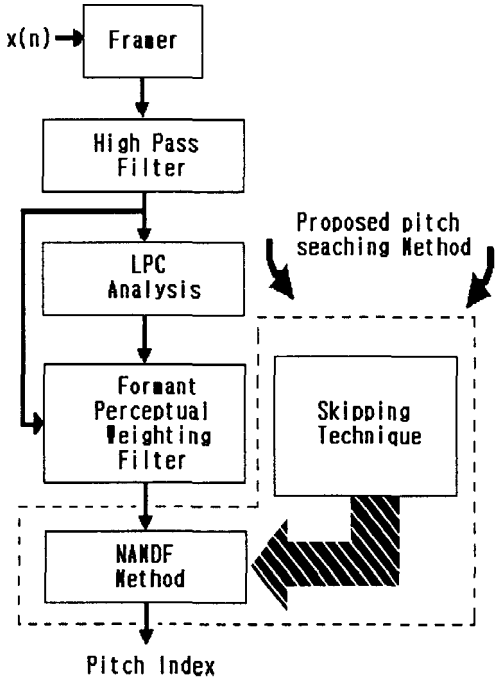

그림 1. 제안한 알고리즘의 블록도

4. 실험 및 결과

제안한 방법의 시뮬레이션을 위해 사용한 음성시료는 뉴스에서 발췌한 일기예보 남/녀 아나운서 음성시료이고, 시료에 대해 한 프레임의 길이를 240 샘플로, 서브프레임을 60샘플 단위로 처리하였다[4].

기존의 방법과 제안한 방법간의 피치 검색 시 계산량을 비교하기 위하여 Visual-C++을 사용하였다.

표 1은 자기상관함수법을 사용한 경우와 제안한 방법에 있어서의 계산량을 비교한 실험 결과이다.

실험에서 사용되어진 음성은 전체 306프레임이고 표 1의 실험결과에서는 프레임 당 연산횟수를 비교하여 나타내었다. 실험에서 얻어진 프레임 당 계산량 감소율은 가/감 연산의 경우 약 29%이고 승/제 연산의 경우에 있어서 약 99.6%로 나타났다. 제안한 알고리즘의 경우에 있어서 계산량 감소율은 전체 약 64.3%로 나타났다. 그림 2는 제안한 알고리즘을 G.723.1 부호화기에 적용한 실험결과이다. 그림 2 (a)는 원 음성신호의 파형을 나타내고 (b)는 원 음성신호의 피치 contour을 나타내었다. (c)는 기존의 G.723.1 부호화기를 통과한 합성음의 피치 contour이며 (d)는 제안한 알고리즘이 적용된 부호화기 합성음의 pitch contour이다. 그림 2에서는 원 음성, 기존의 부호화기를 통과한 합성음과 제안한 알고리즘을 적용한 부호화기의 합성음 파형의 유사성을 관측할 수가 있다.

5. 결 론

본 논문에서 제안한 알고리즘은 NAMDF 방법과 skipping 기법을 이용하였다. 계산량 감소율 측면에서는 약 64%의 감소율을 보였고 기존의 방법과 제안한 방법간의 피치 pitch contour은 거의 유사하였다. 부가적으로 실시된 음질과 신호 대 잡음 비에서도 기존의 방법과 제안한 방법간에는 유사한 실험 결과가 얻어졌다. 즉, 제안한 방법은 계산량을 감소시키면서 다른 열화되는 요소를 갖지 않기 때문에 CELP 부호화기에 광범위하게 적용되어질 수 있다.

표 1. 제안한 알고리즘의 계산량 감소율

비교 / 연산	계산량		계산량 감소율	감소율 합계
	자기상관법	제안한 방법		
+ & −	170163.6	120883.1	29%	64.3%
× & ÷	170928.7	604.3	99.6%	
처리 시간	0.93 sec	0.57 sec		

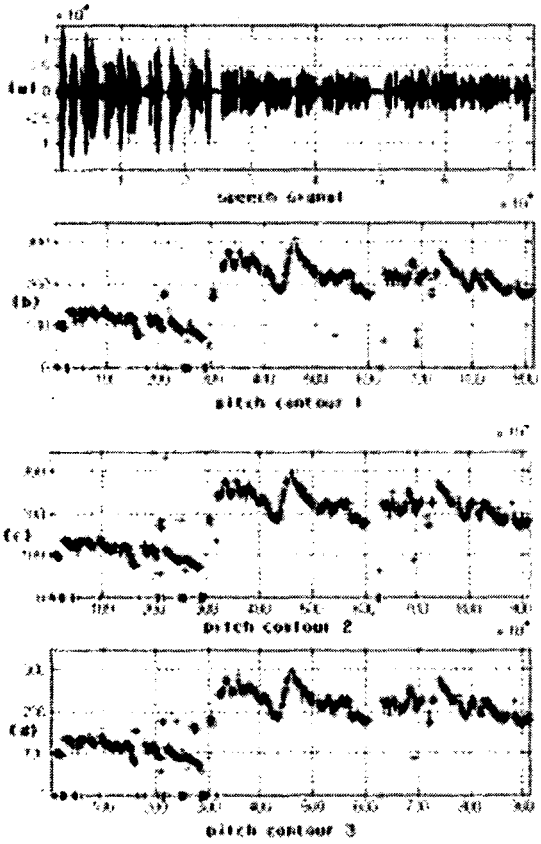

그림 2. 실험에서 얻어진 피치 contour
(a)원음성 신호
(b)원음성 신호의 피치 contour
(c)G.723.1 부호화기 합성음의 피치 contour
(d)제안한 부호화기 합성음의 피치 contour

감사의 글

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

참고문헌

[1] N. S. Jayant and P. Noll, Digital Coding of Waveform-Principles and Applications to Speech and Video, pp.220-221, Prentice-Hall, 1978.

[2] B. Atal, "Efficient coding of LPC parameter by temporal decomposition", Proc. of ICASSP, pp.81-85, 1983.

[3] A. M. Kondoz, Digital Speech, John Willy & Sons, 1994.

[4] L.R. Rabiner and R.W. Schafer, "Digital Proce-ssing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.

[5] W. B. Kleijn et al., "Fast Methods for the CELP Speech Coding Algorithm", IEEE Trans., Acoustics, Speech and Signal Processing, Vol.38, No.8, pp.1330-1341, August 1990.

[6] J. H. LEE, H. Y. JEON, M. BAE, and S. ANN, "A Fast Pitch Searching Algorithm using Correlation Characteristics in the CELP Vocoder", IEEE Communication Society, Proceeding of MILCOM '94, pp.699-702, Oct.2-5 1994.

[7] D. KIM, M. BAE, J. KIM, K. BYUN, K. HAN, H.YOO, "On a Reduction of Pitch Searching Time by Preprocessing in the CELP Vocoder", J. Acoust. Soc, Korea, Vol.13, No.3, pp.33-40, June 1994.