韓國 컴퓨터情報學會 論文誌
第10卷 第6號, 2005. 12.

2005-10-6-2-4

CELP 부호화기에서 가변 윈도우 스펙트럼 분석에 의한 성능 향상에 관한 연구

민 소 연*, 김 은 환**, 배 명 진***

A Study on an Improvement of the Performance by Spectrum Analysis with Variable Window in CELP Vocoder

So-Yeon Min*, Eun-Hwan Kim**, Myung-Jin Bae***

요 약

CELP 계열 음성부호화기는 대략 4.8kbps의 전송률로 음질 면에서 우수한 특성을 지닌다. 이중에서 인터넷폰과 화상회의를 위해 개발되어진 G.723.1 음성부호화기는 5.3kbps ACELP와 6.3kbps MP-MLQ 이중 전송률로 구성된다. 본 논문에서는 CELP 계열 음성부호화기의 음질 개선을 위해 가변 윈도우를 이용한 새로운 스펙트럼 분석 알고리즘을 제안한다. 기존의 방법에서는 스펙트럼 분석시 고정 윈도우를 사용하기 때문에 합성음의 스펙트럼의 경우 왜곡이 발생한다. 그러므로 스펙트럼 누설을 최소화하기 위해 조절되어진 윈도우를 사용하였다. 제안한 알고리즘을 G.723.1 ACELP에 적용한 실험결과에서 스펙트럼 왜곡은 0.084dB 정도 감소, 잔차 에너지는 6.3% 감소하였으며 이에 반해 MOS는 0.1 정도 개선되었다.

Abstract

In general CELP(Code Excited Linear Prediction) type vocoders provide good speech quality around 4.8kbps. Among them, G.723.1 developed for Internet Phone and video-conferencing includes two vocoders, 5.3kbps ACELP(Algebraic-CELP) and 6.3kbps MP-MLQ(Multi-Pulse Maximum Likelihood Quantization). In order to improve the speech quality in CELP vocoder, in this paper, we proposed a new spectrum analysis algorithm with variable window. In CELP vocoder, the spectrum of the synthesised speech signal is distorted because the fixed size windows is used for spectrum analysis. So we have measured the spectral leakage and in order to minimize the spectral leakage have adjusted the window size. Applying this method G.723.1 ACELP, we can get SD(Spectral Distortion) reduction 0.084(dB), residual energy reduction 6.3% and MOS(Mean Opinion Score) improvement 0.1.

▸ Keyword : Variable Window, CELP Vocoder, Fixed Size Window, Spectrum Leakage

• 제1저자 : 민소연
• 접수일 : 2005.10.15, 심사완료일 : 2005.12.02
* 서일대학 정보통신과 전임강사, ** 숭실대학 전자계산원 교수, *** 숭실대학 정보통신 전자공학부 교수

Ⅰ. 서 론

표준 음성 부호화기는 1972년 ITU-T 권고안 G.711로 채택된 64kbps PCM(Pulse Coded Modulation) 방식으로부터 출발하여 32kbps의 ADPCM(adaptive DPCM), 16kbps의 LD-CELP (Low-Delay CELP) 방식으로 표준화되었다.

현재 ITU-T에서는 PCS, IMT-2000등에서 사용할 수 있는 8kbps 음성 부호화기에 대한 표준화 작업으로 1996년에 CS-CELP (Conjugated Structure algebraic CELP)를 G.729로, 그리고 인터넷-폰 및 화상 통신용 음성 부호화기로 ACELP/MP-MLQ (Algebraic CELP /Multi-pulse Maximum Likeli-hood Quantization)의 5.3/6.3kbps dual rate를 G.723.1 권고안으로 선정하였다[1].

그리고 현재 국내에서는 유선망을 통한 화상회의를 목적으로 표준화된 G.723.1 음성부호화기 사용하여 인터넷폰이나 화상회의에 응용하기 위해 많은 연구가 이루어지고 있다. 전화는 64kbps이지만 인터넷을 사용하는 경우, 상황이 나쁘면 음성이 단절되거나 잡음이 섞이므로 안정적인 통화를 보장할 수 없어 전송률을 낮추는 기술과 음질을 향상시키는 기술이 매우 중요하다[1,9,10].

CELP 부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있다[2]. 그리고 CELP 부호화기에서는 음성 신호의 스펙트럼을 LPC 분석을 통해 부호화하는데 고정 윈도우를 사용하여 부호화 한다.

그러나 음성신호의 유성음은 주기성을 가지고 있고 시간에 따라 주기성이 가변적이다. 그런데 고정 윈도우를 적용하면 스펙트럼 왜곡이 생기게 된다.

따라서 본 논문에서는 스펙트럼 왜곡을 최소화 할 수 있는 가변 윈도우를 사용하여 부호화하는 새로운 알고리즘을 제안한다. 본 논문의 구성은, 2장에서 CELP 음성부호화기의 원리에 대해 설명하고 3장에서는 제안한 알고리즘에 대한 설명이 이루어진다. 그리고 4장에서는 실험결과, 5장에서는 결론을 맺는다.

Ⅱ. CELP 음성부호화기의 원리

2.1 개요

CELP(Code Excited Linear Prediction) 부호화기는 코드북 내에 저장된 입력 여기 신호열을 두 개의 시변 선형 회귀(Time-varing Linear Recursive) 필터를 통과시킴으로써 얻은 신호 중 주어진 충실도 판정을 최적화 시키는 것을 선택하도록 구성되어 있다[2]. 이러한 CELP 부호화기는 입력으로 얻어진 음성신호를 분석하여 필요한 파라미터를 추출하고 이를 이용, 음성신호를 합성하여 입력음성 신호와 비교하는 소위 합성에 의한 분석(Analysis By Synthesis)법을 사용함으로써 음질이 매우 우수하다. (그림 1)은 CELP 부호화기의 기본 구조를 나타낸 것이다[2]. (그림 1)에서 s(n)은 입력된 음성 신호이며 x(n)은 코드북 내에 저장되어 있는 입력 여기 신호열을 나타낸다. 이 입력 여기 신호열 x(n)을 b_k를 이용하여 적절하게 크기를 조절한 후, 두 시변 회귀필터 1/P(z), 1/A(z)을 통과시키면 합성 음성신호 $\hat{s}(n)$을 얻는다. 두 음성신호 s(n)과 $\hat{s}(n)$과의 차이를 오차 가중 필터 W(z)에 통과시키면 오차 신호인 e(n)을 얻게 되는데, 이 e(n)을 평균 제곱오차를 이용하여 비교함으로써 가장 적은 오차 신호를 나타내는 x(n)을 얻게 된다. 이때, P(z), A(z)의 계수는 입력된 음성 신호를 이용하여 구하고, 코드북 내의 각 원소들은 백색 가우시안 불규칙 수열들로부터 구성한다.

2.2 스펙트럼 필터

포만트 예측기 또는 단기 예측기(Short-term Predictor)라고도 불리는 스펙트럼 필터는 보통 10차의 차수를 갖는 LPC 계수를 얻기 위하여 자기 상관방법을 사용하게 된다. 다음은 본 논문에 사용된 G.723.1 ACELP의 LPC(Linear Predictive Coding) 분석에 대한 것이다.

G.723.1 ACELP의 LPC 분석은 60샘플의 4개의 부-프레임에서 10차의 선형예측분석이 수행된다. 180표본의 해밍 윈도우(Hamming Window)가 각 부-프레임의 중앙에 위치하게 되며 11개의 자기 상관계수가 윈도우 처리된 신호로부터 계산되어진다. 선형예측계수(LPC)는 Levinson-Durbin

recursion을 사용하여 계산되어지며, 모든 입력 프레임의 각 부-프레임마다 하나씩 계산되어 네 집합의 LPC 계수가 계산되어진다. LPC 합성 필터는 다음과 같이 정의된다 [3,7,8].

$$A_i(z) = \frac{1}{1 - \sum_{j=1}^{10} a_{ij} z^{-j}}, \quad 0 \le i \le 3 \quad \cdots\cdots (2.1)$$

여기서 i는 0과 3사이에서 정의되는 부-프레임 인덱스이며 j는 차수를 나타낸다.

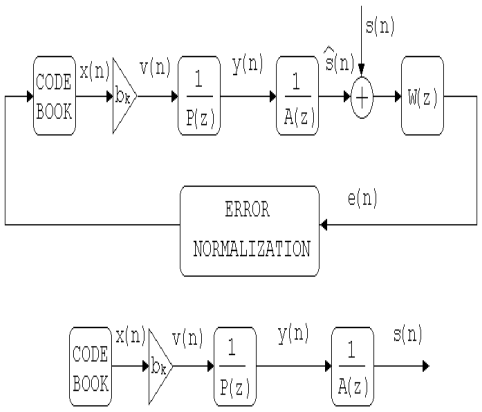

그림 1. CELP 부호화기의 기본 구조
Fig 1. Structure of the CELP Vocoder

III. 제안한 알고리즘

CELP 부호화기의 스펙트럼 필터에서는 고정 크기의 윈도우를 사용한다. 그러나 주기가 변하는 음성신호의 분석에 있어 윈도우의 크기가 고정되어 있으면 스펙트럼 누설현상으로 인한 왜곡이 발생하게 된다. 따라서 본 논문에서는 스펙트럼 왜곡을 최소화하기 위해서 스펙트럼 누설에너지를 구하고 이를 최소화하는 윈도우 크기로 음성신호를 분석하는 알고리즘을 제안한다.

3.1 스펙트럼 누설현상

윈도우 크기를 $N \cdot T$라 하면, $N \cdot T$내에 정수배의 주기를 갖지 않는 신호에 대한 스펙트럼은 원래의 신호주파수 근처에 대칭으로 분산되어 나타나는 현상을 누설현상이라고 한다. 이 현상은 Discrete Fourier Transform이 구간 $N \cdot T$에서 모든 성분이 주기적이라는 가정 하에 Fourier Series 전개를 한 것이기 때문이다[4-6].

음성신호에서 주기성을 갖는 유성음에 대해 일정위치에서 윈도우를 취하면 음성신호의 스펙트럼에 윈도우의 스펙트럼이 컨볼루션(convolution)된 결과가 된다.

$s(\cdot)$, $f(\cdot)$, $w(\cdot)$를 각각 음성신호 및 윈도우가 적용된 음성신호 그리고 윈도우라 할 때 시간영역에서는 식(3.1)과 같이 나타낼 수 있다.

$$f(n) = s(n) w(n-k) \quad \cdots\cdots (3.1)$$

주파수영역에서는 식(3.2)로 표현가능하다.

$$F(e^{j\Omega T}) = S(e^{j\Omega T}) * W(e^{j\Omega T}) \quad \cdots\cdots (3.2)$$

이상적인 윈도우라면 주파수 영역에서 컨볼루션 되어도 원래의 음성 스펙트럼에 근사 되도록 나타날 것이다. 그러나 윈도우의 크기가 음성신호의 주기인 피치의 정수배가 되지 않으면 누설현상이 발생하게 된다. 그렇지만 윈도우의 크기를 피치의 정수배로 적용하기 위해서는 피치를 구하는 일이 선행되어야 하는 어려움이 따르게 된다.

3.2 누설 에너지 검출

윈도우에 의해 발생되는 스펙트럼 누설현상을 최소화하기 위해서 본 논문에서는 윈도우의 크기에 따른 스펙트럼 누설에너지를 시간영역에서 측정하여 그 변화가 최소가 되는 윈도우의 크기를 최적의 윈도우 크기로 결정하였다.

n번째 샘플부터 I 크기의 윈도우를 적용한 음성신호 의 에너지는 식(3.3)과 같이 표시 할 수 있다. 식(3.3)에서 s(n)은 음성신호이므로, s(n)s(n)은 에너지에 해당한다.

$$E_I(n) = \sum_{k=-\infty}^{\infty} s(n) s(n) w_I(n-k) \quad \cdots\cdots (3.3)$$

그리고 윈도우의 시작위치를 이동하면서 식(3.3)을 계산하면 위상이 달라지므로 에너지 값이 다르게 나타난다. 에너지의 변화를 측정하기 위해 rectangular 윈도우를 한 샘플씩 k만큼 이동하면서 적용시킨다. 그리고 구해진 값들의 최대 및 최소값을 구하면 다음과 같다.

$$MaxE_l = Max[E_l(n), E_l(n-1), \cdots, E_l(n-k+1)] \quad \cdots$$

$$MinE_l = Min[E_l(n), E_l(n-1), \cdots, E_l(n-k+1)] \quad \cdots\cdots (3.4)$$

최대에너지와 최소에너지 차이를 $E_{diff}(l)$이라 하면 다음과 같다.

$$E_{diff}(l) = MaxE_l - MinE_l \quad \cdots\cdots (3.5)$$

윈도우의 시점을 달리하면서 위상에 따라 변하는 에너지 차이를 구해서 그 값을 스펙트럼 누설에너지에 근사한 값으로 한다. 그리고 이 값이 가장 작을 때의 윈도우 크기를 이때의 최적 윈도우 크기로 결정한다.

$$l_{opt} = Min[E_{diff}(l)] \quad \cdots\cdots (3.6)$$

l_{opt}는 최적의 윈도우 크기를 나타낸다.

Ⅳ. 실험 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 586에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 처리결과의 성능을 측정하기 위해 다음의 음성시료를 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 사용하였다. 음성 시료는 SNR이 30dB인 환경 하에서 녹음하였다. 실험에서 사용한 발성은 유성음과 무성음이 골고루 사용된 음성시료이다. 음성신호는 개인의 특성에 따라 많은 차이점를 갖고 있으나, 본 실험에서는 음소특성이 중요하게 여겨지므로 다양한 자음과 모음을 반영한 음성시료를 택하였다.

발성1: /인수네 꼬마는 천재소년을 좋아한다./

발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./

발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다./

발성4: /○○대학교 정보통신과 음성통신 연구팀이다./

제안한 알고리즘의 시뮬레이션은 G.723.1 ACELP 부호화기에 C-언어로 구현하여 수행하였다. 제안한 알고리즘의 성능 비교는 G.723.1 Annex A를 통과한 음성과 제안한 알고리즘을 통과한 음성의 SD(Spectral Distortion)와 각 프레임의 잔차-신호(residual)의 에너지를 비교하고, 전체 합성음의 MOS(Mean Opinion Score)를 측정하였다. 식 4.1은 SD를 계산한 식이다.

$|H(e^{jw})|$는 원 음성의 스펙트럼 크기이고 $|\hat{H}(e^{jw})|$는 부호화하여 합성한 음성의 스펙트럼 크기이다[6].

$$SD = \frac{1}{2\pi}\int_{-\pi}^{\pi} [10\log|H(e^{jw})|^2 - 10\log|\hat{H}(e^{jw})|^2]^2 dw \quad \cdots (4.1)$$

부호화기는 240샘플 프레임마다 처리한다(8kHz 샘플링 율에서 30ms). 각 프레임은 DC 성분을 제거하기 위해 하이 패스 필터를 통과하고, 60샘플의 4개의 부-프레임으로 나누어지고 모든 부-프레임에서 입력신호를 사용하여 10차의 선형 예측계수를 계산한다. 음성신호의 단구간 분석을 위해 주로 사용하는 윈도우의 크기는 256 샘플의 크기이다.

앞에서 설명하였듯이 부호화기에서는 240 샘플 프레임 단위로 처리를 하는데 4개의 서브 밴드로 나뉘게 된다. 4개의 서브 밴드 마다 LPC 필터를 통과하게 되는데 LPC 필터는 180 샘플을 단위로 해서 처리를 한다.

그러므로 LPC 필터 크기는 180샘플을 기준으로 하여 ±10, ±20으로 중첩되는 160, 170, 180, 190, 200 크기의 윈도우를 사용하게 된다. 음성신호를 처리함에 있어서 윈도우를 사용하는 경우 중첩을 하여 사용한다. 부호화기의 블록도는 (그림 2)에 나타내었다.

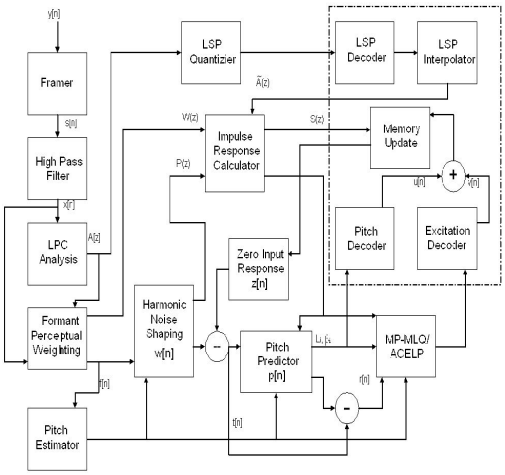

그림 2. G.723.1 음성 부호화기 블록도
Fig 2. Block-diagram of the G.723.1 Vocoder

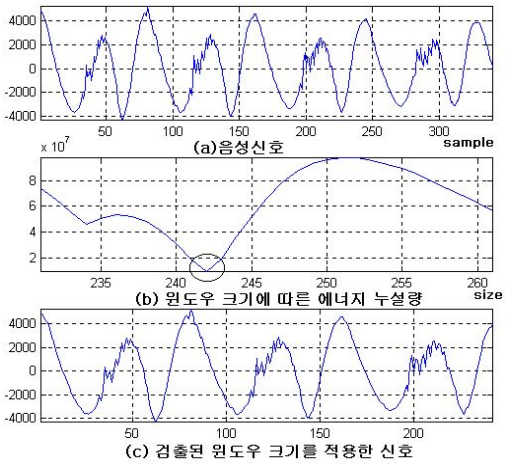

그림 3. 최적의 윈도우 크기 결정
Fig 3. Decision of optimized window size

(그림 3)에서는 최적 윈도우 크기를 결정한 실험결과를 나타내었다. 즉, (a)는 원 음성 신호, (b)는 윈도우 크기를 변화시켜서 에너지 누설량을 비교한 정도를 나타낸다. (c)는 결정된 윈도우를 사용하여 음성신호를 단구간 분석 결과를 나타내고 있다.

<표 1>에서는 기존에 방법과 비교하여, 제안한 알고리즘을 G.723.1 음성부호화기에 적용한 경우 Spectral Distortion이 평균 0.084(dB) 개선되었음을 알 수가 있다.

<표 2>에서는 제안한 알고리즘을 음성부호화기에 적용한 경우, 잔차-신호 에너지가 평균 6.3% 줄어들었음을 나타낸다. 또한 MOS test 결과, 평균 0.1이 높아졌음이 <표 3>에 나타나있다. 일반적으로, 실험에서 사용한 음성신호의 프레임 수가 각기 다르고, 유성음이 무성음에 비해서 에너지가 큰 특성으로 인하여, 스펙트럼 왜곡, 잔차 신호의 에너지 감소량, MOS 결과를 평균값으로 나타내었다.

즉, 제안한 알고리즘에 따른 실험결과를 순차적으로 설명하면 (그림 3)에서는 윈도우 크기를 가변으로 하여 윈도우 크기를 변화 시켰을 때 에너지 누설량에 따라 최적 윈도우 크기를 찾아낸 결과를 나타내었다.

그리고, <표 1, 2, 3>에서는 각각, 고정 윈도우 크기를 사용한 경우와 찾아진 가변 윈도우를 사용한 경우의 스펙트럼 왜곡과 잔차 신호 에너지, MOS test 결과를 비교하였다.

표 1. SD (Spectral Distortion) 비교
Table 1. Analysis of spectral distortion

	G.723.1	제안한 알고리즘	증가율 (dB)
발성 1	0.795	0.762	0.033
발성 2	0.795	0.730	0.065
발성 3	0.812	0.688	0.124
발성 4	1.820	1.706	0.114
Total	0.084 (dB)		

표 2. 잔차신호의 에너지 감소량
Table 2. Energy reduction of residual signal

	발성1	발성2	발성3	발성4
1 - E2/E1	8.51 (%)	4.19 (%)	7.99 (%)	4.51 (%)

*E1 : G.723.1 부호화기의 잔차-신호 에너지
*E2 : 제안한 부호화기의 잔차-신호 에너지

표 3. MOS(Mean Opinion Score) 결과
Table 3. Experimental result of MOS test

	발성1	발성2	발성3	발성4
G.723.1	3.54	3.61	3.53	3.72
제안한 알고리즘	3.6	3.68	3.71	3.81

V. 결 론

CELP 부호화기는 선형 예측을 통한 합성에 의한 분석을 기본원리로 한 부호화기이다. 그러나 스펙트럼 분석 시 고정 윈도우를 사용하여 분석함으로써 주기성이 가변적인 음성신호에서 스펙트럼 왜곡이 발생한다. 따라서 본 논문에서는 스펙트럼 왜곡을 최소화 할 수 있는 가변 윈도우를 사용하여 스펙트럼을 분석하고 부호화하여 음질을 향상시키는 새로운 알고리즘을 제안한다.

본 논문에서는 제안한 알고리즘을 실험하기 위해 화상회의나 인터넷폰을 목적으로 개발된 G.723.1 부호화기를 사용하였다. 실험결과 Spectral Distortion이 평균 0.084(dB) 개선되었고 잔차-신호 에너지는 평균 6.3% 줄어들었으며, MOS는 평균 0.1이 높아졌다. 향후, 제안한 알고리즘에 여러 나라 음성시료를 적용하여 보다 객관적인 실험을 할 예정이다.

참고문헌

[1] 이미숙, "IMT-2000을 위한 음성부호화연구", 가입자망연구소 정보통신연구 제13권 제1호, 1999.3

[2] A.M. Kondoz, Digital Speech, John Wiley & Sons, 1994.

[3] ITU-T Recommendation G.723.1, March, 1996.

[4] 정찬중, 나덕수, 신동성, 배명진, "스펙트럼 누설에너지를 이용한 음성신호의 창함수 적용에 관한 연구", 한국통신학회, 하계 종합 학술발표회 논문집(상), PP. 487-490, 1999.7.

[5] W. B. Klejin et. al, Speech Coding and Synthesis, Elsevier Science B.V., 1995.

[6] 음성 신호 처리 기술, 한국과학기술원, 삼성첨단기술센터, 제 3권 음성부호화, PP.171

[7] 배명진, 디지털 음성분석, 동영출판사.

[8] 배명진, 디지털 음성부호화, 동영출판사.

[9] 김기영, "IPv6에서 멀티캐스트 지원을 위한 핸드 프 기법", 한국 컴퓨터 정보학회 논문지 제10권, 제4호.

[10] 이홍규, "IPv6 기반 자동화된 공격 대응도구", 한국컴퓨터정보학회 논문지, 제10권 제3호.

저 자 소 개

민 소 연

1993년 2월 숭실대학교 전자공학과 공학사

1995년 2월 숭실대학교 전자공학과 공학석사

2003년 2월 숭실대학교 전자공학과 공학박사

2005년 3월~현재 서일대학교 정보통신공학과 전임강사

김 은 환

1990년 2월 숭실대학교 전자계산학과 공학사

1997년 8월 숭실대학교 컴퓨터학과 공학석사

2003년 2월 숭실대학교 컴퓨터학과 공학박사

1997년 9월~현재 숭실대학교 전산원 교수

<관심분야> 정보보호, 암호알고리즘, 네트워크, 인터넷보안

배 명 진

1987년 서울대학교 공과대학 전자공학과 졸업(공학박사)

2005년 12월 현재 숭실대학교 정보통신 전자공학부 교수

<관심분야> 음성합성, 음성분석, 음성코딩, 데이터통신 네트워크, 디지털신호처리 등