2000년도 한국음향학회 학술발표대회 논문집 제19권 제1(s)호

전이구간에서의 Energy 평탄화를 통한 피치정확도 향상에 의한 G.723.1 Vocoder의 음질향상

박원, 김정진, 배명진
숭실대학교 정보통신공학과

The Speech Enhancement of G.723.1 Vocoder by the Improvement of Pitch Accuracy Using a Flattened Energy in a Transient Period

Won Park, JungJin Kim, MyungJin Bae
Dept. of Info. &Telecom. Engr., Soongsil Univ. Seoul 156-743, Korea

wpark@assp.ssu.ac.kr, mjbae@saint.soongsil.ac.kr

Abstract

화상회의 및 인터넷폰을 목적으로 개발된 G.723.1은 CELP계열 보코더와 같이 화자의 개성정보를 위해 피치를 전송하고 있다.

하지만 안정구간과 비 안정구간의 차이를 두지 않고 처리를 하기 때문에 비 안정구간, 특히 전이구간에서 정확한 피치검출이 이루어지지 않는 이유로 음질의 열하가 발생하게 된다.

따라서 본 논문에서는 한 프레임 구간에서 에너지의 기울기로 대략적인 피치이득을 구한 다음 안정구간일 때와 프레임 내의 에너지의 기울기가 문턱값을 넘을 때에는 기존의 방법으로 피치를 구하고 그렇지 않은 경우에는 에너지를 조정하여 피치를 다시 구하는 방법을 사용하였다.

실제 음성시료에 대해 기존의 방법과 제안한 방법을 비교하기 위해 SegSNR 과 MOS를 비교하였을 때 각각 1.302(dB)와 평균 0.045 MOS가 향상되었다.

Ⅰ. 서론

G.723.1은 5.3/6.3kbps의 이중 전송률을 갖는 구조로 현재 인터넷폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며 낮은 전송률에 비해 우수한 음질을 제공하고 있다. 더불어 최적의 전송 환경을 위해 두 개의 전송률을 가지고 있기 때문에 다른 보코더 표준안들에 비해 응용성이 높다[1].

G.723.1 부호화기에 적용되고 있는 피치검색법은 피치지연에 따른 상관관계법을 사용하고있다. 상관관계 피치검색법은 피치가 존재하는 모든 피치지연에 대해 두 신호의 상관관계를 검색하여 최상의 상관관계를 피치지연과 이득으로 결정하게 된다[2]. 하지만 기존의 G.723.1보코더 에서는 주기성을 부호화 할 때 피치를 안정구간과 비 안정구간의 차이를 두지 않고 모든 구간을 처리하여 비 안정구간, 특히 전이구간에서 정확한 피치검출이 이루어지지 않기 때문에 음질의 열하가 발생하게된다.

따라서 본 논문에서는 비 안정구간, 특히 전이구간에서 피치를 정확히 구함으로써 음질을 향상시키고자 한다.

Ⅱ. G.723.1 부호화기의 원리

2-1. 개 요

부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있고 인지 가중화된 오차 신호의 최소화를 시도한다. 부호화기는 한 프레임 240샘플마다 처리가 되어 8kHz 샘플링 표본화에서 30ms가 된다[1].

분석을 위해 각 프레임은 60표본씩 4개의 부프레임으로 나누어지고 모든 부프레임에서 10차의 선형 예측 부호화기

필터는 입력신호를 사용하여 계산되어진다[4]. 마지막 부프레임을 위한 선형 예측 부호화기 필터는 예측 분할 벡터 양자화기(predictive split vector quantizer)를 사용하여 양자화된다. 양자화 되지 않은 선형 예측 부화화기의 계수는 인지 가중화된 음성 신호를 얻기 위한 단구간 인지 가중화 필터를 구축한다.

두 개의 부프레임(120표본)을 처리하기 위한 개회로 피치 주기 예측은 인지 가중화된 음성 신호를 사용하여 계산되어진다. 피치 주기는 18에서 142표본의 구간에서 찾으며 피치 예측은 120표본의 블록 단위로 실행되어진다. 이전에 계산되어 예측된 피치 주기를 사용하여 하모닉 노이즈 세이핑 필터(harmonic noise shaping filter)가 만들어지며 LPC 합성 필터(LPC synthesis filter), 포만트 인지 가중화 필터(formant perceptual weighting filter), 하모닉 노이즈 세이핑 필터의 결합은 임펄스 응답을 생성하기 위해 사용된다. 피치 주기 예측값과 임펄스 응답을 이용하여 폐회로 피치 예측기(closed loop pitch predictor)값이 계산되어진다. 5차 피치 예측기가 사용되는데 피치주기는 개회로 피치 예측값 주위의 차이값이 작은 구간에서 계산되어진다. 피치 예측기는 초기 목적벡터(initial target vector)로 부터 값을 뺄 때 이용되는데 피치 주기와 차이값 둘 다 복호화기로 전송되어진다. 마지막으로 여기신호의 비주기 요소가 근사화 되어진다. 부호화기의 블록도는 그림 2-1에서 나타내었다[12].

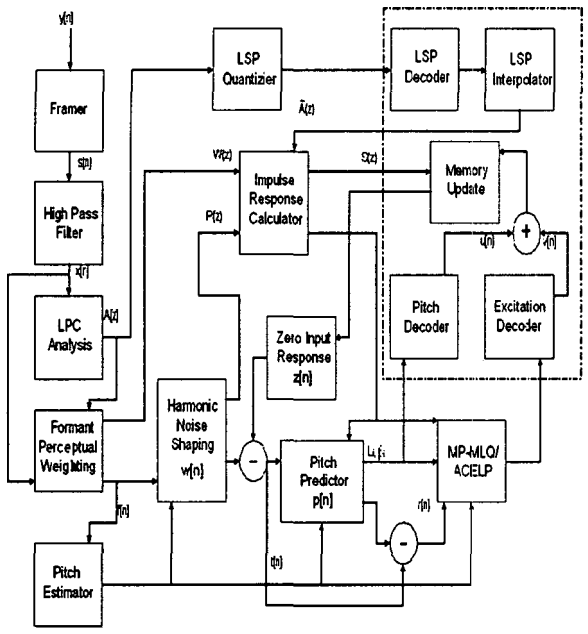

그림 2-1. G.723.1 음성 부호화기 블록도

2-2. 피치검색

피치 예측은 매 프레임당 두 번씩 계산되며, 하나는 처음 두 부프레임에서 다른 하나는 마지막 두 부프레임에서 계산된다. 개회로 피치 주기 예측 L_{OL}은 인지 가중화된 음성신호를 이용하여 계산된다. 최대 상관관계 $C_{OL}(j)$는 피치 주기를 결정하는데 사용되며 다음과 같이 표현된다.

$$C_{OL}(j)=\frac{\left(\sum_{n=0}^{119} f[n]\cdot f[n-j]\right)^2}{\sum_{n=0}^{119} f[n-j]\cdot f[n-j]},\quad 18\le j\le 142 \qquad (2.1)$$

상관관계 $C_{OL}(j)$가 최대가 되는 인덱스 j는 두 부프레임에 적당한 개회로 피치 예측값에서 선택된다.

최적의 인덱스를 찾는 동안 얻게된 선택 값들로부터 피치 배수를 찾게되는 것을 피하기 위해 피치 주기보다 작도록 주어진다. $C_{OL}(j)$의 최대값은 j가 18부터 찾기 시작한다. 모든 최대값 $C_{OL}(j)$을 찾으며 이들 값들은 최적 상태이면서 이전의 최대값 $C_{OL}(j')$과 비교된다. 만약 인덱스 j와 j'의 차이가 18보다 작고 $C_{OL}(j)>C_{OL}(j')$이면 새로운 최대값이 선택된다. 만약 인덱스간의 차이가 18 보다 크거나 같고 $C_{OL}(j')$이 $C_{OL}(j)$보다 1.25dB 를 때는 새로운 최대값이 선택된다.

피치 예측기의 피치 예측 기여(contribution)는 기존의 적응 코드북 기여로 처리한다. 피치 예측기는 5차의 피치 예측기를 사용하며 첫 번째와 세 번째 부프레임에 대해서는 폐회로 피치 지연이 ±1의 범위를 갖는 적당한 개회로 피치 지연값으로부터 결정된다. 두 번째와 네 번째 부프레임에 대해서는 이전 부프레임 지연에서 -1, 0, +1, +2의 차이 값을 갖는 피치 지연값을 사용한다. 이전 부프레임 지연값들에서의 차이값들이 양자화되고 부호화된 피치 지연값은 이점으로부터 L_i로 간주된다. 피치 예측기 이득은 170 엔트리를 갖는 코드북을 사용하여 벡터 양자화를 한다.

Ⅲ. 제안한 방법

기존의 G.723.1에서의 피치 검색은 시간영역에서 규준화된 자기상관관계를 이용하여 모든 프레임의 18부터 142 sample 사이에서 음성을 유성음, 무성음 또는 전이구간에 상관없이 상관관계가 최대가 되는 값을 피치 값으로 선택하였다. 하지만 음소가 변화할 때 나타나는 전이구간에서는 그 주기성이 약하여 피치를 검출하기가 매우 어렵기 때문에 이런 구간에서 에러가 발생하기도 하고 합성음에서 음질의 열하가 발생하기도 한다.

따라서 본 논문에서는 비 안정구간, 특히 전이구간처럼 피치가 일정하지 않은 구간에 대하여 프레임 단위로 에너지의 변화도(기울기)를 구하여 이 기울기를 대략적인 피치 아득으로 보고 이 피치이득 값이 문턱값 내에 있거나 에너지의 기울기가 너무 큰 차이 값을 보이면 기존의 방법을 이용하여 피치이득과 피치값을 구한다. 하지만 그 이외의 구간에서는 에너지를 조정하여 scaling 함으로써 기울기가 대략적으로 1근처에 위치하게 만든 다음에 피치를 검출함으로써 보다 정확한 피치검출을 통해 음질을 향상시키고자 하였다.

아래의 식 (3.2)은 한 프레임을 3등분하여 각각의 에너지를 구하는 식이다.

$$E1 = \sum_{n=1}^{80} S(n)^2$$
$$E2 = \sum_{n=81}^{160} S(n)^2$$
$$E3 = \sum_{n=161}^{240} S(n)^2 \qquad (3.1)$$

다음은 한 프레임 내 에너지의 변화를 알기 위하여 아래의 식 (3.2) 와 같이 E3를 E1으로 나누어 대략적인 기울기를 구한다.

$$Grade = E3/E1 \qquad (3.2)$$

위와 같이 대략적인 에너지의 기울기를 구한 다음 기울기가 1.5 < Grade < 5.5 일 때 즉, 기울기가 증가하고 있을 때와 0.05 < Grade < 0.5 와 같이 기울기가 감소할 때의 원 음성 구간에 scaling factor를 곱해주어 기울기가 1에 가까워지도록 하였다. 다음 식 (3.3)와 (3.4) 은 기울기가 증가할 때와 감소할 때 scaling factor를 구하는 식이다.

$$\rho 1 = \sqrt{E3/E1}$$
$$\rho 2 = \sqrt{E3/E2} \qquad 1.5 < \text{Grade} < 5.5 \qquad (3.3)$$

$$\rho 1 = \sqrt{E1/E3}$$
$$\rho 2 = \sqrt{E1/E2} \qquad 0.05 < \text{Grade} < 0.5 \qquad (3.4)$$

다음 식 3-4는 위에서 구해진 scaling factor를 이용하여 원음성의 프레임이 전이구간일 경우 에너지의 기울기가 증가할 경우와 감소할 경우에 각각의 적용한 예를 보여준다.

$$S'(n) = (s_1(n) \times \rho 1) + (s_2(n) \times \rho 2) + s_2(n)$$
$$1.5 < \text{Grade} < 5.5 \qquad (3.5)$$

$$S'(n) = s_1(n) + (s_2(n) \times \rho 1) + (s_3(n) \times \rho 3)$$
$$0.05 < \text{Grade} < 0.5 \qquad (3.6)$$

s1(n) :	$1 \leq n \leq 80$
s2(n) :	$81 \leq n \leq 160$
s3(n) :	$161 \leq n \leq 240$

아래 그림 3-1은 전이구간에서 에너지의 기울기가 양 일때와 음일때의 원음성과 제안한 방법으로 처리한 파형을 보여주고 있다.

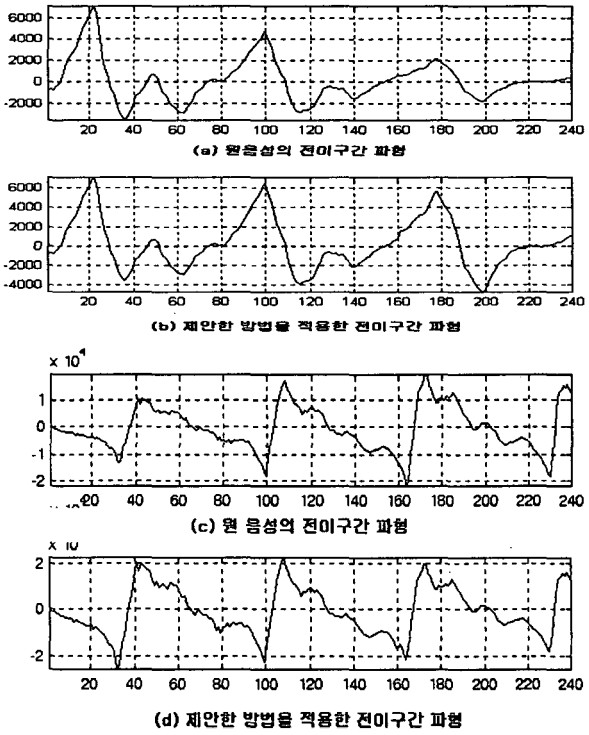

그림 3-1. 전이구간에서 원음성과 제안한 방법을 적용한 파형의 비교

Ⅳ. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 펜티엄-Ⅲ(450MHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 8kHz로 음성을 표본화하였으며 각 시료에 대해 한 프레임의 길이를 240표본으로 하여 처리하였다. 처리결과의 성능을 비교하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자각각 5번씩 발성하여 시료를 사용하였다 :

발성1: /인수네 꼬마는 천재소년을 좋아한다./
발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./
발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다./
발성4: /숭실대학교 정보통신과 음성통신 연구팀이다./
발성5: /May I Help You?./

피치검색은 G.723.1의 피치 검색과정을 C-언어로 구현하여 수행하였다. 성능비교는 기존의 G.723.1의 피치검색과 제안한 방법을 이용한 피치검색을 이용하여 수행하였다.

식 (4.1)은 SegSNR을 구하는 식이다.

$$E_s^2 = \frac{1}{N}\sum_{n=1}^{N} s^2[n]$$

$$E_e^2 = \frac{1}{N}\sum_{n=1}^{N} e^2[n]$$

$$SNR = \frac{E_s^2}{E_e^2}$$

$$SNR[dB] = 10\log_{10}[SNR]$$

$$SNRSEG = \frac{1}{M}\sum_{m=1}^{M} SNR(m)[dB] \qquad (4.1)$$

표 1.은 각 발성에 대해 전이구간 50프레임을 조사하여 G.723.1의 합성음과 제안한 부호화기의 합성음의 SegSNR을 비교한 것이다.

표 1. 전이구간의 SegSNR 비교

	G.723.1	제안한 알고리즘	증가율(dB)
발성 1	13.58	14.55	0.97
발성 2	12.05	14.10	2.05
발성 3	13.99	14.97	0.98
발성 4	12.01	13.29	1.28
발성 5	12.99	14.22	1.23
Total	1.302 (dB)		

표 2. 기존의 G.723.1과 제안된 G.723.1의 MOS SCORE를 비교.

	발성1	발성2	발성3	발성4	발성5
G.723.1	3.86	3.85	3.91	3.87	3.93
제안한 알고리즘	3.91	3.90	3.94	3.91	3.97

V. 결론

최근 화상회의나 인터넷폰 사용자의 증가로 서비스의 질을 높이기 위해 보코더의 음질을 향상시키려는 연구가 활발히 진행되고 있다.

본 논문은 화상회의나 인터넷폰을 목적으로 개발된 G.723.1 부호화기에서 전이구간의 에너지를 평탄화 한 다음 피치를 정확하게 검색함으로써 음질을 향상시키는 알고리즘을 제안한다. G.723.1 부호화기는 유성음에서 피치가 일정한 안정구간과 피치가 변화하는 전이구간의 차이를 두지 않고 처리하여 전이구간에서의 정확한 피치검출이 이루어지지 않는다. 그래서 이러한 문제 때문에 전이구간에서의 음질의 열하가 발생하게 된다. 본 논문에서는 한 프레임에서 에너지의 기울기를 구하여 이 기울기를 대략적인 피치이득이라 보고 전이구간에서 에너지를 평탄화 시켜 1에 근사한 값이 아닐 경우 이 값이 1에 가까운 값을 가지도록 scaling을 통해 에너지의 기울기가 1에 가까워지도록 조절한다. 이러한 방법을 사용하여 정확한 피치값을 구하여 부호화한 결과 전이구간에서의 SegSNR은 평균 1.302(dB)가 개선되었고, MOS는 평균 0.045 가 높아졌다.

VI. 참고 문헌

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

[2] L.R.Rabiner, R.W.Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.

[3] W.B. Kleijn and K.K. Paliwal, Speech Coding and Synthesis, ELSEVIER SCIENCE B.V., pp.6~10, 1995

[4] ITU~T Recommendation G.723.1 March, 1996.

[5] 김정진, 장경아, 배명진, 홍성훈, 심민규, 성유나, "LSP 정보와 에너지 파라미터를 이용한 G.723.1 음성 활동 검출 장치", 신호처리합동학술발표대회 논문집. 1999, 10월.

[6] 나덕수, 노원석, 함명규, 배명진, "피치동기에 의한 음성신호의 전이구간 검출," 한국음향학회, 제 15회 음성통신 및 신호처리 워크샵 논문집, Vol.15, No.1, PP.454-459, 1998년 8월 21-22일

[7] Sharad Singhal and Bishnu S.Atal, "Improving Performance of Multipulse LPC Coders at Low Bit Rates", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1984.