

전이구간에서의 피치보상에 의한 G.723.1 부호화기의 음질 향상 방법

김종국, 배명진

승실대학교 정보통신공학과

mjbae@saint.soongsil.ac.kr

A Improvement Method of the Speech Quality by Pitch Compensation in Transition Region in G.723.1

JongKuk KIM, MyungJin BAE

Dept. of Info. & Telecomm. Engr., Soongsil Univ.

mjbae@saint.soongsil.ac.kr

요약

G.723.1 부호화기는 음성신호의 주기성을 피치와 피치 이득계수로, 스펙트럼 정보를 LSP(Line Spectrum Pair)로 부호화하고 있다. 그런데 주기성을 부호화 할 때 유성음의 피치가 일정한 안정구간과 피치가 변화하는 전이구간의 차이를 두지 않고 처리하여 전이구간에서의 정확한 피치검출이 이루어지지 않는다. 이러한 처리 때문에 전이구간에서의 음질의 열화가 발생하게 된다.

본 논문에서는 전이구간의 피치검출의 정확성을 높여 음질을 향상시킬 수 있는 새로운 알고리즘을 제안한다. 먼저 G.723.1 부호화기에서 검출되는 피치 이득계수를 이용하여 안정구간의 피치 이득계수의 문턱값을 정한다. 그리고 피치 이득 계수가 문턱값을 넘는 부분에 한하여 구해진 피치를 전후 10샘플을 조절하여 피치 이득계수를 다시 구하여 문턱값에 가장 가까운 값을 대표 피치 이득계수로 정하고 그때의 피치와 함께 부호화 한다.

실험 결과 평균 0.6(dB) segmental SNR(Signal to Noise)과 평균0.12 MOS가 향상되었다.

1. 서론

G.723.1 음성 부호화기는 5.3kbps의 ACELP (Algeb

raic Code Excited Linear Prediction)과 6.3kbps의 MP-MLQ(Multi PulseMaximum Likelihood Quantization)의 이중 구조를 가진다[1].

G.723.1 부호화기는 피치필터를 적용하여 음성신호의 피치주기 성분을 부호화한다. 이러한 피치필터에 주로 적용되고 있는 피치검색법은 피치지연에 따른 상관관계법이다. 상관관계 피치검색법은 피치가 존재하는 모든 피치지연에 대해 두 신호의 상관관계를 검색하여 최상의 상관관계를 피치필터의 피치지연과 이득으로 결정하게 된다[2].

그런데 유성음에서 주기성을 부호화 할 때 피치가 일정한 안정구간과 피치가 변화하는 전이구간의 차이를 두지 않고 처리하여 전이구간에서의 정확한 피치검출이 이루어지지 않는다. 이러한 처리 때문에 전이구간에서의 음질의 열화가 발생하게 된다.

본 논문에서는 전이구간의 피치검출의 정확성을 높여 음질을 향상시킬 수 있는 새로운 알고리즘을 제안한다.

2. G.723.1 부호화기의 원리

2-1. 개요

부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있고 인지 가중화된 오차 신호의 최소화를 시도한다. 부호화기는 한 프레임 240 샘플마다 처리한다.

8kHz 샘플링율에서 30ms이다[1].

피치와 피치이득계수는 개회로(Open Loop) 예측과 폐회로(Close Loop) 예측의 두 과정으로 나누어 처리한다. 개회로 예측은 두 개의 부 프레임(120 samples)에서 인지 가중화된 음성 신호를 사용하여 계산한다. 피치주기는 18에서 142표본의 구간에서 찾으며 피치 예측은 120표본의 블록 단위로 실행한다[1].

폐회로 예측을 하기 위해 이전에 계산되어 예측된 피치 주기를 사용하여 하모닉 노이즈 셰이핑 필터(harmonic noise shaping filter)를 만들고, LPC 합성 필터(LPC synthesis filter), 포먼트 인지가중화 필터(formant perceptual weighting filter), 하모닉 노이즈 셰이핑 필터를 결합하여 임펄스 응답을 생성한다. 그리고 피치 주기 예측 값과 임펄스 응답을 이용하여 폐회로 피치 예측(closed loop pitch predictor)값이 계산되어진다. 피치 예측계수는 5차로 사용된다. 부호화기의 블록도는 그림 2-1에서 나타내었다.

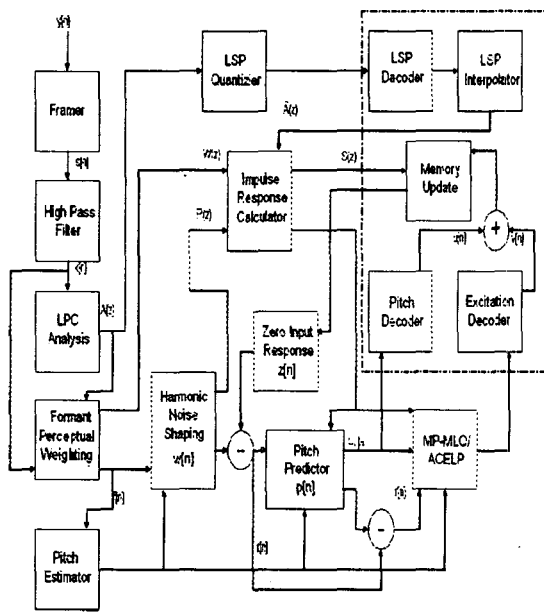


그림 2-1. G.723.1 음성 부호화기 블록도

2-2. 피치검색

개회로 피치주기 예측, L_{OL} 은 인지 가중화된 음성신호를 이용하여 계산된다. 최대 상관관계 $C_{OL}(j)$ 는 피치 주기를 결정하는데 사용되며 다음과 같이 표현된다[1].

$$C_{OL}(j) = \frac{\left(\sum_{n=0}^{119} A_n \cdot A_{n-j} \right)^2}{\sum_{n=0}^{119} A_{n-j} \cdot A_{n-j}} \quad , 18 \leq j \leq 142 \quad (2.1)$$

상관관계 $C_{OL}(j)$ 가 최대가 되는 인덱스 j 는 두 부프레임에 적당한 개회로 피치예측 값에서 선택된다. 최적의 인덱스를 찾는 동안 얻어진 선택 값들로부터 피치 배수를 찾게 되는 것을 피하기 위해 피치 주기보다 작도록 주어진다. $C_{OL}(j)$ 의 최대값은 j 가 18부터 찾기 시작한다. 모든 최대값 $C_{OL}(j)$ 을 찾으며 이들 값들은 최적 상태이면서 이전의 최대값 $C_{OL}(j')$ 과 비교된다. 만약 인덱스 j 와 j' 의 차이가 18보다 작고 $C_{OL}(j) > C_{OL}(j')$ 이면 새로운 최대값이 선택된다. 만약 인덱스간의 차이가 18보다 크거나 같고 $C_{OL}(j)$ 이 $C_{OL}(j')$ 보다 1.25dB 클 때는 새로운 최대값이 선택된다 [1].

폐회로 피치 예측에서는 적응 코드북을 사용한다. 피치 예측기는 5차의 피치 예측기를 사용하며 첫 번째와 세 번째 부프레임에 대해서는 폐회로 피치이득 ± 1 의 범위를 갖는 적당한 개회로 피치이득 값으로 결정한다. 두 번째와 네 번째 부프레임에 대해서는 이전 부프레임 지연에서 -1, 0, +1, +2의 차이 값을 갖는 피치이득 값을 사용하고 이전 부프레임 지연 값들에서의 차이 값들을 양자화한다. 피치 예측기 이득은 170 엔트리를 갖는 코드북을 사용하여 벡터 양자화를 한다. 그리고 피치 예측기의 여기원은 17비트 대수 코드북인 고정 코드북 여기원을 사용한다[1].

3. 제안한 알고리즘

음성신호는 유성음에서 비교적 주기성이 강하게 나타난다. 그러나 음소가 변화할 때 나타나는 전이구간에서는 그 주기성이 약하여 피치를 검출하기가 매우 어렵다[3]. 그래서 피치검출 예러가 대부분 이러한 전이구간에서 발생하게 된다. G.723.1 부호화기에서는 전이구간에 대한 별도의 처리가 없다. 그래서 이러한 전이구간에서 합성음의 음질 열화가 두드러진다.

그림 3-1.은 남성 아나운서의 음성신호에 대한 G.723.1 부호화기의 피치예측 값과 피치이득 값을 나타낸 것이다. 그림(b)에서 안정된 구간의 피치예측 값은 비교적 부드럽게 이어지는데 반해 전이구간에서는 급격하게 변화하는 것을 볼 수 있다. 그림(c)의 피치이득 값도 피치예측 값이 부드럽게 연결되는 안정된 구간에서는 1에 가까운 값으로 나타나지만 피치예측 값이 급격히 변화하는 전이구간에서는 1에서 크게 벗어나는 값으로 나타남을 알 수 있다. 따라서 본 논문에서는 전이구간처럼 피치가 일정하지 않은 구간에 대하여 피치이득 값을 조정하여 피치를 보다 정확히 검출하여 음질을 향상시키는 알고리즘을 제안하고자 한다. 제안한 알고리즘은 먼저 G.723.1 부호화기의 개회로 피치예측기에서 구해진 피치 값

을 이용하여 피치이득 값을 구한다. 피치지연은 개화로 피치예측값의 ± 10 샘플로한다.

식 3.1은 1tap 피치 필터를 나타낸다[2]. 여기서 β 는 τ 를 피치예측 값으로 하는 피치이득 값이다[2][5]. 우리는 τ 를 개화로 피치예측 값으로 하고 β 값을 식 3.2, 3.3과 같이 계산하였다.

$$F(z) = \frac{1}{1 - \beta z^{-(U+\tau)}} \quad (3.1)$$

$$R(\tau + i, 0) = \sum_{m=0}^{N-1} r(m - \tau - i)r(m) \quad (3.2)$$

$$V(i, j) = \sum_{m=0}^{N-1} r(m - \tau - i)r(m - \tau - j)$$

$$\beta = \frac{R(\tau, 0)}{V(0, 0)} \quad (3.3)$$

$$= \frac{\sum_{m=0}^{N-1} r(m)r(m+\tau)}{\sum_{m=0}^{N-1} r^2(m)}$$

먼저, 피치이득 값 β 를 계산하여 $0.75 < \beta < 1.25$ 이면 안정구간으로 간주한다. 그리고 에너지가 묵음구간 보다 크면서 β 값이 $0.75 < \beta < 1.25$ 범위를 벗어나는 프레임은 구해진 피치예측 값을 변화시키면서 β 값을 계산한다.

새롭게 계산된 피치이득 값 β 가 $0.75 < \beta < 1.25$ 에 들어오면 그때의 피치예측 값을 폐회로 피치예측기로 전달한다. 만일, β 가 위의 범위에 들어오지 않을 경우 1에 가장 가까운 피치이득 값이 계산되는 지연 값을 새로운 피치예측 값으로 결정한다.

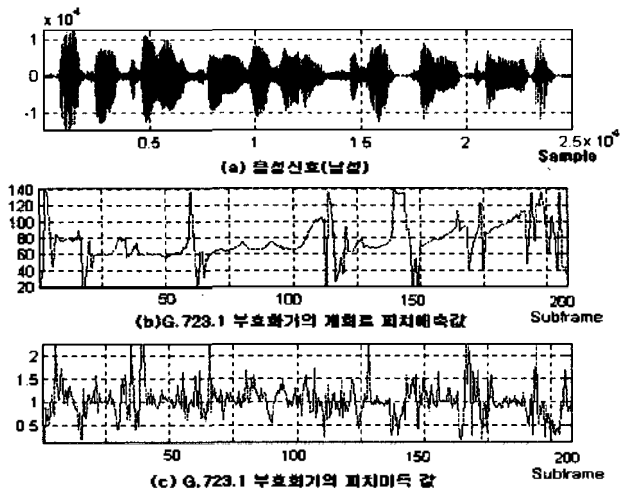


그림 3-1. G.723.1 부호화기의 피치예측 값과 피치이득 값

4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC(586)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 처리결과 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 음성시료는 SNR이 30dB인 환경하에서 녹음하였다. 음성시료는 다음과 같다.

- 발성1: /인수배 꼬마는 천재소녀를 좋아한다./
- 발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./
- 발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다./
- 발성4: /송실대학교 정보통신과 음성통신 연구팀이다./

제안한 알고리즘의 시뮬레이션은 C-언어로 구현하여 수행하였다. 성능 비교는 G.723.1 Annex A를 통과한 음성과 제안한 알고리즘을 통과한 음성의 SNRSEG (segmental SNR)을 비교하고, 전체 합성음의 MOS 측정하였다. 식 4.1은 SNRSEG를 계산한 식이다. $s(n)$ 는 원 음성신호이고 $\hat{s}(n)$ 는 합성한 신호이다[7].

$$s(n) = s(n) - \hat{s}(n)$$

$$E_s = \frac{1}{N} \sum_{n=1}^N s^2(n) - \left(\frac{1}{N} \sum_{n=1}^N s(n) \right)^2$$

$$E_e = \frac{1}{N} \sum_{n=1}^N e^2(n) - \left(\frac{1}{N} \sum_{n=1}^N e(n) \right)^2 \quad (4.1)$$

$$SNR(dB) = 10 \log_{10} \left(\frac{E_s}{E_e} \right)$$

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

그림 4-1은 G.723.1 부호화기와 제안한 부호화기의 피치예측 값과 피치이득 값을 비교한 것이다. 제안한 부호화기의 피치예측 값은 G.723.1의 그것에 비해 크게 변하지 않았지만 피치이득 값이 전이구간에서 보다 1에 가까워졌음을 알 수 있다. 표 1은 각 발성에 대해 SNRSEG를 나타낸 것이다.

표 1. SNRSEG 비교

	G.723.1	제안한 알고리즘	감소율(dB)
발성 1	2.4359	1.1022	1.3337
발성 2	1.9540	1.5565	0.3975
발성 3	1.8432	1.6424	0.2008
발성 4	2.1167	1.6916	0.4251
Total		0.5893(dB)	

표 2. MOS 결과

	발성1	발성2	발성3	발성4
G.723.1	3.56	3.63	3.53	3.72
제안한 알고리즘	3.65	3.68	3.7	3.80

5. 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 피치예측 값을 보상함으로써 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 열화가 발생하게 된다. 본 논문에서는 개회로에서 구해진 피치예측 값을 이용하여 피치이득 값을 계산하고 1에 근사한 값이 아닐 경우 이 값이 1에 가까운 값을 가지도록 피치예측 값을 조절한다. 이러한 방법을 사용하여 피치예측 값을 조절하여 부호화한 결과 segmental SNR이 평균 0.6dB 정도 개선되었고, MOS는 평균 0.12이 높아졌다.

참고 문헌

- [1] ITU-T Recommendation G.723.1, March, 1996.
- [2] A.M. Kondoz, "Digital Speech", John Wiley & Sons, 1994.
- [3] W. B. Klejin et. al, "Speech Coding and Synthesis", Elsevier Science B.V., 1995.
- [4] L.R.Rabiner, R.W.Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.
- [5] 나덕수, 노원석, 함명규, 배명진, "피치동기에 의한 음성신호의 전이구간 검출," 한국음향학회, 제 15회 음성 통신 및 신호처리 워크샵 논문집, Vol.15, No.1, PP.454-459, 1998년 08월 21-22일.
- [6] 나덕수, 정찬중, 박영호, 배명진, "LSP를 이용한 음성신호 성분분리에 의한 CELP 보코더의 전송률 감소에 관한 연구", 한국음향학회, 학술발표대회논문집, 1999, 8월.
- [7] "음성 신호 처리 기술", 한국과학기술원, 삼성첨단기술센터, 제 3권 음성부호화, PP.171

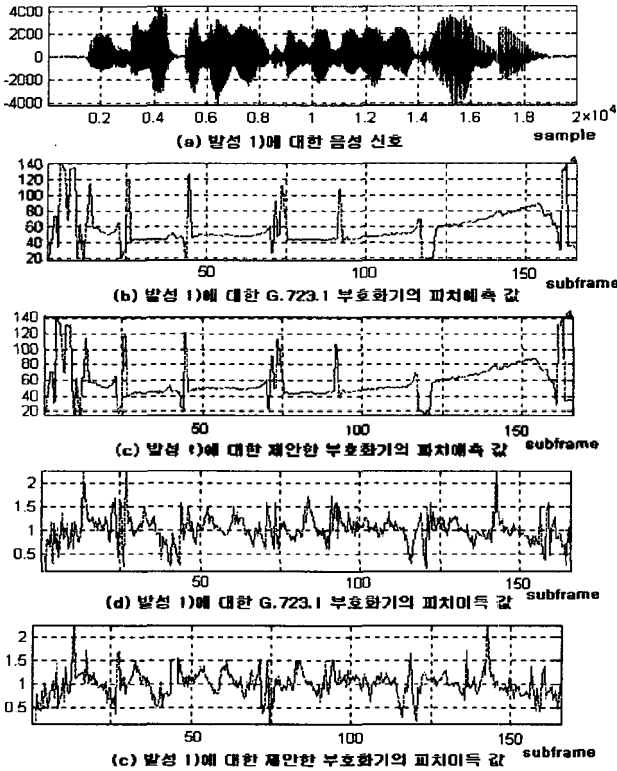


그림 4-1. G.723.1 부호화기와 제안한 부호화기의 피치예측 값과 피치이득 값의 비교

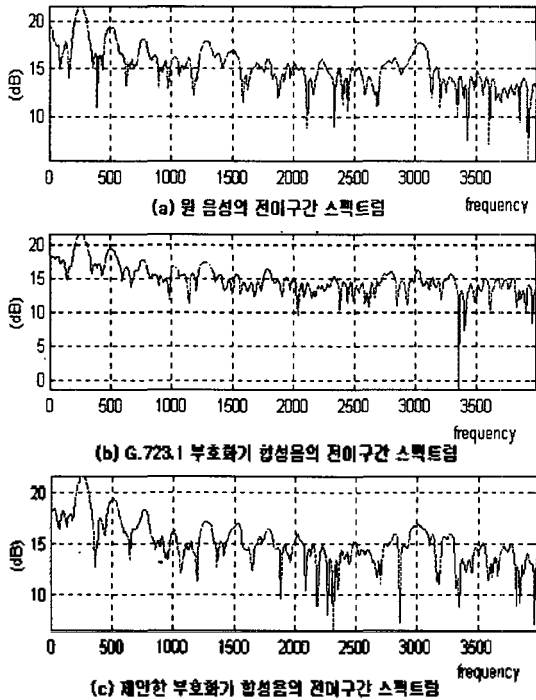


그림 4-2. 전이구간의 스펙트럼 비교