

# LSP를 이용한 인지가중필터의 성능개선에 관한 연구

정 현 옥, 김 의 성, 배 명 진

승실대학교 정보통신공학과

전화 : (02) 824-0906 / 팩스 : (02) 820-0018

## A Study on the Improvement of PWF Performance Using the LSP.

HyunUk JUNG, IkSung KIM, MyungJin BAE

Dept. Information and Telecommunication Engr., Soongsil University

E-mail : t2studio@nate.com

### Abstract

최근 음성 부호화기의 연구방향은 저전송률, 저복잡도와 더불어 가변전송률 음성부호화기에 대한 연구로 진행되고 있다. 지금까지 제안된 저전송률 음성부호화기로는 스펙트럼 모델링을 이용한 MBE 계열과 혼성부호화 방식의 CELP 계열이 있다. 그 중에서 가장 많은 연구가 이루어지고 있는 방식이 CELP 방식이다. 이 방식은 4.8Kbps 내외의 전송율에서 양호한 음질을 얻을 수 있다.

본 논문에서는 평균자승오차값을 최소화하여 계산량을 줄이고 음질을 향상시킬 수 있는 새로운 알고리즘을 제안한다. 먼저 G.723.1 부호화기에서 인지가중필터를 거친 신호를 LSP를 이용하여 각 포먼트의 위치를 검출하여 Pole점만 비교하여 Zero점의 영향을 최소화 하였고 평균자승오차값을 최소화 하여 문턱값에 가장 가까운 값을 대표 피치이득계수로 정하고 그때의 피치와 함께 부호화한다.

### 1. 서론

음성을 저장하는 방법에는 혼성부호화법이 있으며 여 기 신호를 저장하는 방법에 따라 여러 가지 방법이 제안 되어있다. 이중에 CELP법에서는 ABS-LPC구조가 사용 되고 있다[1][2][3]. 이러한 CELP 보코더에서는 계산량 이 방대한 단점은 있지만[4][5], 저전송률 부호화기의 구 현을 가능케 하며, 또한 인지가중필터를 사용함으로써 적은 비트 할당에도 주관적으로 느끼는 음질을 향상시 킬 수 있다. 코딩이나 디코딩시에 발생하는 예측 오차는

인지가중필터를 사용함으로써 최적화 할 수 있으며, 입력 신호의 주파수 응답을 변화시키기 위해서 인지적으로 덜 중요한 에러가 있는 스펙트럼 포락을 상쇄 할 수 있는 선형 필터는 매우 유용하게 사용된다.

### 2. G.723.1 부호화기의 원리

#### 2-1. 개요

부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있고 인지가중화된 오차 신호의 최소화를 시도한다. 부호화기는 한 프레임 240샘플마다 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 인지가중화된 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 예측값과 임펄스 응답을 이용하여 폐회로 피치 예측(closed loop pitch predictor)값이 계산되어진다.

피치 예측계수는 5차로 사용된다. 부호화기의 블록도는 그림 2-1에서 나타내었다.

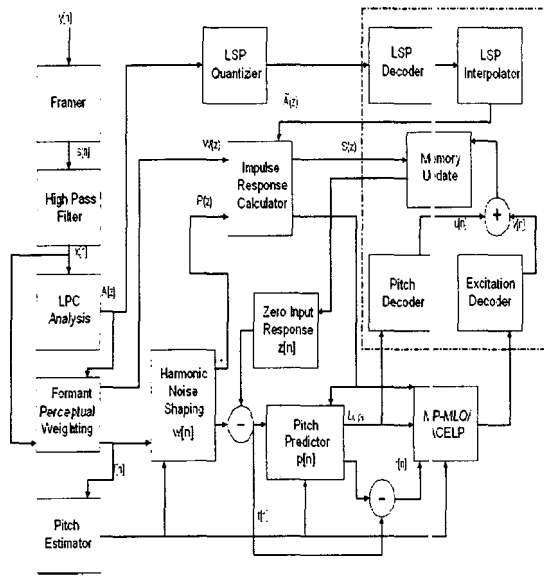


그림 2-1. G.723.1 음성 부호화기 블록도

## 2-2 인지가중필터

음성 부호화기에 많이 적용되는 방법은 합성에 의한 분석법이며, 부호화과정은 입력 음성과 합성된 음성간의 가장된 에러를 최소화하는 파라미터를 선택하는 과정이다. 에러최소화과정은 간략성 등의 이유로 코드북내에서 최적의 인덱스를 찾기 위한 방법으로 평균자승제곱오차가 일반적으로 사용되고 있다. 그리고 저장과 전송시에 낮은 전송율을 유지하기 위해 적은 비트가 사용되고 있기 때문에 과정을 완벽하게 비교하기가 무척 어렵다.

이런 상황에서도 인간의 인지도에 좀더 효과적으로 사용되고 있는 것이 인지 가중화된 필터이다. 지금까지 많은 연구가 진행되고 있지만 잔여 신호중 기준으로 선택할 만한 최적의 방법이 아직 까지 제시되고 있지 않다. 현재 ABS-LPC구조에서 인지가중필터(Perceptual Weighting Filter)가 가장 널리 쓰이고 있다. 인지가중필터에 관한 식은 아래와 같다[6].

$$W(z) = \frac{A(z)}{A(z/\gamma)} \quad (2.1)$$

$$= \frac{1 - \sum_{i=1}^p a_i z^{-i}}{1 - \sum_{i=1}^p a_i \gamma^i z^{-i}} \quad 0 \leq \gamma \leq 1$$

여기서  $A(z)$ 는 포먼트 예측오차필터이고, 인자값( $\gamma$ )

는 인지가중화된 파라미터이다. 인자값( $\gamma$ )의 효과는 중심포먼트 주파수를 변화 시키지 않는다. 그러나  $\nabla f$ 에 의한 포먼트의 대역폭은 식 (2.2)와 같이 넓어지게 된다.

$$\nabla f = \frac{-f_s}{\beta} \ln \gamma \quad (Hz) \quad (2.2)$$

여기서  $f_s$ 는 샘플링 주파수이다. 인지가중필터의 전형적인 주파수응답이 그림 2-2에 나타나있다.

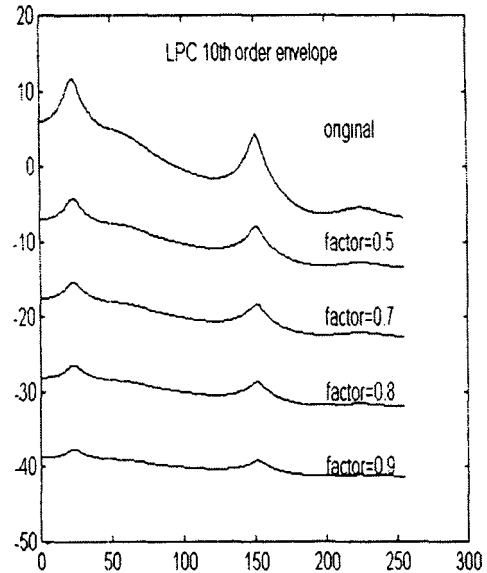


그림 2-2 Factor 값에 따른 LPC 포락선

그림 2-1에서 보듯이 가중필터는 LPC분석에 의해 결정된 포먼트의 진폭 값을 전체적으로 일정하게 만드는 역할을 한다. 포먼트의 진폭값을 작게 하기 위해서 0부터 1사이의 인자값( $\gamma$ )을 사용한다. 청취도 테스트결과 가장 적당한 인자값( $\gamma$ )은 0.8에서 0.9정도이다[7]. 음성 부호화기에 적용된 합성필터는 식 2-3과 같다.

$$H(z) = \left(\frac{1}{A(z)}\right) * W(z) = \frac{1}{A(z/\gamma)} \quad (2.3)$$

비록 인지가중필터가 에러최소화 전단계에 위치하여 사용될 수 있지만 계산의 편리를 위하여 가중필터를 두 가지(branch)로 나누어서 위치시킬 수 있다.

$$\frac{1}{A_w(z)} = \frac{1}{A(z)} \cdot W(z) \quad (2.4)$$

$$= \frac{1}{1 - \sum_{i=1}^p a_i \gamma^i z^{-i}}$$

LPC계수와 인자값( $\gamma$ )을 이용하여 LPC포락선을 구한 후 에러 최소화과정을 수행하는 것이 보통이다.

### 2-3. LSP 파라미터의 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC 분석이 이루어 진다[1].

$$H(z) = 1/A_p(z) \quad (2.5)$$

$$\text{where } A_p(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.6)$$

$H(z)$ 는 LPC 필터이고  $p$ 는 필터의 차수이다.

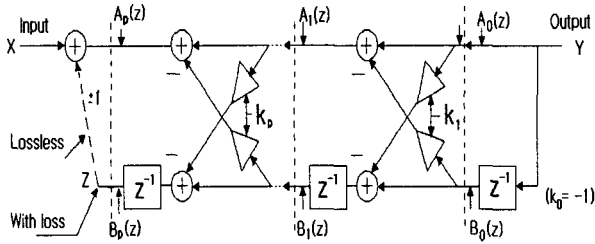


그림 2-3. PARCOR structure of LPC synthesis

LSP 파라미터를 유도하기 위해서 PARCOR (Partial Correlation) 필터를 이용하여 식(2.5)과 식(2.6)를 표현하면 다음과 같다.

$$A_{p-1}(z) = A_p(z) + k_p B_{p-1}(z) \quad (2.7)$$

$$B_p(z) = z^{-1} [B_{p-1}(z) - k_p A_{p-1}(z)]$$

여기서  $A_0(z) = 1$ ,  $B_0(z) = z^{-1}$  이고

$$B_p(z) = z^{-(p-1)} A_p(z^{-1}) \quad (2.8)$$

그림 2-3의 PARCOR 구조는 손실이 없는 음파관에서 음파의 전달로 이해된다. 시스템은 단지 역방향(backward) 에너지 모양에서  $Z$  종점에서 손실이 있다. 이러한 음관은  $Z$  종점의 출력이  $k_{p-1} = \pm 1$ 의 경로를 통해 입력의 종점으로 귀환될 때 완전한 무손실이 된다. 각각의 공명 값인  $Q$ 는 무한해지고 에너지 분포 스펙트럼은 몇 개의 선 스펙트럼에 집중된다[2][6].  $k_{p-1} = -1$  조건의 귀환은 입력종점에서 완전히 폐쇄되고  $k_{p-1} = +1$  은 무한 자유공간상으로 개방된다[1][6].

그림 2-3에서  $k_{p-1} = \pm 1$  인 전달함수를  $P_{p-1}(z)$ 와  $Q_{p-1}(z)$ 로 나타내면 다음과 같다.

$$k_{p-1} = 1 \text{ 일때, } P_{p-1}(z) = A_p(z) - B_p(z) \quad (2.9)$$

$$k_{p-1} = -1 \text{ 일때, } Q_{p-1}(z) = A_p(z) + B_p(z)$$

$$\Rightarrow A_p(z) = \frac{1}{2} [P_{p-1}(z) + Q_{p-1}(z)]$$

두 개의 근( $k_{p-1} = \pm 1$ )을 알고 있으므로  $P_{p-1}(z)$ 의  $Q_{p-1}(z)$ 의 차수를 줄일 수 있다. 즉,

$$P'(z) = \frac{P_{p-1}(z)}{(1-z)} = A_0 z^p + A_1 z^{(p-1)} + \dots + A_p \quad (2.10)$$

$$Q'(z) = \frac{Q_{p-1}(z)}{(1-z)} = B_0 z^p + B_1 z^{(p-1)} + \dots + B_p$$

$$\text{조건 : } A_0 = 1, B_0 = 1$$

$$A_k = (a_k - a_{p-1-k}) + A_{k-1} \quad (2.11)$$

$$B_k = (a_k - a_{p-1-k}) - A_{k-1} \quad \text{for } k = 1, \dots, p$$

LSP는  $0 \leq \omega_i \leq \pi$  인 범위에서  $P'(z)$ 와  $Q'(z)$ 을 통해 얻어진 근의 각(angular) 위치를 나타낸다. LSP는 다음과 같은 두가지 성질을 지닌다.

첫째,  $P'(z)$ 와  $Q'(z)$ 는 단위원 상에 놓여 있다.

둘째,  $P'(z)$ 와  $Q'(z)$ 의 근들이 단위원 상에 번갈아 나타난다.

### 3. 제안한 알고리즘

기존의 방법은 시간영역에서 LPC 계수를 얻은 후 주파수 영역으로 변환하여 포먼트 포락을 얻게된다. 이때 제 1포먼트의 에너지가 다른 포먼트의 에너지값에 비해 상당히 크기 때문에 에러 최소화 과정에서 전체 프레임의 에러값을 좌우하게 되는 문제점이 있다. 이 문제점을 해결하기 위하여 인지가중필터 방식을 사용한다. 하지만 정확한 포먼트값을 얻을 수 없고, 일반적으로 사용하고 있는 LPC 전극 모델링은 영점모델링을 하지 않기 때문에 Zero점에 의한 영향을 최소화 해야한다.

본 논문에서 제안한 방법은 시간영역에서 LPC 계수를 얻은 후 원 신호와 합성신호의 차를 인지가중필터에 통과시켜서 LSP를 이용하여 각 포먼트의 위치를 검출하여 Zero점의 영향을 최소화하기 위해 Pole점만 비교하여 계산량의 감소와 주간적으로 느끼는 음질을 향상시키는 새로운 방법을 제안하였다.

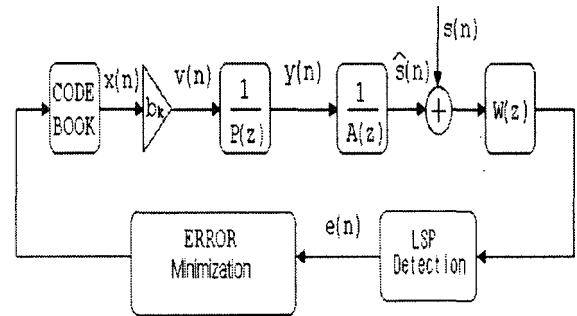


그림 3-1 제안한 방법의 블록도

## 4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC Pentium III(1.1GHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 처리결과와 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 음성 시료는 SNR이 30dB인 환경 하에서 녹음하였다. 음성 시료는 다음과 같다.

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

실험은 MatLab과 C-언어를 사용하여 prototype으로 구현하였다. 제안한 부호화기의 전송률은 8kbps 샘플링되고 16비트 양자화된 PCM신호를 평균 96.8% 압축할 수 있었고, 남성과 여성에 따라 또는 목음과 무성음의 길이에 따라 유동적이었다. 제안한 부호화기는 4kbps 초반의 전송율을 나타내었다. 제안한 부호화기의 음질을 평가 위해 5.3kbps의 전송율을 가지는 G.723.1 ACELP 부호화와 MOS(Mean Opinion Score)를 비교하였다. 각 부호화기의 실험 발성에 대한 MOS는 20대 남녀 각 5명이 각 발성에 대해 10번씩 무작위 청취하여 평가한 점수를 평균한 것이다.

표 4-1.은 제안한 알고리즘과 G.723.1 부호화를 사용하여 계산시간을 비교한 것이다. 제안한 알고리즘의 LSP 변환 방법이 기존의 방법 보다 평균 48%이상 빠르다는 것을 보여주고 있다. 그러나, 변환된 LSP 계수를 비교해 보면 기존의 것과 동일하였다. 즉, LSP 계수의 값은 변하지 않고 계산시간만 48% 감소하였다. 그리고 표 4-2 역시 제안한 알고리즘과 G.723.1 부호화를 사용하여 MOS를 측정할 결과 Pole점만 비교하여 평균자승오차값을 최소화하는 방법이 음질이 향상된 것을 알 수 있다.

표 4-1. 기존의 방법과 제안한 방법의 계산시간 비교

	기존의 방법 (단위:초)	제안한 방법 (단위:초)	감소율 (단위:%)
발성(1)	120.15	62.39	48.07
발성(2)	74.04	38.48	48.03
발성(3)	73.22	37.68	48.54
발성(4)	119.90	62.78	47.63

표 4-2. 기존의 방법과 제안한 방법의 MOS 비교

	G.723.1 (5.3kbps)	제안한 부호화기
발성(1)	3.84	3.9
발성(2)	3.76	3.9
발성(3)	3.8	3.8
발성(4)	3.7	3.8

## 5. 결론

최근에 가장 많이 사용되고 있는 음성부호화기중에 하나인 CELP부호화기에서는 적은 비트가 할당되기 때문에 이러한 적은 비트 할당에도 주관적으로 느끼는 음질을 향상시키기 위해서 인지가중화된 필터를 사용한다. 그러나 기존방법은 Pole점과 인지적으로 덜 중요한 Zero점을 같이 비교하는 방식이다.

본 논문에서는 인지가중필터를 거친 신호를 LSP를 이용하여 각 포먼트의 위치를 검출하여서 Zero점을 제외하고 Pole점만 비교하여서 최소의 에러값을 찾아내는 데 있어서 제 1 포먼트의 영향을 덜 받고 제 2포먼트 또는 제 3 포먼트의 영향으로 향상된 음질을 갖는 최소의 에러값을 찾아낼 수 있게 된다. 또한 Pole점만 비교하기 때문에 기존의 Pole점과 Zero점을 비교하는 평균자승오차값을 최소화하여 계산량을 줄일 수 있고 음질 역시 향상된 것을 들 수 있었다.

## 참고 문헌

- [1] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signal, Prentice-Hall, 1978.
- [2] A.N. Ince, Digital Speech Processing(speech coding, synthesis, and recognition), Kluwer Academic Publishers, 1992.
- [3] G. Fant, Acoustic Theory of Speech Production, The Hague, The Netherlands, Mouton, 1970
- [4] M.R. Schroeder and B.S. Atal, Code-Excited Linear Prediction(CELP) : High-Quality at Low Bit Rates, Proc. Int. Conf. on Acoust, Speech and Signal Processing, pp.25.1.1 - 25.1.4, 1985
- [5] J. RHEEM, M. BAE, S. ANN, Reduction of Pitch Searching Rate in CELP Vocoder, J.,Acoust., Soc., Korea, Vol.12, No.2E, pp.33-37, July 1993
- [6] A.M.Kondoz, Digital Speech, John Wiley & Sons, 1994.
- [7] M.Schroeder and B.Atal. "Predictive coding of speech signals and subjective error criteria". IEEE Trans. on ASSP, 1979, 247-254
- [8] APCO Project 25 Vocoder Description, Digital Voice System, Inc., Ver.13,15,Jul,1993.