

# 유무성음 판정에 따른 LSF 코사인테이블 재구성에 의한 음질향상

최성영, 배명진

승실대학교 정보통신공학과

## Speech Enhancement by Reconstruction of Cosine Table for LSF Roots According to the Voiced/Unvoiced Decision

SeongYoung Choi, MyungJin BAE,

Dept. of Telecomm. Engr., Soongsil Univ.

mjbae@saint.soongsil.ac.kr

### Abstract

CELP 계열 보코더중 인터넷 폰 및 화상회의를 목적으로 개발된 G.723.1 보코더에서는 LPC를 LSP로 변환하기 위해 LPC 계수를 이용한 다항식을 구성한 다음 근을 검색하는 방법을 사용하고 있다. 근 검색시에는  $256/\pi$ 의 범위안에서 동일한 간격을 갖는 코사인 테이블을 구성하여 순차적으로 검색하게 된다.

LSF의 근들은 포먼트가 존재하는 대역에서 근들이 나타나게 되므로 유성음의 경우 저주파수 대역에서 무성음의 경우 고주파수 대역에서 많이 분포하게 된다. 하지만 G.723.1에서 사용하는 코사인 테이블은 음성신호의 특성을 고려하지 않고 균등한 간격을 갖는 값들을 사용함으로써 음질을 저해할 수 있는 요소를 갖고 있다.

따라서 본 논문에서는 음성의 특성을 고려한 코사인 테이블을 재구성함으로써 음질을 향상시킬 수 있었으며 주관적 음질평가인 MOS 시험결과 평균 1.8 정도의 음질향상을 가져올 수 있었다.

### I. 서 론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP(Code Excited Linear Prediction)구조이다. 이 방식은 4.8kbps내외의 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA등 여

러 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. 특히 PCS 및 전화기 라인상에서의 인터넷을 통한 화상회의를 위하여 낮은 전송율에서 고음질을 가지는 코덱이 많은 주목을 받고 있다[1].

이중 유선망을 이용한 화상회의, 인터넷폰을 목적으로 ITU-T에서 표준화된 G.723.1에서는 LPC 계수를 전송하기 위해 양자화 에러에 강하고 시스템의 안정성이 보장된다는 특성을 가지고 있는 LSF로 변환하는 과정을 거치게 된다. LPC를 LSF로 변환하기 위해서는 먼저 LPC 계수를 이용하여 다항식을 구성하고 이 다항식의 근을 구하여 LSF로 변환하게 된다. G.723.1에서는 다항식의 근들을 코사인 테이블로 구성하여 순차적으로 검색하는 방법을 사용하고 있다. 이때 사용되는 코사인 테이블의 값은  $256/\pi$  사이에서 균등한 간격으로 구성하게 된다.

일반적으로 음성신호가 유성음인 경우 포먼트는 저주파수 대역이 많이 분포하며 무성음의 경우에는 고주파수 대역이 많이 분포되어 있다. 따라서 LSP 근들은 유성음의 경우 저주파수 대역에서 그리고 무성음의 경우 고주파수 대역에서 많이 나타나게 되는 것을 알 수 있다. 따라서 본 논문에서는 이러한 특성을 이용하여 음성신호의 구간별 특징을 구한 다음 각각의 특징에 맞는 코사인 테이블을 재구성하여 음질을 향상시키는 방법에 대해 제안한다.

## II. LSF 파라미터 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC(Linear Predictive Coding)분석이 이루어져야 한다[1].

$$H(z) = 1/A_p(z) \quad (2.1)$$

$$\text{where } A_p(z) = 1 + \sum_{k=1}^p \alpha_k z^{-k} \quad (2.2)$$

여기서  $H(z)$ 는 LPC 필터이고  $p$ 는 필터의 차수이다. LSP 파라미터를 유도하기 위해서 PARCOR(Partial Correlation) 필터를 이용해서 식(2.1)과 식(2.2)를 표현하면 다음과 같다.

$$A_{p-1}(z) = A_p(z) + k_p B_{p-1}(z) \quad (2.3)$$

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

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

$$B_p(z) = z^{-(p+1)} A_p(z^{-1}) \quad (2.4)$$

그림 2-1.에 보이는 것처럼 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 [3][4].

$k_{p+1} = -1$  조건의 귀환은 입력종점에서 완전히 폐쇄되고  $k_{p+1} = +1$  은 무한 자유공간상으로 개방된다 [1][2].

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

$$k_{p+1} = 1 \text{ 일때, } P_{p+1}(z) = A_p(z) - B_p(z) \quad (2.5)$$

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

$$\Rightarrow A_p(z) = \frac{1}{2} [P_{p+1}(z) + Q_{p+1}(z)] \quad (2.6)$$

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,

$$P(z) = \frac{P_{p+1}(z)}{(1-z)} = A_0 z^p + A_1 z^{(p-1)} + \dots + A_p \quad (2.7)$$

그리고

$$Q(z) = \frac{Q_{p+1}(z)}{(1-z)} = B_0 z^p + B_1 z^{(p-1)} + \dots + B_p \quad (2.8)$$

조건 :

$$A_0 = 1, B_0 = 1 \quad (2.9)$$

$$A_k = (\alpha_k - \alpha_{p+1-k}) + A_{k-1}$$

$$B_k = (\alpha_k - \alpha_{p+1-k}) - A_{k-1} \quad (2.10)$$

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 가지고 있다.

첫째,  $P(z)$ 와  $Q(z)$ 는 단위원 상에 놓여 있다.

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

### Real root method

일반적으로  $P(z)$ 와  $Q(z)$ 의 다차 방정식의 해를 구하는 방식에 따라 여러 가지 변환법이 개발되었다. 그러나 이러한 변환법 중 real root 방법이 비교적 간단하고 이해하기 쉬워 주로 사용되어지고 있다[2].

$P(z)$ 와  $Q(z)$ 의 계수는 대칭적이기 때문에 식(2.7)의 차수는  $p/2$ 로 줄어든다.

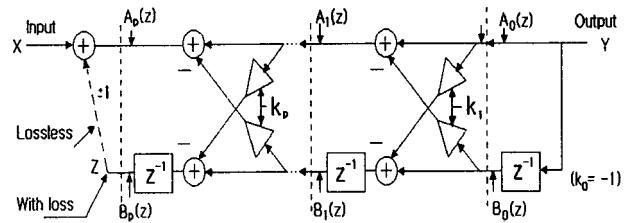


그림 2-1. PARCOR structure of LPC synthesis

$$\begin{aligned}
P(z) &= A_0 z^p + A_1 z^{p-1} + \dots + A_1 z^1 + A_0 \\
&= z^{p/2} [A_0 (z^{p/2} + z^{-p/2}) + \\
&\quad A_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + A_{p/2}]
\end{aligned} \tag{2.11}$$

$$\begin{aligned}
Q(z) &= B_0 z^p + B_1 z^{p-1} + \dots + B_1 z^1 + B_0 \\
&= z^{p/2} [B_0 (z^{p/2} + z^{-p/2}) + \\
&\quad B_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + B_{p/2}]
\end{aligned} \tag{2.12}$$

모든 근이 단위원 상에 있기 때문에, 단지 아래와 같이 정의하고 단위원 상에서 식(2.11)의 값을 구할 수 있다.

$$\text{Let } z = e^{j\omega} \text{ then } z^1 + z^{-1} = 2 \cos(\omega) \tag{2.13}$$

$$\begin{aligned}
P(z) &= 2e^{j\omega p/2} [A_0 \cos(\frac{p}{2}\omega) + \\
&\quad A_1 \cos(\frac{p-2}{2}\omega) + \dots + \frac{1}{2} A_{p/2}]
\end{aligned} \tag{2.14}$$

$$\begin{aligned}
Q(z) &= 2e^{j\omega p/2} [B_0 \cos(\frac{p}{2}\omega) + \\
&\quad B_1 \cos(\frac{p-2}{2}\omega) + \dots + \frac{1}{2} B_{p/2}]
\end{aligned} \tag{2.15}$$

$x = \cos \omega$ 를 대입해서 식(2.14)와 식(2.15)을  $x$ 에 대해서 풀 수 있다. 예를 들어서  $p = 10$ 이면 다음과 같이 얻어진다.

$$\begin{aligned}
P_{10}(x) &= 16A_0 x^5 + 8A_1 x^4 + (4A_2 - 20A_0)x^3 \\
&\quad + (2A_3 - 8A_1)x^2 + (5A_0 - 3A_2 + A_4)x \\
&\quad + (A_1 - A_3 + 0.5A_5)
\end{aligned} \tag{2.16}$$

유사하게,

$$\begin{aligned}
Q_{10}(x) &= 16B_0 x^5 + 8B_1 x^4 + (4B_2 - 20B_0)x^3 \\
&\quad + (2B_3 - 8B_1)x^2 + (5B_0 - 3B_2 + B_4)x \\
&\quad + (B_1 - B_3 + 0.5B_5)
\end{aligned} \tag{2.17}$$

LSP는 식(2.18)에 의해서 구해진다.

$$LSP(i) = \frac{\cos^{-1}(x_i)}{2\pi T}, \text{ for } 1 \leq i \leq p \tag{2.18}$$

이 방법은 다른 변환 방법보다 비교적 간단하나 계산시간이 어느 정도 걸릴지 예상할 수 없다는 단점이 있다[2].

### III. 코사인 테이블의 재구성

음성신호를 스펙트럼 분석하면 그림 3-1에서와 같이 유성음의 경우 저주파수 대역에 높은 에너지가 존재하며 무성음의 경우에는 고주파수 대역에 높은 에너지가 존재하게 된다. 따라서 LSP는 유성음의 경우에는 저주파수 대역에 많이 분포하게 되며 무성음의 경우에는 고주파수 대역에 많이 분포하게 된다.

G.723.1 보코더에서는  $256/\pi$ 의 범위 안에서 균등한 간격을 갖는 근들을 코사인 테이블 구성하여 다항식의 근을 순차적으로 검색하는 방법을 사용하고 있다. 본 논문에서는 LPC 계수를 LSF로 변환하기 전에 유무성음 판정을 먼저 수행하게 된다. 그런 다음 유성음의 경우  $P'(z)$ 와  $Q'(z)$ 를 구하기 위한 512개의 코사인테이블 엔트리 중  $0 \sim \pi/2$  대역까지는 384개의 근이 균등하게 존재하게 하고  $\pi/2 \sim \pi$  사이에는 128개의 근이 균등하게 존재하는 코사인 테이블을 재구성하였다. 무성음의 경우에는 이와 반대로  $0 \sim \pi/2$  대역까지는 128개의 근이 균등하게 존재하게 하고  $\pi/2 \sim \pi$  사이에는 384개의 근이 균등하게 존재하는 코사인 테이블을 재구성하였다.

즉 유성음의 경우에는 저주파수 대역의 검색구간을 세밀하게 하고 무성음의 경우에는 고주파수 대역의 검

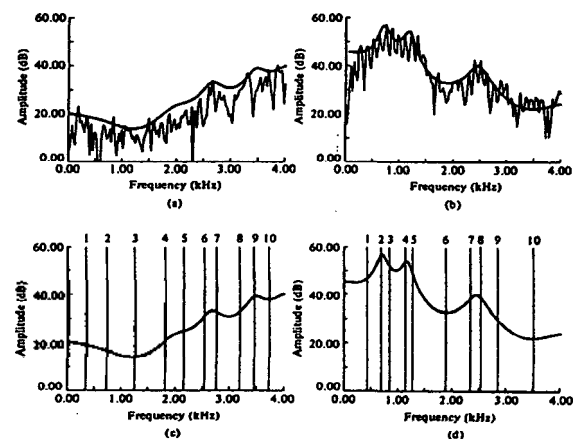


그림 2-2. 선형스펙트럼상의 예  
(a) 자음 /s/ (b) 모음 /a/  
(c),(d) /s/와 /a/에 대한 LPC 분석과 LSP

색구간을 세밀하게 함으로써 보다 정확한 LSF 근을 구하도록 하였다.

유무성음 판정은 에너지와 ZCR을 이용하여 판정하였으며 판정의 모호함에 따른 음질 저하를 방지하기 위해 이런 경우에는 기존의 G.723.1 코사인 테이블 이용하였다.

#### IV. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 586(233MHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. G.723.1에서는 8kHz로 음성을 표본화한 음성을 입력으로 하며 각 시료에 대해 한 프레임의 길이를 240표본으로 하여 처리하였다. 처리결과의 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 음성 시료는 SNR이 20dB인 환경하에서 녹음하였다. 음성 시료는 다음과 같다.

- 발성1: /인수네 꼬마는 천재소년을 좋아한다./
- 발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./
- 발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다./
- 발성4: /승실대학교 정보통신과 음성통신 연구팀이다./
- 발성5: /공일이삼사오육칠팔구/

제안한 알고리즘의 시뮬레이션은 C-언어로 구현하여 수행하였다. 성능 비교는 G.723.1을 통과한 음성과 제안한 알고리즘을 통과한 음성에 대해 MOS test를 수행하여 측정하였다. 실험 결과 G.723.1에 비해 메모리가 4K Word 증가되었다. 표1은 주관적 음질평가 결과를 나타내고 있으며 실험 결과 평균 약 1.8 정도의 음질 향상을 가져올 수 있었다.

표 1. MOS test 결과

	시료 1	시료 2	시료 3	시료 4	시료 5	평균
G.723.1	3.70	3.78	3.77	3.73	3.74	3.74
제안한 알고리즘	3.98	3.94	3.93	3.89	3.87	3.92

#### V. 결론

화상회의, 인터넷폰을 목적으로 ITU-T에서 표준화된 G.723.1에서는 LPC 계수를 전송하기 위해 양자화 에러에 강하고 시스템의 안정성이 보장된다는 특성을 가지

고 있는 LSF로 변환하는 과정을 거치게 된다. LPC를 LSF로 변환하기 위해서는 먼저 LPC 계수를 이용하여 다항식을 구성하고 근을 구하여 LSF로 변환하게 된다. G.723.1에서는 다항식의 근들을 코사인 테이블로 구성하여 순차적으로 검색하는 방법을 사용하고 있다.

일반적으로 음성신호가 유성음인 경우 포먼트는 저주파수 대역이 많이 분포하며 무성음의 경우에는 고주파수 대역이 많이 분포되어 있다. 따라서 LSP 근들은 유성음의 경우 저주파수 대역에서 그리고 무성음의 경우 고주파수 대역에서 많이 나타나게 되는 것을 알 수 있다. 따라서 본 논문에서는 이러한 특성을 이용하여 음성신호의 구간별 특징을 구한 다음 각각의 특징에 맞는 코사인 테이블을 재구성하여 음질을 향상시키는 방법에 대해 제안한다.

유무성음 판정은 에너지와 ZCR 파라미터를 사용하였으며 판정이 모호한 경우에는 기존 G.723.1 테이블을 이용하였다. 실험결과 평균 약 1.8정도의 MOS Score를 향상시킬 수 있었다.

#### 참고 문헌

- [1] A.M. Kondoz, "Digital Speech", John Wiley & Sons, 1994.
- [2] 배명진, "디지털 음성분석", pp.95-120, 동영출판사, 1998. 4
- [3] John R. Deller, Jr., John G. Proakis, John H.L. Hansen, "Discrete-Time Processing of Speech Signals", pp.124-125, Maxwell Macmillan International, 1993.
- [4] Sadaoki Furui, "Digital Speech Processing, Synthesis, and Recognition", pp129, MARCEL DEKKER, INC. 1991.
- [5] ITU-T Recommendation G.723.1, March, 1996.
- [6] 나덕수, 정찬중, 박영호, 배명진, "LSP를 이용한 음성신호의 성분분리에 의한 CELP 보코더의 전송률 감소에 관한 연구", 한국음향학회, 학술발표대회논문집, 1999, 8월
- [7] N. S. Jayant and P. Noll, Digital Coding of Waveform-Principles and Applications to Speech and Video, pp.220-221, Prentice-Hall, 1978.
- [8] W. B. Kleijn et. al, "Speech Coding and Synthesis", Elsevier Science B.V., 1995.