

# LSP를 이용한 음성신호의 성분분리에 의한 CELP 보코더의 전송률 감소에 관한 연구

나 덕수, 박영호, 정찬중, 배명진

승실대학교 정보통신공학과  
156-743 서울시 동작구 상도동 1-1  
dsna@assp.soongsil.ac.kr

## A Study on a Reduction of the Transmission Bit Rate by the U/V Decision Using LSP in the CELP Vocoder

DuckSu Na, MyungKyu Ham, YoungHo Park, MyungJin Bae

Dept. Information and Telecommunication Engr., Soongsil University  
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA  
dsna@assp.soongsil.ac.kr

### 요 약

기존의 CELP 보코더에서, 무성음에 대한 별도의 처리 없이 유성음과 동일하게 처리하였다. 유성음과 무성음은 발생모델측면에서 임펄스열과 랜덤 잡음으로 각각 다름에도 불구하고 동일하게 처리함으로써 합성음에서 음질의 저하 및 계산량과 전송률 측면에서 손실을 가져왔다. 또, U/V(Unvoiced /Voiced) 분류기를 사용하는 경우에는 U/V 분류기의 성능에 따라 합성음의 음질저하의 정도의 차이가 심하다.

본 논문에서는 에리울과 전처리 계산량을 최소화 할 수 있는 U/V 분류기를 사용하여 CELP 보코더에서 전송률을 감소시키는 방법을 제안한다. CELP 보코더에서는 스펙트럼 정보를 LPC 파라미터로 추출한 후 다시 전송형 파라미터인 LSP(Line Spectrum Frequency)로 변환한다. 새로운 U/V 분류기는 이 LSP 파라미터를 이용한다. LSP 파라미터의 주파수영역 분포도와 간격정보를 이용하여 U/V를 결정하게 된다. 제안한 방법을 5.3kbps ACELP에 적용하여 성능 평가를 실시하였다. 실험결과 음질의 저하 없이 5.6% (280bps)의 전송률을 감소할 수 있었다.

### 1. 서 론

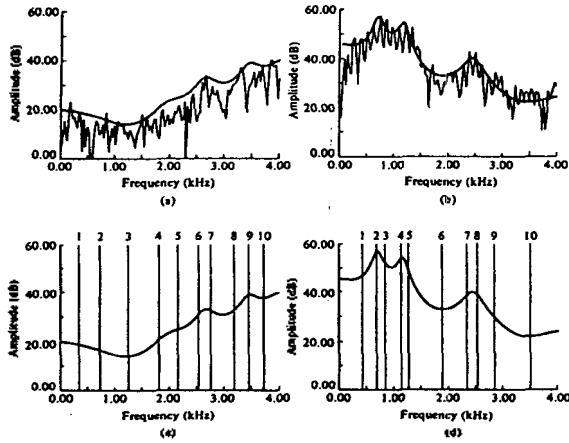
과거의 음성 정보를 전송하는 방법은 한정된 일부

사용자와 특정 분야에 적용되었기 때문에 전송률은 크게 고려되지 않았으며 고음질의 음성을 수신단에 보낼 수 있으면 되었다. 그러나 과거에 비해 멀티미디어와 이동 통신의 발달로 인해 특정 집단이나 개인에게 제공되었던 서비스들이 이제는 다수의 대중에게도 제공하게 되었고 그 숫자 역시 기하 급수적으로 늘어나게 되었다. 그로 인해 지금까지 적용되어왔던 전송률로는 사용자 집단의 숫자를 충족시킬 수 없게 되었다. 또한 전송률을 저하시켜 동일 채널 상에 사용자 수를 증가시킬 경우에 발생하는 음질 열하 역시 문제가 되었다. 이와 같은 배경하에서 음성 부호화기, 즉 보코더의 개발이 이루어지게 된 것이다.

CELP형 보코더는 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 하지만, 무성음에서도 유성음과 같은 분석을 통해서 데이터를 전송하므로 보다 많은 계산량이 필요하게 되고 전송률 측면에서도 손실을 가져오게 된다. 또 U/V 분류기를 사용하는 경우 이것의 성능에 따라 음질 저하의 차이가 심하였다.

본 논문에서는 기존의 CELP 보코더에서 사용되는 LSP 파라미터의 특성을 이용하여 유성음과 무성음을 분리하여 전송률을 낮추는 방법을 제안하고자 한다. 먼저 2장에서는 LSP 파라미터를 이용하여 성분분리하는 방법에 대하여 설명하고 3장에서는 제안한 성분분리해 대한 실험 및 결과 그리고 CELP 보코더에 적용한 결과를 설명하고 4장에서 결론을 맺는다.

## 2. LSP 파라미터를 이용한 성분 분리



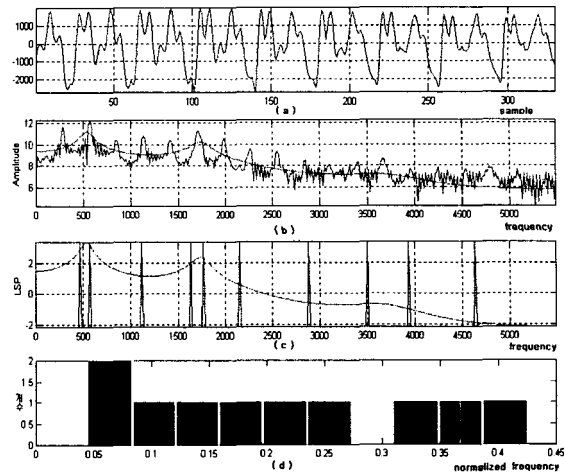
그림<2-1> LSP 파라미터의 예

- (a) 자음 /s/ (b) 모음 /a/  
(c),(d) /s/와 /a/에 대한 LPC 분석과 LSP

LSP 파라미터는 일정한 스펙트럼 민감도와 저전송률 부호화에서 낮은 스펙트럼 왜곡, 좋은 선형보간특성을 가진다. LSP 파라미터를 얻기 위해서는 먼저 선형예측 방법에 의해 LPC(Linear Prediction Coefficient)를 구하고 이를 변환하여야 한다. 변환방법은 복소근(Complex root) 방법, 실근(Real root) 방법, Ratio filter 방법, 등이 있으나 실근 방법이 주로 사용된다[3].

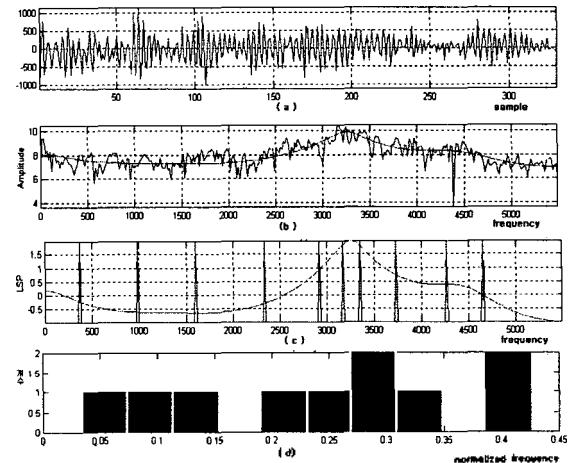
그림 <2-1>은 음성신호의 무성음과 유성음에 대한 스펙트럼과 LPC 분석 그리고 LSP 파라미터를 보여주고 있다. 유성음의 스펙트럼은 (b)와 같이 1kHz 이내에서 제1포먼트가 나타나고 3개 이상의 포먼트가 존재한다. 반면에 무성음의 스펙트럼은 (a)와 같이 저주파수 영역의 에너지가 낮고 고주파수 영역의 에너지가 높다. 이와 같은 스펙트럼모양으로 LSP의 모양도 유성음과 무성음에서 뚜렷한 차이를 가진다. 먼저 유성음에서는 포먼트로 인해 저주파수 영역에서 보다 많은 LSP가 분포하고 그 간격도 고주파수이 LSP 보다 좁게 나타난다. 반면에 무성음에서는 고주파수 영역에서 LSP가 많이 분포하고 그 간격도 저주파수영역의 LSP보다 좁게 나타난다.

위와 같은 유/무성음에 따른 LSP 파라미터의 특성으로 인해 성분분리가 가능해진다. 먼저 LSP의 유/무성음에 따른 분포의 차이를 이용한다. 샘플링 주파수를  $F_s$ 라고 할 때  $F_s/4$  이하의 주파수영역에 존재하는 LSP 개수를 NL이라 하고,  $F_s/4$  이상의 주파수영역에 존재하는 LSP 개수를 NH라 한다면, NL이 NH보다 큰 경우



그림<2-2> 유성음에 대한 LSP 분포

- (a) 유성음 파형 (b) 음성파형의 스펙트럼  
(c) LPC 분석과 LSP (d) LSP 분포



그림<2-3> 무성음에 대한 처리예

- (a) 무성음 파형 (b) 음성파형의 스펙트럼  
(c) LPC 분석과 LSP (d) LSP 분포

는 음성신호의 스펙트럼이 저주파 쪽에서 봉우리(pole)가 많이 나타나는 모양이어서 유성음의 스펙트럼 특징을 나타낸다고 간주한다. 즉 유성음의 제 1 포먼트와 제 2 포먼트가 주로 저주파수 영역에 존재하기 때문이다.

이와 반대로 NH가 NL보다 큰 경우는 무성음을 나타낸다고 결정한다. 즉 무성음의 스펙트럼은 주된 포먼트가 고주파수영역에 나타나기 때문이다. 하지만 /i/, /I/, /ε/, /æ/ 와 같은 유성음은 제 2 포먼트, 제 3 포먼트 또는 제 4 포먼트가 고주파수 영역에 존재하여 유성음인면서도 NH가 NL보다 크게 나타난다. 이와 같은 경우에는 제 1 포먼트의 존재 여부로써 무성음인지 유성음인지를 결정하게 된다. 즉, LSP 파라미터들의 간격을 조사

하여  $F_s/4$  이하의 영역에서 좁은 간격을 나타내는 LSP 들이 존재하면 유성음으로 간주한다.

그림 <2-2>는 유성음에 대한 LSP 분포를 나타낸다. (d)의 Histogram에서 10개의 막대 중 앞에서 5개의 막대가 NL을 구성하는 것이고 나머지 5개의 막대가 NH를 구성하는 것이다. NL이 NH보다 큰 것을 알 수 있어 Histogram만으로도 유성음임을 알 수 있다. 반면에 그림 <2-3>의 (d)에서는 NH가 NL보다 크게 나타남을 알 수 있다. 즉 유성음에서는 LSP가 저주파수 영역에 많이 분포하고 무성음에서는 고주파수 영역에 많이 분포함을 알 수 있다.

### 3. 실험 및 결과

제안한 방법을 실험하기 위해서 먼저 IBM PC(233 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남자와 여자가 연구실 환경(30dB 의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit 로 양자화하여 사용하였다. 발성한 문장은 다음과 같다.

- 발성1) "인수녀 꼬마는 천재소년을 좋아한다."
- 발성2) "창공을 날으는 인간의 도전은 끝이 없다."
- 발성3) "예수님께서 천지창조의 고운을 말씀하셨다."
- 발성4) 일기예보 아나운서 음성시료

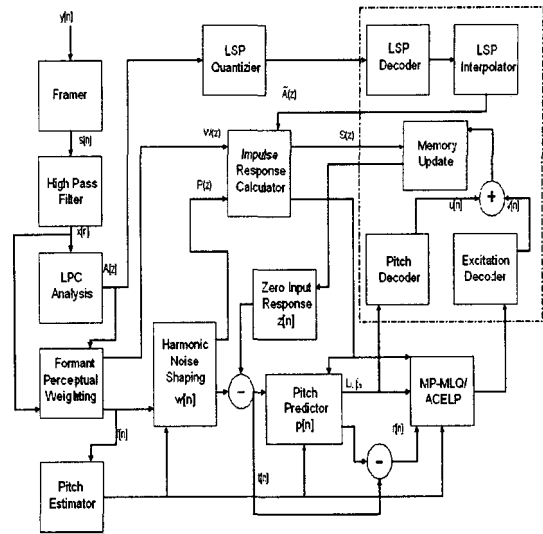
제안한 LSP 파라미터를 이용한 성분분리 방법을 C-언어로 구현하여 5.3kbps ACELP (ITU-T 표준안 G.723.1) 보코더에 적용하였다. 실험에 사용한 보코더는 그림 <3-1>과 같다. 프레임 크기는 30msec이고 부프레임은 5msec이다. 성분분리 과정은 그림 <3-2>에 나타내었다.

먼저 입력신호에서 묵음부분과 음성부분을 분리과정은 기존의 G.723.1 ACELP에서 사용하는 VAD(Voice Activity Detection)을 사용하였다. 이것으로 검출된 음성부분에 대해서만 유/무성음 결정을 수행하였다. VAD가 1이면 음성부분인데 이런 구간에 대해 먼저 NL과 NH를 비교하여 NL이 크면 유성음으로 간주한다. 그렇지 않으면 LSP 파라미터중 간격이 최소가 되는 주파수 영역을 찾는다. 그림 <3-2>에서  $\min(intv)$ 는 LSP 파라미터들의 간격 중 최소가 되는 주파수 나타낸다. 여기서  $intv$ 는 LSP 파라미터들의 간격정보이다.

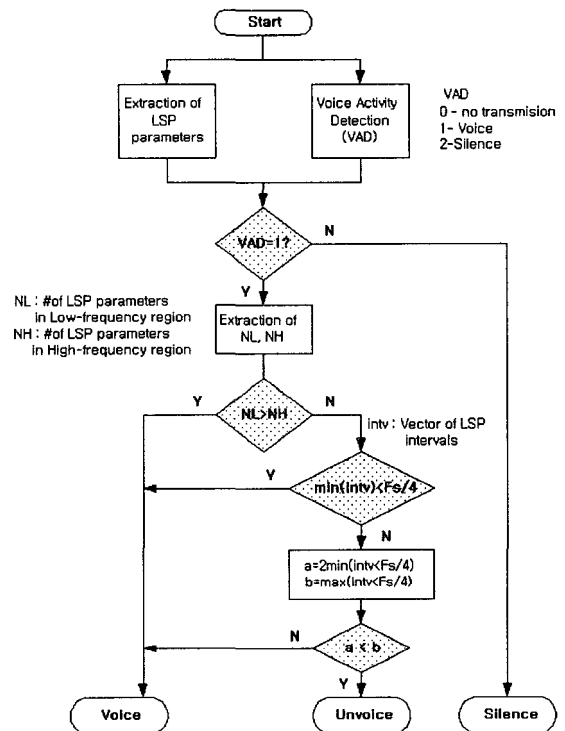
$$P = [p_1, p_2, p_3, \dots, p_{10}] \quad (3.1)$$

$$intv(i) = |p_{i+1} - p_i|, \quad 1 \leq i \leq 9$$

P는 LSP 파라미터이고,  $intv$ 는 간격정보를 나타낸다.



그림<3-1> 실험에 사용한 부호화기의 블록도



그림<3-2>제안한 성분분리 방법에 대한 순서도

$\min(intv)$ 가  $F_s/4$ 보다 작으면 제1포먼트가 저주파수 영역에서 나타난다고 생각하고 비록  $NL < NH$  이지만 유성음으로 결정한다.  $\min(intv < F_s/4)$ 는  $F_s/4$ 의 주파수 영역에서 나타나는 LSP의  $intv$ 중 가장 작은 것을 의미하고  $\max(intv < F_s/4)$ 는 가장 큰 것을 의미한다.

$\min(\text{intv}(F_s/4))$ 의 두배를 a라 하고  $\max(\text{intv}(F_s/4))$ 를 b라 한다면  $a < b$ 이면  $F_s/4$ 이하의 주파수 영역에서 좁은 간격의 LSP 파라미터들이 존재함을 반영하는 것으로 생각하고 유성음으로 결정하고  $a > b$ 이면 무성음으로 결정한다.

이렇게 결정된 유/무성음에 대해 각기 다른 부호화 방법을 적용하여 전송률을 줄이게 된다. 유성음은 기존의 방식을 사용하여 부호화 하고 무성음에 대해서는 LSP 파라미터와 Gain만 전송한다. LSP는 한 프레임을 4개의 부 프레임으로 나누고 각 부프레임별 LSP를 6 bit VQ로 양자화하여 24bit만 전송하고 신호에 대한 Gain을 6bit 전송한다. 그리고 이를 제어할 비트를 2bit 전송한다. 이렇게 하여 무성음 프레임에 대해서는 32비트를 전송한다. 기존의 G.723.1에서는 무성음에 대해서도 158bit를 전송하는데 비해 126bit를 줄이게 된다.

실험한 결과는 전송률과 음질로 평가하였다. 표 4-1은 G.723.1의 VAD만 사용한 보코더의 전송률과 LSP 파라미터를 이용하여 성분분리를 수행한 보코더의 전송률을 비교한 것이고 표 4-2는 음질을 비교한 것이다. 전송률은 제안한 방법이 기존의 5.3kbps ACELP보다 5.6%(280bps) 감소하였고 음질열하는 거의 없었다.

표 4-2 전송률 비교

	G.723.1 (5.3kbps)	Proposed Method	Degradation bps
발성 1	5.251	4.991	0.260
발성 2	4.656	4.447	0.209
발성 3	5.044	4.724	0.320
발성 4	4.999	4.670	0.329

표 4-3 기존 방법과 제안한 방법의 MOS 비교

	G.723.1 (5.3kbps)	Proposed Method
발성 1	3.84	3.8
발성 2	3.76	3.7
발성 3	3.8	3.78
발성 4	3.7	3.7

## 4. 결 론

음성부호화 시 처리될 음성이 유성음, 무성음 또는 묵음부분인지의 정보를 통해 중요도를 달리하거나 각 성분에 보다 적합한 파라미터를 사용하는 방법을 통해

비트를 줄이거나 음질을 향상시킬 수 있다. 따라서 음성 신호처리에서 성분분리의 전처리과정은 매우 중요하다.

본 논문에서는 CELP 보코더에서 LP 분석시 얻어지는 LSP 파라미터를 사용하여 유/무성음을 결정하여 전송률을 줄일 수 있는 방법을 제안하였다. 저주파수 영역과 고주파수 영역의 LSP 분포와 포먼트의 존재여부를 결정하는 LSP 간격정보를 이용하여 음성신호의 성분을 분리하였다. 이렇게 분리된 무성음 구간 대해 LSP 파라미터와 Gain정보만 전송함으로써 전송률을 낮출 수 있었다.

CELP형 보코더인 G.723.1의 5.3kbps에 제안한 방법을 적용한 결과 음질의 열하없이 5.6%의 전송률을 감소시킬 수 있었다.

## 6. 참 고 문 헌

- [1] Mabo Robert Ito, Robert W. Donaldson, "Zero-Crossing Measurements for Analysis and Recognition of Speech Sounds", IEEE Trans. on A.A.E. Vol. AU-19, No. 3 pp. 235-242, Sep., 1971.
- [2] B.S Atal, L.R.Rabiner, "A Pattern Recognition Approach to Voiced-Unvoiced-Silence Classification with Application to Speech Recognition", IEEE Trans. on ASSP, Vol. ASSP-24, No. 3. June, 1976.
- [3] A. M. Kondoz, "Digital Speech", John Wiley & Sons Ltd, 1994.
- [4] 배성근, 백금란, 배명진, 안수길, "음성신호의 진폭분포를 이용한 유/무성음 검출에 대한", 한국음향학회. 학술논문발표회 논문집, 제 12호, 제 1(s)호, 1993.
- [5] H. Kobatake, "Optimization of Voiced/Unvoiced Decisions in Noise Environments", IEEE Trans. on ASSP, Vol. ASSP-35, No. 1, pp. 9-18, Jan., 1987.
- [6] S.G. Knorr, "Reliable Voiced/Unvoiced Decision", IEEE Trans. on ASSP. Vol. ASSP-27, No. 3, June, 1979
- [7] L.R. Rabiner and R.W. Schafer, "Digital processing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.