

# CELP형 보코더에서 피치 검색오류의 개선 (On an Improving the pitch Searching Error of CELP Type Vocoder)

배 명 진\*, 장 호 성\*\*  
(Myung Jin Bae\*, Ho Sung Jang\*\*)

## 요 약

부호 여기된 선형예측(CELP) 음성보코더들은 4800bps의 낮은 전송율에서도 좋은 음질을 제공한다. CELP형 보코더의 피치검색법의 문제점중에 하나는 피치 검출시의 조오류(gross error)에 의해 피치 예측이득이 저하된다는 점이다. 본 논문에서 우리는 CELP보코더의 피치 예측이득을 개선하는 한 새로운 피치검색법을 제안하였다. 제안한 방법은 예비피치들을 검출하여 이들중 피치 예측이득이 최대인 값을 선정하는 방법을 적용하였다. 이 방법을 여러 화자의 발성에 대해 적용한 결과 피치 예측이득을 6.1% 정도 개선할 수 있었다.

## ABSTRACT

Code Excited Linear Prediction(CELP) speech coders exhibit good performance at data rates as low as 4800bps. One of the major drawback to the pitch search method in CELP type coders is their low pitch prediction gains occurred by gross errors. In this paper, we propose a new pitch search method that improves the pitch prediction gain of the CELP type vocoder. The basic idea is to reduce the pitch errors by selecting a maximum gain of the pitch prediction for preliminary pitches. As employing the proposed method to several utterances, we can get approximately 6.1% improvement in the pitch prediction gain.

## I. 서 론

음성신호를 메모리에 저장하거나 전송하기 위한 음성 부호화법에는 크게 파형부호화법, 신호원부호화법, 혼성부호화법 등이 있다.

파형부호화법은 음성파형 자체의 반복적이고 불필요한 잉여성분을 제거한 후에 저장 및 합성하는 방법이며 DPCM, ADM, ADPCM 등이 있다. 최근에는 디지털 신호처리 전용칩의 제조기술과 파형부호화법의 분석 및 합성 알고리즘이 잘 개발되어 32kbps 전송율을 갖는 ADPCM의 표준화가 실현되었다. 또한 파형부호화법은 인간의 개성과 감정을 대변해 주는

여기정보와 언어의 의사전달을 나타내는 여파기정보를 분리시키지 않고 처리하므로 고품위의 음질과 개성을 유지할 수 있으나, 테이타량이 방대하기 때문에 채널의 대역폭이나 저장용 메모리가 많이 소요된다는 단점이 있다.

신호원부호화법은 음성의 발성모델에 근거하여 여기정보와 여파기정보를 분리시켜서 부호화하는 방법으로 LPC, PARCOR, LSP, 포맷트부호화법, MBE법 등이 있다. 이들 알고리즘은 10kbps이내로 전송율을 낮출 수 있기 때문에 낮은 전송대역폭에 효율적인 방법이다.

신호원부호화의 메모리 효율성과 파형부호화의 자연성 및 명료성을 유지하기 위해 이 두가지 부호화법을 결합시킨 혼성부호화법으로는 MLPC, RELP, VELP, CELP 등의 기법이 제안되어져 있다. 혼성

\* 숭실대학교 정보통신공학사

\*\* 홍익대학교 전자전산기공학과

접수일자: 1993년 5월 9일

부호화법에서 포먼트정보는 선형 예측부호화법으로 보통 부호화하고, 그 나머지 잔여신호에 대해 어떻게 부호화하느냐에 따라 RELP법, VELP법, MPLP법, CELP법, 등이 제안되어져 있다. 혼성부호화법들 중에서도 최근 이동통신용 부호화법으로 채택되고 있는 것은 CELP(code excited linear prediction)법이다.

상기의 음성 부호화법들에서는 최적의 음질을 선택하기 위해 포먼트성분이 걸러진 잔여신호에 대해 피치 주기성분을 제거하는 피치 검색과정을 적용하고 있다. 피치 주기성분을 제거하기 위해서는 원래의 음성파형과 임의의 주기로 합성된 파형과의 상관관계가 최대인 값을 결정하게 된다. 그러나 이렇게 결정된 피치치연은 실제 피치주기의 반주기, 배주기, 삼배주기 등이 검출되는 부정확성이 발생하며 이때 피치필터의 이득은 크게 저하된다.

따라서 본 논문에서는 CELP보코더의 피치 검색시에 나타나는 피치 검출오차를 최소화 하는 방법을 새로이 제안하고자 한다.

## II. CELP 보코더의 원리

그림 2-1은 CELP 음성 부호화기의 블록도이다. 포먼트 합성필터는 10차의 LPC 전극형 구조가 보통 적용된다. LPC계수는 양자화시에 왜곡이 크기 때문에 LSP계수로 변환되어 부호화되고, 복호화시에 다시 LPC계수로 변환시킨다. LPC계수들은 20ms 한 프레임마다 부호화하고, 인터플레이션을 통해 5ms의 각 부프레임 마다 다르게 공급해준다. 여기위 파라미터는 또한 매 5ms 부프레임 마다 새롭게 바꾸어 준다.

CELP부호/복호기는 2개의 여기원을 사용한다. 첫 번째는 장기(피치) 예측상태 또는 적용 코드북이다. 두번째 여기원은 여기 코드북에서 취한다. 저 전송율의 경우에는 코드북의 코드워드는 128개 정도이다. 이들 2개의 여기원은 그들에 해당하는 이득항이 곱해진 다음에 합해진다. 이것은 결합된 여기시퀀스가 된다. 각 부프레임의 여기출력은 다음 부프레임에서 활용할 수 있도록 적용코드북의 장기 필터상태를 바꾸는데 적용된다.

CELP보코더의 부호화법에서는 포먼트정보가 걸러지고 남은 잔여신호에 대해 벡터양자화기법을 적용하여 부호화 하기 때문에, 잔여신호 성분 전송에 적용되는 데이터는 코드북의 인덱스가 된다. 따라서 전

송율을 4.8Kbps이하로 낮출 수 있으며, 이러한 파라미터를 오류정정용 부가코드와 함께 전송하면 전송 잡음에도 강인한 부호화법이 된다. 또한 이 부호화법은 합성에 의한 분석법을 적용하여 최적의 음질이 유지되도록 반복적으로 분석하기 때문에 주어진 저 전송율에서도 음질이 우수하게 된다.

CELP 보코더는 매번 음성을 합성해서 비교해야 하므로 복잡한 구조를 갖는다. 특히 부호화시에 계산량을 많이 요구하는데, 코드북에서 입력 여기신호를 찾는 과정과 피치필터 P(Z)의 계수를 구하는 과정에서 대부분의 계산시간을 소모하게 된다.

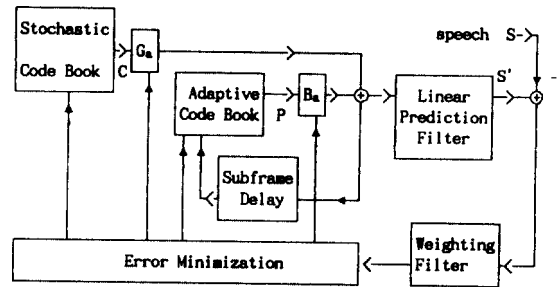


그림 2-1. CELP음성 부호화기  
Fig 2-1. a CELP speech coder.

## III. 피치검색법

피치검색 과정은 음성신호의 장기 상관관계에 해당하는 피치 주기정보를 얻는 과정이다. 이것은 8 KHz로 표본화된 음성의 경우 피치분석은 5ms 마다 한번씩 수행한다. 피치분석은 피치 지연조건을 최적으로 만족하는 값을 순차적인 비교를 통해 결정하게 된다. 이때 얻어진 최적의 피치치연을 이용하여 상관관계 비율은 피치이득이 된다. 이러한 피치 검색과정은 코드북검색과 함께 CELP 부호화기의 음질에 영향을 미치는 중요한 부분이다.

지금까지 피치검색의 개선을 위해 제안된 방법들은 자기 여기된 구조[6], 확장 적용코드북 구조[8], 델타 피치검색 구조[9], 등이 있는데, 이들은 보통 피치주기의 검색시 인근 피치주기간의 상관관계를 이용하여 피치 검색시간을 줄이는 방법을 사용하고 있다.

피치 검색과정은 원래 음성과 합성된 음성간의 피치 지연조건을 최적으로 만족하는 값을 반복적인 비교를 통해 결정하는 과정이다. 즉, 이것은 원래의 음

성신호에 대해 시간지연을 점차 변경하면서 자기 상관관계치를 검출하고, 이들중 최대의 상관관계치를 나타내는 시간지연을 피치주기로 검출하는 과정이 된다. 이렇게 계산된 상관관계치는 시간지연에 따라 그림 3-1(b) 처럼 나타난다. 여기서 시간지연이 음성과형의 주기성의 정수배에 해당될 때마다 상관관계가 최대치를 이루게 된다.

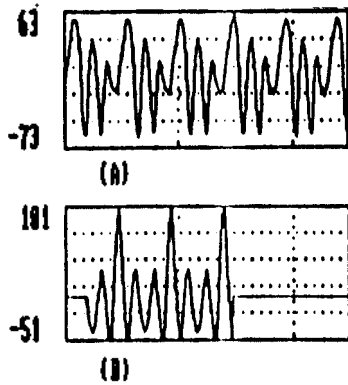


그림 3-1. 시간지연에 따른 상관관계값.

(a)음성과형 (b)자기 상관관계 파형

Fig 3-1. Auto correlation for time delay.

피치 검색과정에서 시간 지연범위는 인간이 실제로 발생가능한 피치범위인 2.5ms에서 25ms까지 수행해야 하나, 검색시간과 부호화율을 고려하여 2.5ms(20표본)에서 18.375ms(147표본) 까지만 검색한다. 그러나 이 검색 구간에서도 피치주기가 최대 7개 까지 나타날 수 있는 긴 구간이기 때문에, 이 구간내에서 음성과형의 진폭과 피치주기가 많이 변화될 수 있다.

CELP코더의 피치검색시에는 보통 고정된 부프레임 길이에 대해 상관관계를 계산하여 사용하고 있지만, 음성신호의 피치가 부프레임의 길이에 정수배가 아니기 때문에 검출오차가 많이 발생하게 된다. 즉, 피치 검색구간에 대해 상관관계가 최대인 값을 피치주기로 결정하게 되면, 음성과형의 실제 피치주기가 아닌 반주기, 배주기, 삼배주기 등의 검출오차가 발생하게 된다. CELP코더에서 음성신호의 피치가 정확히 검출되지 않으면 피치필터로 여과되지 않고 남은 잔여신호가 커져서 이를 불규칙 코드북단에서 흡수해야 하는 어려움이 발생하게 된다.

#### IV. 피치 검출오차의 개선

잔여신호  $s(n)$ 을 피치필터에 통과시킬 때 얻어지는 피치 예측이득  $G(L)$ 은 다음과 같다:

$$G(L) = \frac{\sum_{n=0}^{M-1} s(n)^2}{\sum_{n=0}^{M-1} (s(n) - s(n-L))^2} \quad (4-1)$$

여기서  $L$ 은 피치지연을 나타내고,  $M$ 은 부프레임의 길이를 나타낸다. 피치이득이 최대인 피치 지연값  $L$ 을 구하려면, 피치 검색구간에 대해 다음의 상관관계치  $E(L)$ 을 계산하여 이 값이 최대일 때의  $L$ 을 구해야 한다:

$$E(L) = \frac{E_{xy}^2}{E_{yy}} = \frac{\sum_{n=0}^{M-1} (s(n) s(n-L))^2}{\sum_{n=0}^{M-1} (s(n-L) s(n-L))} \quad (4-2)$$

이때 상관관계치는 피치 주기마다 100%에 근접한 값이 얻어지며, 유사한 정도가 어느 정도인지는 피치검색 구간내에 있는 파형의 주기성과 파형 진폭의 변화에 따라 다르다. 그리고  $E(L)$ 이 최대일 때의  $L$  값에서 피치필터의 계수  $b = E_{xy}/E_{yy}$ 를 취하게 된다.

그렇지만 식 4-1에서 계산하는 피치이득은 입력되는 신호에 무관하게 부프레임의 길이  $M$ 구간(보통 5ms)에 대해서 얻어지게 된다. 인간이 발생 가능한 실제 피치주기의 범위는 2.5ms에서 25ms이기 때문에 식 4-1에 의해 계산되는 피치이득은 한 피치구간의 성질을 적절히 만족시킬 수 없게 된다. 이 때문에 식 4-2을 통해 구한 피치지연 값은 한 피치지연내의 예측이득을 나타내기 보다는 부프레임 구간내에서의 최적 피치이득을 나타내게 된다. 이 때문에 식 4-1에 의해 검출된 피치이득은 실제의 피치구간에 대해 구한 피치이득 보다 낮아지게 된다.

본 논문에서는 이상의 피치 검색오차에 따른 피치필터의 예측오차를 최소화하기 위해 피치검색시에 상관관계를 감시하는 방법을 제안하였다. 식  $E_{xy}/E_{yy}$ 에 대해 시간지연 20표본에서 부터 1씩 증가시킬 때 마다 계산되는 상관관계값이 0을 넘으면서 양의 봉우리를 구성하는 경우에 이를 예비 피치구간으로 고려하여 봉우리 침두에서의 피치지연값과 이득을 보관

한다. 이렇게 하여 피치 검색구간 전체(편의상 20표본에서 147표본까지)에 대해 예비피치들을 감시한 다음에 이들 예비피치와 이득에 대해 가장 바람직한 피치를 검출하게 된다.

바람직한 피치주기의 결정은 앞에서 검출된 예비피치들의 피치지연과 계수를 사용하여 예비피치 구간내에서의 피치필터의 이득을 사용하게 된다. 즉, 다음 식과 같이 피치필터의 예측이득  $G(L_i)$ 을 구한 다음에, 이 값이 최대인 피치주기와 계수를 결정하게 된다.

$$G(L_i) = \frac{\sum_{n=0}^{L_i-1} (s(n) s(n))}{\sum_{n=0}^{L_i-1} (s(n) - b_i s(n-L))^2} \quad (4-3)$$

여기서  $L_i$ 와  $b_i$ 는  $i$ -번째 예비피치의 피치지연과 이득값이다. 예측된 피치이득이 최상이면, 이때 걸러지고 남은 잔류신호는 최소가 되기 때문에, 이때 잔류신호를 부호화하는 코드북 검색과정에서는 최소의 부호로 부호화를 수행할 수 있게 된다.

### V. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC/486DX2(50)이며 여기에 음성신호를 입출력하기 위한 AD/DA변환기를 인터페이스하여 8KHz의 표본율로 표본당 12비트의 데이터를 입력하였다. 처리결과의 성능을 평가하기 위해 다음의 대표적인 문장을 20세의 여성, 22세의 남성, 28세의 남성 화자가 각 5번씩 발성하여 시료로 사용하였다:

- 발성 1: "인수네 꼬마는 천재소년을 좋아한다."
- 발성 2: "예수님께서 천지창조의 교훈을 말씀하셨다."
- 발성 3: "송실대 정보통신공학과 음성처리 연구실이다."
- 발성 4: "공일이삼사오육칠팔구."

피치 검색과정은 그림 5-1과 같이 일반적인 CELP 보코더의 피치 검색과정을 수정하여 C-언어로 구현하였다. 이 처리과정에서 양의 상관관계 봉우리를 검색하는 블럭, 예비피치의 시간지연과 상관이득을 계산하는 블럭, 그리고 피치필터의 이득을 구하는 블럭은 우리가 제안하여 별도로 추가한 과정이다. 여기서  $1/A(z)$ 는 포먼트 여파기의 전달함수이고,  $A(z)/A$

$(z/a)$ 는 인지적인 가중화된 여파기응답이며, ZIR은 이전 상태의 영입력 여파기응답을 나타낸다. 그리고  $y_L(n)$ 은 피치지연  $L$ 에 대해 합성된 음성파형이고,  $E_{xy}$ 는 원래음성과 합성된 음성과의 교차 상관관계치이고,  $E_{yy}$ 는 합성된 음성의 자기 상관관계치이다.

피치검색의 성능을 비교하기 위해 처리과정은 크게 두가지로 나누어 수행하였다. 우선 피치검색 구간(20표본에서 147표본까지)에 대해 피치지연  $L$ 을 1씩 증가하여 검색하는 순차 피치검색법을 수행한 경우의 결과파형을 그림 5-2(b)와 (c)에 나타내었다. 그림 5-2(b)는 그림 5-2(a)의 파형에 대해 기존의 순차 피치검색법에 의해 구해진 피치변화도이고, 그림 5-2(c)는 이 때 얻어진 피치필터 이득을 나타낸다. 여기서 피치의 반주기, 배주기, 삼배주기가 많이 검출되었으며 이로인해 피치필터의 예측이득이 많이 저하되고 있음을 알 수 있다.

두번째로 처리한 과정은 반주기 피치의 검출을 억제하기 위해 우리가 제안한 방법이다. 순차 피치검색과정에서 양의 상관관계를 이루는 곳이 관찰되면 최상의 상관관계가 얻어지는 위치를 예비 피치지연과 이득으로 고려하게 된다. 이러한 예비피치들이 전체 피치 검색구간에 대해 찾아지면, 예비피치들 중에서 최적의 피치지연을 검출한다. 최적의 피치지연은 이들 예비피치와 이득을 그 예비피치 구간동안 피치필터에 통과시켰을 때, 피치이득이 최대인 피치필터의 지연값과 계수값을 선택하게 된다. 그림 5-2(d)에는 이렇게 검출된 피치변화도를 나타내었다. 여기서 피치의 배주기나 삼배주기가 검출되었을 지라도 기존의 방법에 비해 많이 제거되었다. 또한 이에 대한 피치필터의 예측이득을 그림 5-2(e)에 나타내었는데, 피치필터의 예측이득은 전반적으로 개선되어 나타났다.

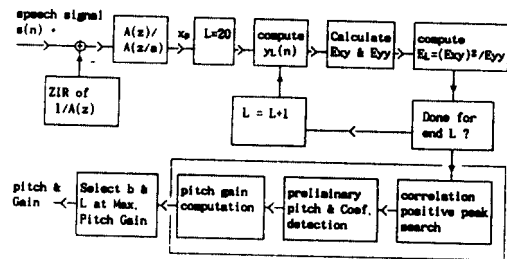


그림 5-1. 제안한 피치 검색과정  
Fig 5-1. The pitch search algorithm proposed in this paper.

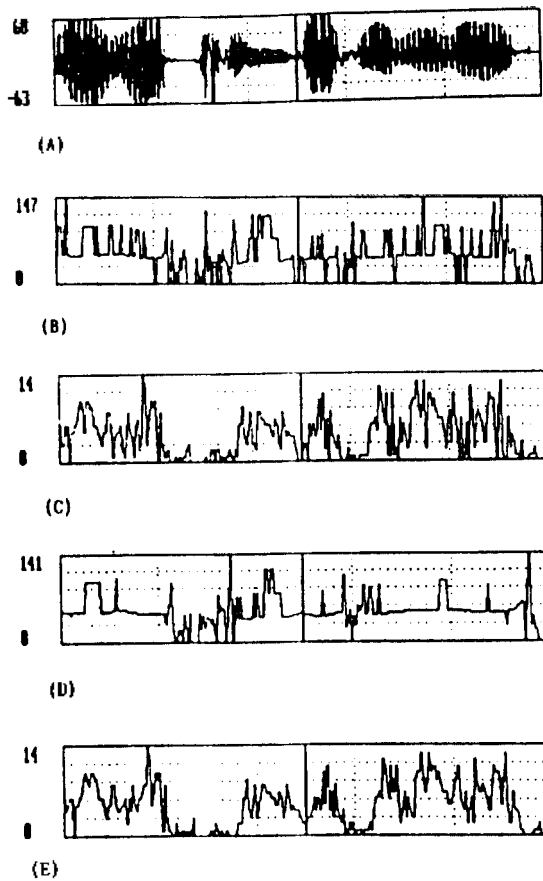


그림 5-2. 발성 1에 대해 처리한 결과: (a) 음성신호파형, (b) 순차 피치검색의 피치변화도, (c) 순차 피치검색에 따른 피치예측, (d) 개선된 피치변화도, (e) 개선된 피치예측.

Fig 5-2. A result for utterance 1.

기존 피치검색법에 대해 본 논문에서 제안한 피치검색법이 각 발성별, 화자별로 개선된 피치필터의 이득정도를 퍼센트율로 표 5-1에 제시하였다. 전반적으로 여성화자에 비해 남성화자의 경우가 더 많은 개선을 보이고 있는데, 이것은 여성화자의 피치가 피치검색구간인 부프레임 길이의 단방에 놓여있기 때문이다. 그리고 각 발성에 대해 개선된 양을 퍼센트로 표시한 것은 피치필터의 예측이득이 입력 음성신호의 레벨에 민감하기 때문에 표준화하는 의미에서 백분율로 나타내었다. 제안한 방법은 기존의 피치 검색법에 비해 계산량이 약 7.8%정도 증가되었으나, 여성화자인 경우 4.9%, 남성화자인 경우 7.3% 정도로 평균 6.1%의 예측이득이 개선되었다.

표 5-1. 피치 검출오차에 따른 피치필터의 예측이득

Table 5. Prediction gains of the pitch filter according to pitch detection errors.

(단위: dB)

	개선 이득율 (%)		평균
	여성	남성	
발성 1	6.0	11.5	8.8
발성 2	3.7	5.2	4.5
발성 3	6.1	5.4	5.8
발성 4	3.9	7.2	5.6
총 계	4.9	7.3	6.1

## VI. 결 론

CELP모코디의 피치검색에는 상관관계식을 적용하기 때문에 한 프레임내에서의 과형의 진폭변화와 피치변화에 아주 민감한 변화를 나타낸다. 따라서 피치 검색구간에 대해 부프레임 길이 단위로 계산된 상관관계치가 최대인 값을 피치주기로 결정하게 되면, 음성파형의 실제 피치주기가 아닌 반주기, 배주기, 삼배주기 등의 검출오차가 발생하게 된다. 따라서 본 논문에서는 피치 검출오차를 최소로 하는 방법을 새로이 제안하였다.

제안한 방법은 피치검색시에 상관관계치가 봉우리를 이루는 예비피치를 검출한 다음에 이들에 대해 실제의 피치를 검출하는 방법이다. 즉, 단순 피치검색 과정에서 상관관계값이 양의 봉우리를 구성하게 되면, 이들 예비 피치구간으로 고려하여 봉우리 정점에서 피치지연값과 이득을 보관한다. 이렇게 하여 얻어진 이들 예비피치와 이득에 대해 피치필터의 이득이 최대인 피치를 검출하는 방법으로 피치 검출오차를 개선하였다.

제안한 방법으로 피치를 검출하였을 때 기존 피치 검색법에 비해 피치필터의 예측 이득이 평균 6.1% 정도 개선되었다. 이때 기존의 피치검색법에 비해 증가된 처리량은 전체 계산량의 7.8% 이하로 무시될 정도이다.

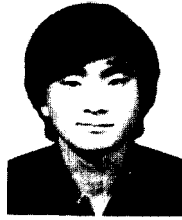
## 참 고 문 헌

1. J.D. Markel and A.H. Gray, Jr., *Linear Prediction of Speech*, Springer Verlag, New York, 1976.
2. L.R. Rabner and R.W. Schafer, *Digital Processing of Speech Signal*, Prentice-Hall, 1978.
3. A.N. Ince, *Digital Speech Processing(speech coding, sy-*

*thesis, and recognition*), Kluwer Academic Publishers, 1992.

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.-25.1.4, 1985.
5. G. Davidson and A. Gersho, "Complexity Reduction Methods for Vector Excitation Coding," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1986.
6. R.C. Rose and T.P. Barnwell III, "Quality Compression of Low Complexity 4800bps Self Excited and Code Excited Vocoders," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1987.
7. A.L. Guyader, D. Massaloux, and J.P. Petit, "Robust and Fast Code-Excited Linear Predictive Coding of Speech Signals," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.
8. J. Menez, C. Galand, M. Rosso, and F. Bottau, "Adaptive Code Excited Linear Predictive Coder (ACELPC)," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.
9. J.P. Campbell, Jr. V.C. Welch, and T.E. Tremain, "An Expandable Error-Protected 4800 bps CELP Coder(U.S. Fedral Standard 4800 bps Voice Coder)," Proc. Int. Conf. on Acoustics, Speech and Signal Processing, 1989.
10. W.B. Kleijn et al., "Fast Methods for the CELP Speech Coding Algorithm," IEEE Trans. Acoustics, Speech and Signal Processing, Vol.38, No.8, pp.1330-1341, Aug. 1990.
11. R.C. Rose and T.P. Barnwell, "Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders," IEEE Trans. Acoustics, Speech and Signal Processing, Vol.38, No.9, pp. 1489-1503, Sep. 1990.

▲배 명 진



1956년 5월 20일생  
 1977년 3월~1981년 2월 : 숭실대학교 전자공학과 (공학사)  
 1981년 3월~1983년 2월 : 서울대학교 대학원 전자공학과(공학석사)  
 1983년 3월~1987년 8월 : 서울대학교 대학원 전자공학과 수료(공학박사)

1986년 3월~1992년 8월 : 호서대학교 공과대학 전자공학과 조교수

1992년 8월~현재 : 숭실대학교 공과대학 정보통신공학과 부교수

관심분야 : 음성신호 처리

▲장 호 성



1949년 2월 15일생  
 1967년 3월~1971년 2월 : 수도공과대학 공업교육과 전기전공(공학사)

1974년 9월~1977년 2월 : 한양대학교 대학원 동력통신전공(공학석사)

1986년 3월~1991년 8월 : 단국대학교 대학원 전자공학과(공학박사)

1978년 3월~1992년 2월 : 홍익공업전문대학 전자공학과 부교수

1992년 3월~현재 : 홍익대학교 과기대학 전자전산기공학과 조교수

관심분야 : 디지털신호처리, 영상 및 음향처리.