

음성 데이터에 학습된 여기 코드북을 갖는 CELP 음성부호화기

正會員 이 인 성*, 전 선 화**

CELP Speech Coder with Optimized Excitation Codebook by Real Speech Data

Insung Lee*, Sunwha Jeon** *Regular Members*

※본 연구는 정보통신부의 대학기초연구지원사업(과제번호 : U96-158)의 지원으로 수행된 것임.

요 약

본 논문에서는 CELP 알고리즘에서 사용하고 있는 겹쳐지는 여기 코드북을 실제 음성 데이터를 사용하여 학습화하는 과정을 설명한다. 여기 코드북은 형태와 이득으로 구성되는데 형태 부분과 이득 부분을 각각 학습화된다. 코드북의 형태를 학습화하기 위해 비선형 비용함수는 정의되었고 비선형 방정식 최적화 알고리즘을 사용하여 비용함수를 최소화하는 코드북 변수들은 검색되어진다. 코드북 이득 양자화는 일반적 Lloyd 방법을 사용하여 최적화되어진다. 학습된 코드북의 사용은 기존의 QCELP 코드북을 사용하는 음성 부호화기 보다 약 1.2 dB SNR과 0.7 dB SEGSNR 이상의 성능 개선을 이룰 수 있었고 성능 개선은 비공식 음질 시험에서 확인되었다.

ABSTRACT

The training process of overlapped excitation codebook in a Code Excited Linear Prediction(CELP) coder using real speech data is described in this paper. The excitation codebook consists of the shape part and gain part in which each part is trained by real speech data. The nonlinear cost function is defined for the error criterion of optimization process. The elements of overlapped codebook is searched by the use of a nonlinear function optimization method. The general Lloyd algorithm is used for training of codebook gain quantizer. The trained excitation codebook by real speech data showed the performance improvement of 1.2 dB signal to noise ratio(SNR) and 0.7 dB segmental SNR(SEGSNR). The performance of new codebook was evaluated in an informal listening test.

*충북대학교 전자공학과

**충북대학교 공과대학부설 정보통신공학과

論文番號: 97010-0113

接受日字: 1997年 1月 13日

I. 서 론

현재 사용되고 있는 아날로그 이동통신 시스템의 가입자 용량의 제한성을 해결하기 위해 디지털 이동통신 시스템의 개발이 전세계적으로 활발히 이루어지고 있다. 디지털 이동통신 시스템에서 음성신호의 디지털 변환과 전송 데이터량을 줄이기 위해 여러 음성 부호화 기술이 사용되어지고 있다. 여러 부호화 방식중 최근에 개발된 분석/합성 구조의 Code Excited Linear Prediction(CELP) 음성 부호화 방식이 많이 사용되고 있다. 유럽의 TDMA(Time Division Multiple Access) 방식의 이동통신 시스템에서는 13 kbps 전송률을 갖는 RPE-LTP(Regular Pulse Excited Long Term Prediction)[1]와 북미와 일본에서는 8 kbps와 6.7 kbps에서 동작하는 VSELP(Vector Sum Excited Linear Prediction)[2] 음성 부호화기를 사용하고 있다. 미국과 한국의 CDMA(Code Division Multiple Access)방식의 이동통신 시스템에서는 미국 Qualcomm사에 의해 제안된 8 kbps 가변전송률 QCELP(Qualcomm Code Excited Linear Prediction) 알고리즘을 사용하고 있다 [3]. 98년부터 서비스되는 개인휴대통신 시스템(PCS)에서는 13 kbps 가변전송률 QCELP 음성 부호화기가 사용될 예정이다[4].

이러한 여러 종류의 CELP 음성 부호화기에서 가장 큰 차이는 여기(Excitation) 신호를 위한 벡터 양자화기의 코드북 구조와 이에 따르는 검색 방법에 있다. 이러한 변형된 구조의 여기 벡터 양자화 코드북을 사용함으로써 코드북 저장 메모리를 줄일 수 있고 코드북 검색시 계산량을 크게 줄일 수 있다. VSELP 알고리즘에서는 여기 벡터 신호를 기본 벡터의 합으로 구성하였고 기본 벡터는 실제 음성 데이터에 학습되어 사용되어진다. 한국의 CDMA 이동통신 시스템에서 사용되는 QCELP 알고리즘 및 미국의 연방 표준 FS1016 음성 부호화기 DOD CELP[5]등은 코드북 검색의 계산량과 코드북 저장 메모리량을 줄이기 위해 겹쳐지는(Overlapped) Stochastic 코드북[6]을 사용한다. 일반적으로 QCELP 알고리즘의 겹쳐지는 여기 코드북은 가우시안 불규칙수 발생기로부터 얻은 값을 사용한다. 가우시안 불규칙수 발생기로부터 얻은 값을 사용한 여기 코드북은 일반적으로 음성 여기 신호는 백색 특성을 갖는다는 특성 때문에 비교적 실

제 음성을 잘 표현하고 있다. 그러나 유성음에서는 여기신호는 백색 잡음이 아닌 성분이 많이 포함하고 있고, VSELP 음성 부호화기 설계 과정에서 보았듯이 실제 음성 데이터에 학습화된 코드북이 더 좋은 음질을 갖음을 알 수 있다. 따라서 본논문에서는 여기 코드북을 실제 한국어 음성 데이터에 학습화 함으로써 겹쳐진 여기 코드북을 갖는 CELP 음성 부호화기의 음질을 개선하려 한다. 본 논문에서는 겹쳐진 여기 코드북을 학습화하는 방법을 제시하고 실제 CELP 알고리즘에 적용하여 음질 개선을 주관적, 객관적 방법으로 확인한다.

II. 겹쳐진 여기 코드북을 갖는 CELP 음성 부호화기

본 논문에서는 겹쳐진 여기 코드북을 갖는 CELP 부호화기로서 QUALCOMM사에서 제안한 QCELP 음성 부호화기를 사용하였고 겹쳐진 여기 코드북의 학습화 과정은 QCELP 알고리즘 상에서 이루어진다. QCELP 음성 부호화 방식의 가장 큰 특징은 음성신호의 에너지량에 따라 전송률이 변하는 가변 전송률 부호화 방식이다. 이 방식은 양방향 통신방식인 통화 시스템에서 묵음시간이 전체 통화시간의 60-70% 정도를 차지한다는 점을 이용하여 음성 품질의 저하없이 낮은 전송률로써 음성신호를 부호화할 수 있다는 장점이 있다. 이는 CDMA 디지털 이동통신 시스템에서 채널 용량을 1.7 배 늘려주는 중요한 역할을 한다. 데이터 전송률은 한 프레임의 음성신호 에너지 값을 주변 잡음의 양에 따라 결정되는 3종류의 문턱값과 비교함으로써 결정된다. 전송률에 따라 각 파라미터에 할당되는 비트수와 갱신률을 달리한다.

QCELP 알고리즘의 인코딩 과정은 그림 2에 나타나 있다. QCELP 음성 부호화 방식의 인코딩시 입력 신호는 8 kHz로 샘플링한 다음 160 샘플을 갖는 20 ms 프레임으로 나누어진다. 자기상관 함수 방법을 이용하여 10차 LPC 계수를 계산하고 이렇게 결정된 LPC 계수는 LSP(Line Spectrum Pairs) 값으로 변환 후 각 전송률에 할당된 비트로 양자화된다. 각 피치와 코드북 부프레임에서 LSP값은 현재와 바로 전 프레임의 LSP값들을 보간함으로써 얻어진다. QCELP 알고리즘은 분석/합성 부호화 방식으로 입력 음성신호와 합

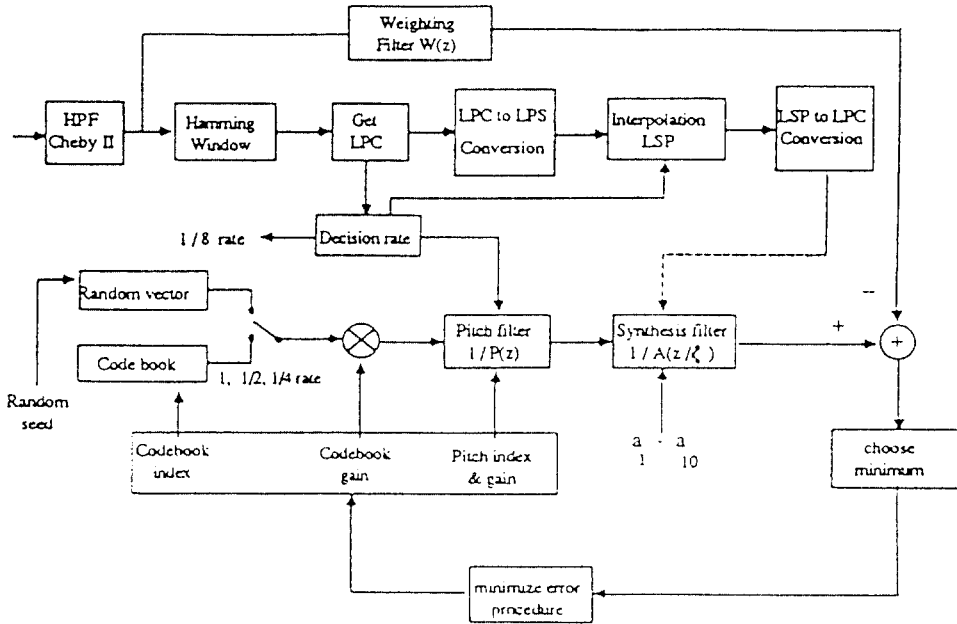


그림 1. QCELP 인코딩 과정

성 음성신호를 비교하여 그 오차를 줄이는 피치 파라미터와 코드북 파라미터를 검색하여 부호화한다.

III. 복잡도 줄인 여기(Excitation) 코드북 검색

QCELP 알고리즘은 코드북 검색의 계산량과 코드북 저장 메모리량을 줄이기 위해 128개의 값을 갖는 중복된 Stochastic 코드북을 사용한다. 코드북은 101개의 영의 값을 갖는 요소와 27개의 영이 아닌 요소로 구성된다. 128개의 값을 갖는 코드북에서 각 전송률에 따라 벡터 크기가 다른 128개의 여기 코드벡터가 만들어진다. 현재 인덱스의 여기 코드벡터는 바로 전 인덱스 코드벡터의 요소들을 한자리씩 이동하고 마지막 성분만 새로운 값으로 채워짐으로써 구성된다. 여기 코드북 검색 과정은 그림 2에 보여진다.

코드북의 순환적인 특성때문에 가중 합성필터 출력은 다음과 같이 간단하게 얻을 수 있다.

$$y(n) = \begin{cases} h(0)c((-l) \bmod 128) & n=0 \\ y_{l-1}(n-1) + h(n)c((-l) \bmod 128) & 1 \leq n < N_h \\ y_{l-1}(n-1) & N_h \leq n < L_c \end{cases} \quad (1)$$

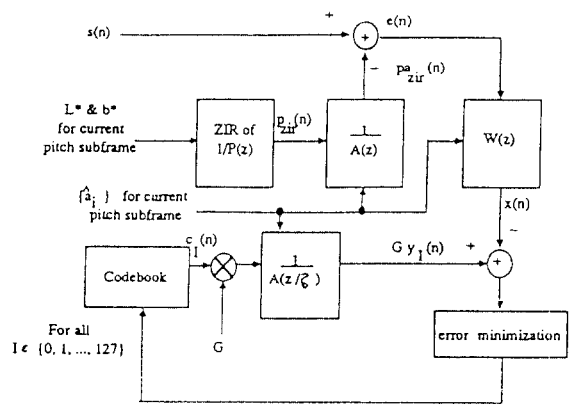


그림 2 여기 코드북 검색 과정

또한 $c((-l) \bmod 128) = 0$ 일때 합성필터 출력은 더욱 간단하다.

$$y(n) = \begin{cases} 0 & n=0 \\ y_{l-1}(n-1) & 1 \leq n < L_c \end{cases} \quad (2)$$

여기 코드북 검색시에도 적용 코드북 검색시와 같이 합성필터 출력의 에너지값과 상호상관값을 계산하여야 한다. 에너지값 E_{yyt} 의 계산은 적용 코드북 검색시와 같이 순환적으로 계산되어질 수 있다. 코드북 부프레임 크기 L_c 가 합성 필터의 임펄스 응답 길이 N_h 보다 클때 복잡도는 감소될 수 있다. N_h 를 20으로 하였을 때 코드북 부프레임 크기가 각각 40, 80샘플인 4 kbps, 2 kbps, 1 kbps의 전송률일때 에너지값 계산량을 줄일 수 있다.

$$E_{yyt} = \sum_{n=0}^{N_h-1} y_t^2(n) + S_t(n) \quad (3)$$

여기서 $S_t = S_{t-1} + y_t^2(N_h) - y_{t-1}^2(L_c - 1)$ 과 같다. 피치 부프레임의 크기 L_c 는 각 4 kbps, 2 kbps, 1 kbps 전송률에서 40, 80, 160 샘플이므로 가장 합성필터 임펄스 응답의 길이 N_h 를 20으로 하였을때 계산량은 상당히 줄어들 수 있다. 또한, $c((-l) \bmod 128) = 0$ 일때 E_{yyt} 의 계산은 다음과 같이 간단한 형태로 된다.

$$E_{yyt} = E_{yyt-1} - y_{t-1}^2(L_c - 1) \quad (4)$$

이 두가지 에너지값의 순환적인 계산은 직접적인 에너지값 E_{yyt} 의 계산시 보다 8 kbps에서 65%, 4 kbps에서 80%, 2 kbps에서 90%, 1kbps에서 95%의 곱하기량을 줄일 수 있다.

IV. 음성 데이터에 학습화된 여기 코드북

QCELP 알고리즘은 코드북 검색의 계산량과 코드북 저장 메모리량을 줄이기 위해 128개의 값을 갖는 겹쳐지는(Overlapped) Stochastic 코드북을 사용한다. 가우시안 랜덤 발생기로부터 만들어진 128 요소값은 1.2의 값에서 제로 클리핑(Zero Clipping)을 한다. 이렇게 만들어진 여기 코드북은 101개의 영의 값을 갖는 요소와 27개의 영이 아닌 요소로 구성된다. QCELP 알고리즘에서 128개의 여기신호 벡터를 만들어내는데 현재 인덱스의 여기 코드벡터는 바로 전 인덱스 코드벡터의 요소들을 한자리씩 이동하고 마지막 성분만 새로운 값으로 채워짐으로써 구성된다. 여기 신호 벡터 사이의 상관성을 줄이기 위해 DOD CELP에서는 두개의 성분을 새로운 값으로 채워진다. 이와같

은 2 샘플 달라진 겹쳐진 코드북은 256개의 요소를 요구한다.

QCELP 알고리즘은 가우시안 랜덤 발생기로부터 여기(Excitation) 코드북을 만들었고, 이웃 코드 벡터 사이에 단지 1 샘플만 차이가 나므로 코드북 메모리량은 줄일 수 있으나 이웃 코드 벡터 사이에 상관성이 비교적 높아 음질 저하를 발생시킬 수 있다. 좀 더 실제 음성에 잘 적용하기 위해 여기 코드북을 실제 한국어 음성 데이터에 학습화를 수행한다. 여기 코드북 값들은 중복된 코드북 구조때문에 비선형 방정식을 최적화함으로써 코드북을 얻을 수 있다. 128개의 변수를 갖는 비선형 코스트 함수 F는 다음과 같이 주어진다.

$$F(c_1, c_2, \dots, c_{128}) = \frac{1}{N} \sum_{n=1}^N \|x(n) - G(n)y(n)\|^2 \quad (5)$$

여기서 $G(n)$ 는 코드북 이득이고, $y(n)$ 는 코드북 벡터를 합성필터에 통과한 합성음성 신호 벡터이고, $x(n)$ 는 원하는 신호벡터를 나타낸다.

최적화 알고리즘은 코스트 함수 F를 최소화하는 코드북 변수 c_1, c_2, \dots, c_{128} 을 반복적으로 찾음으로써 학습된 코드북을 얻을 수 있다. 이 과정을 위해 비선형 방정식 최적화 알고리즘으로 Quasi-Newton 방법이나 함수의 미분 값을 구하지 않는 Powell 방법 등을 사용할 수 있다[6]. 최적화하는 과정에서 코드북은 101개의 영의 성분을 갖는다는 제한을 두어 최적화를 수행하였다. 8 kHz로 샘플링된 한국 음성 데이터를 사용하여 코드북 학습화 과정을 수행하였다.

각 변수의 최적화하는 과정시 128개의 코드북 요소 중 절대치 값이 큰 순서로 27개만 그대로의 값을 갖고 나머지 코드북 변수는 영으로 하여 코스트 함수 값을 계산한다. 최적화 알고리즘은 다음과 같이 수행되어진다.

1) QCELP 인코딩 알고리즘, 시작 코드북, 학습화에 사용될 한국어 음성 데이터를 최적화 알고리즘에 제공한다.

2) 지엽적 최소값을 찾을때까지 최적화 알고리즘을 사용하여 코드북을 최적화시킨다. 각 코드북 변수 최적화 과정 시 코스트 함수 계산은 절대치 값이 큰 순서로 27개의 변수만 그대로의 값을 사용하고 나머지 요소 값들은 영으로 한 코드북을 사용하여 계산한다.

3) 코드북에서 영의 위치는 고정시키고 영이 아닌 27개의 변수를 다시 한번 최적화시킨다. 2)의 과정에서 얻은 27개의 영이 아닌 요소를 갖는 코드북은 최적화 알고리즘의 시작 코드북으로 사용된다.

최적화 과정에서 처음의 코드북은 가우시안 랜덤 발생기로부터 얻은 값들을 사용한다. 코드북은 9개의 한국어어를 사용하여 학습화하였다. 사용된 한국어 문장은 4개의 여성 음성 데이터 5개의 남성 음성 데이터이다. 코스트 함수는 여러 지역적 최소값을 갖을 수 있기 때문에 여러 시작 코드북을 사용하여 최적화 함으로써 좋은 성능을 갖는 코드북을 얻을 수 있다.

코드북의 형태뿐만 아니라 코드북 이득 양자화기도 실제 음성 데이터로 학습시킬 수 있다. 코드북 이득 양자화를 위해 4 비트 이득 양자화기를 설계하였다. 총 가중 오차는 다음과 같이 코드북 이득의 함수로 표현할 수 있다.

$$D_i = \sum_{n \in N_i} \|x(n) - \eta(n)G_i y(n)\|^2 \quad (6)$$

여기서 최소의 가중 오차를 갖는 코드북 이득을 얻기 위해 미분을 수행하면

$$\frac{\partial D_i}{\partial G_i} = 0 \quad (7)$$

식(7)을 정리하면 다음과 같은 식을 얻을 수 있다.

$$G_i = \frac{\sum_{n \in N_i} \eta(n)x^T(n)y(n)}{\sum_{n \in N_i} \|y(n)\|^2}, \quad i = 1, \dots, 4 \quad (8)$$

실제의 QCELP 알고리즘에서 코드북 이득은 그림 3과 같이 DPCM 구조로 구성되어 있다. 양자화 후 코드북 이득은 다음과 같이 주어진다.

$$20 \log_{10}(G) = Q_G + P_G \quad (9)$$

로그 함수로 변환된 코드북 이득은 예측기의 값과 코드북 이득 양자화기의 출력 값의 합으로 주어진다. 따라서, 코드북 이득 양자화기는 다음과 같이 주어진다.

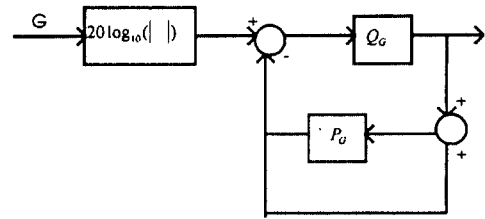


그림 3. DPCM 구조의 코드북 이득 양자화기

$$Q_G = 20 \log_{10} \left[\frac{\sum_{n \in N_i} \eta(n)x^T(n)y(n)}{\sum_{n \in N_i} \|y(n)\|^2} \right] - P_G(n) \quad (10)$$

코드북 이득 양자화기는 식(10)을 일반화된 Lloyd 알고리즘[7]에 적용하여 반복적으로 한국 음성 데이터로부터 계산되어진다. 여기서 $\eta(n)$ 는 코드북 이득의 부호이고, P'_G 는 DPCM 구조의 이득 양자화기에서 예측 이득을 나타내고, N_i 는 각 양자화 영역 크러스터를 표현한다. 4 비트 코드북 이득 양자화기는 영을 기준으로 대칭으로 구성되기 때문에 8개의 영역으로 나뉜다. 본래의 QCELP에서 일정한 간격을 갖는 코드북 이득 양자화기는 비균등 양자화기로 구성되어지며 균등 양자화기보다 약간의 성능 개선을 이룰 수 있었다.

이와 같이 코드북과 이득만을 교체하기 때문에 성능 개선을 위해 알고리즘을 바꿀 필요도 없고, 부가적으로 소요되는 복잡도나 메모리도 전혀 없다. 즉 학습화된 코드북으로의 단순한 대체로서 쉽게 성능 개선이 가능하다.

V. 성능평가 및 결과

개선된 여기 코드북의 성능을 평가하기 위해 8 kHz로 샘플링된 4개의 한국어 문장과 4개의 영어 문장을 사용하였고, 다음과 같다.

문장 1: 울림퓹은 전 인류의 축제입니다. (여자)

문장 2: 말 한마디로 천량 빛을 갚는다. (여자)

문장 3: 어제 산 물건이 벌써 고장이 났다. (남자)

문장 4: 일에서 십까지의 합은 오십오입니다. (남자)

Sent 1: The pipe began to rust while new. (Female)

Sent 2: Add the sum to the product of these three.
(Female)

Sent 3: Oak is strong and also gives shade. (Male)

Sent 4: Thieves who rob friends deserve jail. (Male)

학습화된 코드북과 본래의 QCELP 코드북의 사용 시 가변 전송률에서 QCELP의 SNR과 SEGSNR의 값은 표1에 나타나 있다. 학습화에 사용되지 않은 한국어 4문장, 영어 4문장을 사용하여 성능 비교를 수행하였다 평균적으로 SNR과 SEGSNR이 한국어의 경우 1 dB와 0.7 dB 개선되었고, 영어의 경우 1.44 dB와 0.7dB 각각 개선되었다. 표 1에서 특히 주목되는 것은 비록 한국어에 코드북을 학습하였으나 영어에서도 좋은 음질을 갖는다는 것을 알 수 있는데 이는 비록 다른 언어로 학습될지라도 언어의 기본 요소는 서로 유사하기 때문에 이를 전혀 고려하지 않는 QCELP의 Stochastic 코드북 보다는 좋은 성능을 보이는 것으로 생각되어진다.

당 연구실 내에서 8명의 참석자에 의해 이루어진 청취 테스트에서도 학습화된 코드북을 사용한 음성 부호화기가 훨씬 선명한 음질을 나타낸다는 결과를 보였다. 청취 테스트의 결과는 표 2에 나타나 있다. 대부분의 사람들은 새로운 코드북이 낫거나 같다고 표시하였다. 또한 새로운 코드북은 실제 시험용 CDMA 이동통신 시스템(Roving Test System)에 적용하였다. 이 실험에서도 대부분의 사람들은 학습화된 코드북을 갖는 음성 부호화기가 나은 음질을 갖는다고 선택하였다.

표 1. 학습된 코드북과 Qualcomm 코드북의 SNR/SEGSNR

	Stochastic 코드북		학습화된 코드북	
	SNR	SEGSNR	SNR	SEGSNR
한국어	15.30	12.61	15.90	13.16
영어	12.49	10.61	13.95	11.38

표 2. 두 코드북의 청취 테스트 결과

	Stochastic 코드북	학습화된 코드북	같다
선호도	7.9%	49.5%	42.6%

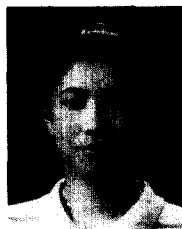
VI. 결 론

QCELP 알고리즘이나 4.8 kbps DOD CELP등은 메모리량과 계산량을 줄이기 위해 겹쳐지는 코드북 구조를 사용한다. 이러한 구조의 코드북은 단혀진 형태의 해를 구할 수 없어 실제 음성 데이터로 학습화를 수행하지 못했다. 본 논문에서는 비선형 방정식 최적화하는 방법을 사용하여 여기 코드북의 학습화를 수행하였다. 또한 본 논문에서는 코드북의 형태 뿐만 아니라 코드북 이득 양자화기를 학습화하는 방법을 제시하였다. 학습화된 코드북은 Stochastic 코드북보다 약 0.7 dB SEGSNR과 1.2 dB SNR 이상의 성능을 보였고 주관적 음질 테스트에서도 좋은 음질을 갖음을 보였다.

참 고 문 헌

1. J. E. Natvig, "Evaluation of Six Medium Bit-Rate Codecs for Pan-European Digital Mobile Radio System," IEEE J. Selected Areas in Commun., vol-6, pp. 346-352, 1988.
2. I. Gerson and M. A. Jasiuk, "Vector Sum Excited Linear Prediction (VSELP) Speech Coding at 8 kbps", in Proc. IEEE Int. Conf. Acoust., Speech, Sig. Processing, pp. 461-464, April 1990.
3. QUALCOMM Inc., TIA/EIA IS-96 Interim Standard, "Speech service option standard for wideband spread spectrum digital cellular system", April, 1993.
4. J. Campbell, V. Welch, and T. Tremain, "An expandable error-protected 4800 bps CELP coder," in Proc. IEEE Int. Conf. Acoust., Speech, Sig. Processing, pp. 735-738, 1989.
5. QUALCOMM Inc., TIA/EIA Interim Standard, "High rate speech service option standard for wideband spread spectrum digital cellular system", Feb., 1995.
6. D. Lin, "Speech Coding using Efficient Pseudo-Stochastic Block Codes," in Proc. IEEE Int. Conf. Acoust., Speech, Sig. Processing, pp. 1354-1357, April 1987.

7. W. B. Kleijin, D. J. Krasinski, and R. H. Ketchum, "Fast methods for the CELP speech coding algorithm," IEEE Trans. on Acoustics, Speech, and Signal Processing, pp. 1330-1342, Vol. 38, No. 8, Aug. 1990.
8. M. Avriel, "Nonlinear Programming: Analysis and Methods," Prentice-Hall, 1976.
9. M. Sabin and R. M. Gray, "Global convergence and empirical consistency of the generalized Lloyd algorithm," IEEE Trans. on Information Theory, vol. IT-32, pp. 148-155, March 1986.



全 善 華(Sunhwa Jeon) 정회원
1968년 2월 20일생
1992년 2월:충북대학교 전자공학과 졸업(공학사)
1994년 2월~현재:충북대학교 정보통신공학과 석사과정 재학중
1993년 2월~1995년 9월:충북대학교 공과대학부설 정보통신산업연구소 연구원
※주관심분야:음성 및 영상신호 알고리즘, VLSI 설계



李 寅 誠(Insung Lee) 정회원
1960년 6월 8일생
1983년 2월:연세대학교 전자공학과 졸업(공학사)
1985년 2월:연세대학교 전자공학과 졸업(공학석사)
1992년 12월:Texas A&M University 전자공학과 졸업(Ph.D.)

1986년 5월~1987년 7월:한국통신 연구개발 전임연구원
1986년 9월~1992년 8월:Texas A&M University 연구조교
1993년 2월~1995년 9월:한국전자통신연구소 이동통신기술연구단 선임연구원
1995년 10월~현재:충북대학교 전파공학과 조교수
※주관심분야:음성 및 영상신호 압축, 이동통신, 적응필터