

IS-127 Enhanced Variable Rate Codec의 성능 평가

정희원 정성교*, 박상욱*, 최용수**, 윤대희*

Performance Evaluation For IS-127 Enhanced Variable Rate Codec

Sung-Kyo Jung*, Sang-Wook Park*, Yong-Soo Choi**, and Dae-Hee Youn* *Regular Members*

요 약

IS-127 EVRC(Enhanced Variable Rate Codec)는 우리 나라와 북미의 CDMA 디지털 셀룰러 시스템에 채택되어 서비스 중에 있으며 8 kbps에서 우수한 성능을 갖는 부호화기이다. 본 논문에서는 EVRC의 성능을 전송 에러와 채널 상태에 대해 평가한다. 부호화된 신호의 전송 과정에서 발생하는 채널 에러는 종종 재생 신호에 심각한 왜곡을 야기한다. 채널 에러에 대한 EVRC 부호화기의 성능을 평가하기 위해 전송 에러에 대한 비트별 민감도를 측정하였다. 측정을 위하여 수신단으로 전송되는 비트 열에 인위적으로 에러를 발생시키고 이를 SNR, segSNR, BSD와 같은 객관적 음질 평가 척도를 사용하였다. 그리고 채널 상태를 나타내는 FER(Frame Erasure Rate)에 대한 음질 평가도 수행하였으며 FER과 주관적 음질 평가와의 높은 상관관계를 확인하였다.

ABSTRACT

IS-127 EVRC (Enhanced Variable Rate Codec) has been adopted as a standard coder for the CDMA digital cellular system, and known to provide good call quality at 8 kbps. In this thesis, the performance of EVRC is evaluated for the transmission errors and the channel conditions.

Severe distortions in the reconstructed speech are often caused by the errors in the transmission channel. To assess the sensitivity of EVRC to the channel errors, the output bitstream is artificially perturbed and the quality is measured using objective methods such as SNR, segSNR, and BSD. The quality of the reconstructed speech is also tested with FER (Frame Erasure Rate) and the correlation with the subjective evaluation is investigated.

I. 서 론

최근 들어 북미와 우리 나라의 디지털 셀룰러 시스템에서 가입자들에게 음성 신호의 보다 양호한 품질을 제공하기 위해 IS-127 EVRC (Enhanced Variable Rate Codec)^[1]으로 알려진 새로운 음성 부호화 알고리즘을 채택하여 서비스 중에 있다.

EVRC는 기존의 디지털 셀룰러 시스템에서 사용하던 8kbps QCELP(Qualcomm CELP)와 동일한 전송률을 갖지만 개인 휴대 통신인 PCS에서 서비스하고 있는 13 kbps QCELP에 버금가는 양호한

음질을 제공하고 있다. 실제 CDMA 채널을 통과하면서 왜곡된 신호는 통화 품질을 저하시키는 주된 요인이 된다. 이러한 채널 전송 과정에서 발생하는 왜곡에 효과적으로 대처한다면 양질의 신호를 가입자들에게 제공할 수 있다.

본 논문에서는 컴퓨터 모의 실험을 통해 EVRC 음성 부호화기의 채널 에러에 대한 성능 평가를 수행한다. 우선 객관적 음질 평가 방법을 사용하여 전송 에러에 대한 비트별 민감도를 측정한다. 그리고 다양한 FER(Frame Erasure Rate) 환경에 대해 주관적 음질 평가를 수행하고 FER과 주관적 음질

* 연세대학교 전기전자공학과 음향,음성,신호처리연구실(skjung@assp.yonsei.ac.kr),

** LG전자 디지털네트워크연구소

논문번호 : 00115-0402, 접수일자 : 2000년 4월 2일

* 본 연구는 현대전자산업 주식회사가 지원한 CDMA 핵심 기술 개발을 위한 공동 연구의 일부분으로 수행되었습니다.

평가와의 상관관계를 연구한다. 이를 바탕으로 FER 을 주관적 음질을 예측하기 위한 객관적 음질 평가 의 척도로 사용하는 방안에 대해 살펴본다.

본 논문의 구성은 다음과 같다. 우선 2장에서는 EVRC 음성 부호화에 대해 간략히 살펴보고 3 장에서는 주/객관적 음질 평가 방법과 이 두 방법에 대한 상관관계를 제시한다. 그리고 4 장에서는 채널 에러에 대해 EVRC 음성 부호화의 성능을 평가 하기 위해 전송 에러에 대한 비트별 민감도를 구한 다. 그리고 다양한 FER 환경에 대한 주/객관적 음 질 평가를 통해 FER과 주관적 음질과의 상관관계 에 대해 살펴본다. 끝으로 5 장에서 결론을 맺는다.

II. EVRC 음성 부호화기

IS-127 EVRC는 최근 우리 나라 CDMA 디지털 셀룰러 시스템에서 채택되었으며 부호화 알고리즘은 부호화 효율이 높은 일반화된 분석-합성 방법에 사용하는 RCELP(Relaxation CELP)^[2] 알고리즘을 기 반한다. 그리고 제한된 채널을 효과적으로 사용하기 위해 세 가지의 전송률을 가변적으로 사용한다. 전 송률은 입력 음성신호에 따라 부호화기 내부에 있 는 RDA(Rate Determination Algorithm)에서 결정 된다. 음성은 최대 전송률(Rate 1)과 중간 전송률 (Rate 1/2)로 부호화되고 배경 잡음을 최소 전송률 (Rate 1/8)로 부호화된다. 고역 필터링과 잡음 억제 과정을 통과한 입력 음성 신호는 20 ms의 프레임 단위로 선형 예측 분석 과정을 거치게 된다. 선형 예측 계수는 양자화 과정과 전송 과정에서 효과적인 LSP(Line Spectrum Pairs) 주파수로 변환한다. 잔차 신호 도메인에서 개회로 검색으로 지연을 추 정하고 장구간 예측 이득을 구한다.

적용 코드북의 지연 컨투어는 현재와 과거 피치 를 선형 보간하여 구한 피치 컨투어를 사용한다. 목 적 신호를 만들기 위해 이전 과정에서 수정된 잔차 신호와 현재 프레임의 지연 컨투어를 가지고 잔차 신호 수정을 위한 목적 신호를 만들고 이 신호에 최적으로 정합되도록 원래의 잔차 신호를 수정한다.

수정된 잔차 신호를 합성 필터에 통과시켜 수정 된 음성 신호를 구하고 가중 필터를 사용하여 수정 된 음성 신호를 가중한다. 수정된 가중 음성 신호에 서 적용 코드북의 기여를 제거하여 고정 코드북 탐 색을 위한 목적 신호를 생성한다. 고정 코드북은 ACELP(Algebraic CELP) 코드북을 사용하며 코드 북 탐색은 서브프레임 단위로 행한다. 코드북터는

최대 전송률인 경우에는 ±1의 크기를 갖는 8 개의 펄스를, 중간 전송률인 경우에는 3개의 펄스를 포함 한다.

III. 음질 평가 방법

체감 음질은 사용자가 귀로 직접 느끼는 부분으 로 주관적 음질 평가에 의해 측정된다. 주관적 음질 평가는 미리 녹음된 데이터를 사용하여 평가 대상 자들에 의하여 직접 이루어지므로, 많은 물적 비용 과 시간을 필요로 한다. 이러한 문제를 해결하기 위 하여 객관적 음질 평가 척도를 이용하여 주관적 음 질을 예측하는 연구가 활발히 진행되어 왔다.

3.1 주관적 음질 평가 방법

ITU-T는 Rec. P.800^[3]과 Rec. P.830^[4]에 주관적 음질 평가에 대하여 기술하고 있다. 주관적 음질 평 가는 서로 대화를 하며 상호간에 음질을 평가하는 Conversation-opinion 평가와 단방향으로 이루어지 는 Listening-opinion 평가로 이루어지는데, 일반적 으로 후자가 주로 행해진다. Listening-opinion 평가 는 주로 ACR(Absolute Category Rating)을 사용하 고, 평가 기준으로 MOS (Mean Opinion Score) 5 등급을 사용한다.

ACR에 사용되는 음성 시료는 최소 남녀 각각 2 명의 화자에 의하여 녹음되어야 하고 음성 부호화 기의 입력 레벨로는 -26 dBov가 가장 적당하다. 또 평가를 위한 듣기 레벨은 79 dB SPL에서 오차가 10 dB이내여야 한다.

3.2 객관적 음질 평가

체감 음질을 측정하기 위해 사용되는 주관적 음 질 평가는 많은 물적 비용과 시간을 필요로 한다. 이러한 문제를 해결하기 위하여 가장 많이 사용되는 객관적 평가 방법은 식 (1)과 (2)처럼 주어지는 신호 대 잡음비를 나타내는 SNR(Signal-to-Noise Ratio)과 구간별 SNR의 통계적 특성을 나타내는 segSNR(Segmental SNR)^[5]이다.

$$SNR = 10 \log_{10} \frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} [x(n) - y(n)]^2} \quad [dB], \quad (1)$$

$$segSNR = \frac{1}{L} \sum_{l=1}^L SNR_l \quad [dB], \quad (2)$$

여기서 $x(n)$, $y(n)$, N 은 각각 원음, 왜곡된 음

성, 샘플 수이다. SNR_k 은 k 번째 구간에서의 SNR 이며 L 은 프레임 수이다. 음성 신호는 비정적 신호이므로 SNR 보다는 구간별로 SNR 의 통계적 특성을 이용하는 $segSNR$ 이 주관적 음질과 상관관계가 높다. 그러나 SNR 과 $segSNR$ 은 신호 파형의 크기와 위상 성분에 매우 민감하므로, 주관적 음질을 예측하기에는 적합하지 않다. 스펙트럼 포락선 성분의 대수적 차이를 나타내는 선형 예측 켈스트럼 거리(LPC-CD: Linear Predictive Coefficient - Cepstral Distance)[5]는 선형 예측 계수를 이용하여 식 (3)과 같이 구한다.

$$LPC-CD \approx [2 \sum_{i=1}^K (C_x(i) - C_y(i))^2]^{\frac{1}{2}}, \quad (3)$$

여기서 $C_x(i)$, $C_y(i)$ 는 각각 원 음성과 왜곡된 음성의 선형 예측 켈스트럼 계수이고 K 는 선형 예측 계수의 차수이다.

LPC-CD 외에도 바크(Bark) 주파수 영역에서 스펙트럼의 차이를 나타내는 바크 스펙트럼 거리(BSD: Bark Spectral Distance)와 같이 쉽게 평가할 수 있는 객관적 음질 평가 척도를 이용하여 DAM (Diagnostic Acceptability Measure), MOS 등과 같은 주관적 음질을 예측하는 연구가 활발히 진행되어 왔다⁶⁾.

3.3 객관적 음질 평가와 주관적 음질 평가의 상관관계

주관적 음질 평가는 비용과 시간이 많이 소모되므로 주관적 음질과 상관관계가 높은 객관적 음질 평가 척도를 사용하여 주관적 음질을 예측하는 것이 보다 바람직하다. 객관적 음질 평가 척도로부터 주관적 음질을 예측하기 위해서는 아래와 같은 이차 MOS 예측함수가 널리 사용되고 있으며 비교적 우수한 성능을 나타낸다.

$$\widehat{MOS} = ax^2 + bx + c, \quad (4)$$

여기서 x 는 객관적 음질 평가 척도이며, a , b , c 는 least square regression 방법을 사용하여 구한다.

음질 예측 함수의 성능 평가를 위해서 아래 식을 사용하여 상관 계수를 구한다⁷⁾.

$$r = \sqrt{\frac{\sum (\widehat{MOS}_i - m_y)^2}{\sum (MOS_i - m_y)^2}}. \quad (5)$$

여기서 m_y 는 평균 MOS 값이고 \widehat{MOS}_i 는 예측된 MOS 값이다. 위 식은 식 (4)를 사용하여 MOS

를 예측했을 때 예측된 MOS 값과 실제 MOS 값과의 오차 정도를 의미한다. $r=1$ 은 MOS 예측 오차가 없음을 의미한다.

IV. 채널 에러에 대한 EVRC 성능 평가

4.1 파라미터별 에러 민감도

일반적으로 CELP 부호화기는 음성 대역에서 발생하는 음성이 아닌 신호나 주변 잡음에 의해 손상된 음성 신호 등과 같은 입력 신호에 대해서는 강인하지만 부호화된 신호를 전송할 때 발생하는 전송 에러에 대해서는 민감하다⁸⁾. 이러한 채널 에러는 복원된 음성 신호에 대해 심각한 왜곡을 야기하기도 한다. 보호 비트(protection bit)를 첨가하여 채널 에러의 효과를 줄일 수도 있지만 코더 자체가 갖고 있는 채널 에러의 민감도를 개선하는 것이 더욱 효과적이라고 할 수 있다. 채널 에러는 채널 환경에 따라 버스트 에러 혹은 랜덤 에러가 발생할 수 있다. 버스트 에러는 인터리빙 기법에 의해 보다 랜덤한 형태로 표현할 수 있으므로 랜덤 에러에 의해 발생하는 영향에 대해 초점을 맞추고자 한다.

표 1. 최대 전송률인 경우의 EVRC 비트 할당.

Parameters	Framesize (samples)	Bits	Bit Indices
LPC parameter	160	1+28	1, 2-29
Pitch delay	160	7+5	30-36,37-41
Subframe parameter	53	3+35+5	42-44,45-79, 80-84
- ACB gain	53	3+35+5	85-87,88-122, 123-127
- FCB shape	54	3+35+5	128-130,131-165, 166-170
- FCB gain			
reserved bit	160	1	171

본 논문에서는 EVRC 부호화기의 전송 에러에 대한 민감도를 살펴보기 위해서 파라미터별 에러 민감도를 측정하였다. 전송률은 최대 전송률로 고정하여 모든 프레임에 대해 동일수의 비트로 부호화한 다음 매 프레임마다 동일한 위치의 비트에 대해 에러를 발생시키고 에러에 대한 민감도를 객관적인 음질 평가 방법을 사용하여 측정하였다. 표 1은 최대 전송률인 경우의 EVRC 비트 할당을 보여준다. 프레임 단위로 LPC와 피치 파라미터를 수신단에 전송하고 서브프레임 단위로 적응 코드북 획득과 고정 코드북 인자 및 이득을 전송한다.

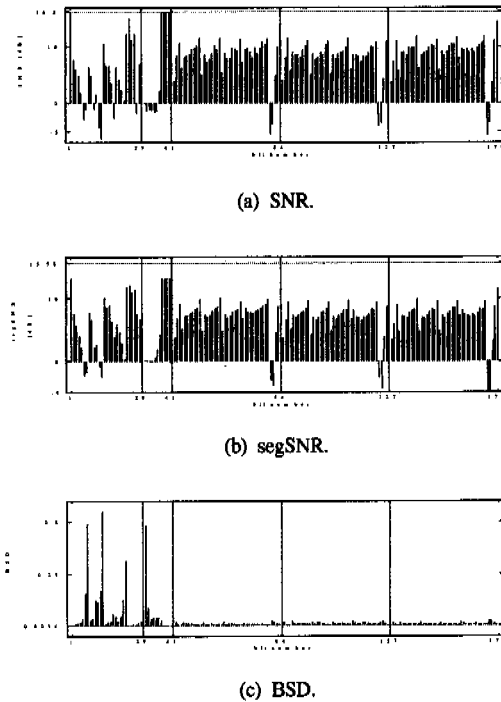


그림 2. 비트별 에러 민감도

그림 1은 각각 인위적으로 발생한 비트 에러에 대한 SNR, segSNR, BSD의 측정 결과를 보여주고 있다. 우선 SNR과 segSNR을 측정한 결과를 살펴보면, 에러에 가장 민감한 파라미터는 고정 코드북 이득이며 그 다음으로는 LPC 파라미터, 적응 코드북 인자, 적응 코드북 이득, 고정 코드북 인자의 순으로 에러에 민감하였다. SNR과 segSNR이 크기와 위상의 차이에 민감하기 때문에 고정 코드북 이득이 에러에 대해 가장 민감하게 나타났다. 이러한 민감도 측정 결과는 실제로 느끼는 주관적 음질과 거리가 있다. 주관적 음질과 높은 상관관계를 가지는 것으로 알려진 BSD 측정 결과를 보여준다. SNR 측정 결과와 달리 BSD 측정 결과에서는 LPC 파라미터가 에러에 가장 민감한 것으로 나타났다. 이는 LPC 파라미터가 채널 에러에 의해 손상되면 스펙트럼 거리가 커지기 때문이다. 그 다음으로 적응 코드북 인자, 고정 코드북 이득, 적응 코드북 이득, 고정 코드북 인자의 순서로 에러에 민감하였다.

EVRC에서는 LPC 계수를 LSP 파라미터로 변환하고 양자화하여 복호화단에 전송한다. 최대 전송률인 경우에는 4 개의 분리 코드북을 사용하여 각각 6, 6, 9, 7 비트를 할당하여 LSP 파라미터는 양자화한다. BSD 측정 결과에서 알수 있듯이 4 개의

분리 코드북에 할당된 비트의 MSB보다는 LSB에서 에러에 민감하며 높은 주파수 영역의 LSP보다 낮은 주파수 영역의 LSP가 에러에 더 민감하다는 결과를 얻었다. 피치 파라미터는 LSB보다는 MSB에서 발생한 에러가 음질에 더 큰 영향을 준다. 고정 코드북 인자와 적응 코드북 이득도 LSB보다는 MSB가 에러에 더 민감한 것으로 관찰되었다. 이와 같은 측정 결과는 일반적인 CELP 부호화기의 민감도를 측정하는 기존의 연구 결과^[23]와 유사하다.

4.2 FER에 대한 EVRC의 주/객관적 음질 평가 프레임 에러 검출 모듈은 데이터 전송률을 결정하고 프레임 에러를 검출하기 위하여 패킷 유형 지표를 사용한다. 프레임 에러는 실제 채널 상태에서 버스트 에러로 나타나므로 디지털 셀룰러 채널의 다양한 상태에 대한 모의 실험을 위해 GEC(Gilbert Elliott Channel) 모델을 사용하였다.

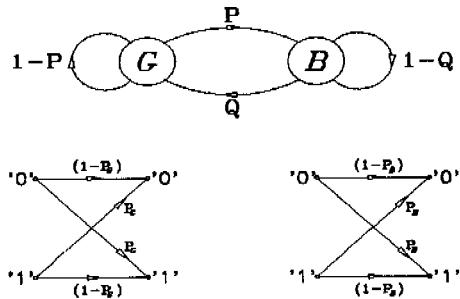


그림 2. Gilbert Elliott Channel Model (GEC).

그림 2는 GEC 모델로, G(Good)와 B(Bad or Burst) 상태로 구성되고 각 상태에 대한 천이 확률(transition probability)과 회기 확률(remaining probability) 등 4 가지 파라미터로 구성된다. 확률 P는 상태 G에서 상태 B로의 천이 확률이고, 확률 Q는 상태 B에서 상태 Q로의 천이 확률이다. 또 각 상태에서 비트 에러가 발생하는 확률은 각각 P_G 와 P_B 이다. 전체 BER은 식 (6)과 같다.

$$BER = \frac{P}{1-\gamma} P_B + \frac{Q}{1-\gamma} P_G \quad (6)$$

여기서 $\gamma = 1 - (P + Q)$ 이며 비트 에러의 상관계수이며 $\gamma \approx 1$ 은 버스트 채널을 의미한다. 본 실험에서는 버스트 채널 상태에 대한 모의 실험에서 $\gamma = 0.99$ 를 사용하였다.

주관적 음질 평가를 위해 남녀 화자 각각 1 명이

대해 약 3 초 길이의 음성을 녹음하였고, modified-IRS(Intermediate Response System)^[4] 가중을 사용하여 전화 대역 필터링을 하였다. 또 각 음성은 -26 dBov로 에너지 정량화를 한 후 EVRC 부호화기로 부호화하여 비트 패킷을 생성하였다. 생성된 비트 패킷은 위에서 설명한 GEC 모델을 사용하여 FER을 발생시킨 후 다시 복호화되었다. FER은 버스트한 상태로, 4 %에서 53 % 사이에 분포시켰다. ACR은 10 명의 평가자에 의하여 헤드셋을 사용하여 MOS를 평가하였다.

지털 셀룰러 시스템의 다양한 채널 상태에서 주관적 음질을 예측하는데 사용할 수 있는 보다 바람직하고 효과적인 평가 척도라 할 수 있다.

V. 결론

본 논문에서는 현재 우리 나라의 CDMA 디지털 셀룰러폰용 음성 부호화기인 QCELP(Qualcomm CELP)의 음질을 개선시키기 위해서 새롭게 채택된 EVRC 부호화기의 채널 에러에 대한 성능 평가도 수행하였다.

음질 평가 연구를 바탕으로 채널 에러에 대한 EVRC 성능 평가를 두 가지 측면에서 수행하였다. 부호화된 신호를 채널을 통해 전송할 때 발생하는 전송 에러는 파라미터에 따라 복원된 음성 신호에 대해 발생하는 왜곡의 정도가 다르다. 열악한 채널 환경에 대해 발생한 비트 에러를 복구하지 못하는 경우 프레임이 삭제하게 된다. 에러에 민감한 파라미터에 대해 에러 민감도를 개선하는 것이 채널 에러의 영향을 효과적으로 줄이는 방법이다. 채널 에러에 대한 민감도 측정 결과, 일반적인 CELP 부호화기와 같은 결과를 얻었다. 복원된 신호의 음질 저하가 심한 파라미터는 LPC 파라미터, 적응 코드북 인자, 고정 코드북 인자. 적응 코드북 이득, 고정 코드북 이득의 순으로 측정되었다. 이는 주관적 음질과 상관관계가 높은 BSD의 측정 결과와 일치한다. 또한 다양한 FER 환경에서 주/객관적 음질 평가를 수행하고 객관적 음질 평가 척도와 주관적 음질과의 상관관계를 측정한 결과, 심리 음향을 고려한 BSD와 무선 채널의 에러를 나타내는 FER이 주관적 음질 평가 척도인 MOS와 높은 상관관계를 나타내었다. BSD와는 달리 사용되는 음성이 아니라 무선환경에서의 채널의 상태에 따라 정해지는 FER을 주관적 음질을 예측하는 데 사용할 수 있는 객관적 음질 평가 척도로 적합하다.

실제 디지털 이동 전화 시스템에서 음질은 본 논문에서 살펴본 FER뿐만 아니라 이동 전화망과 연결되는 PSTN과 반향 제거기(echo canceller) 등의 여러 가지 요인과 관련이 있으므로 디지털 이동 전화 환경에서의 보다 폭 넓은 음질 평가 연구가 필요하다.

참고 문헌

[1] IS-127 Enhanced Variable Rate Codec, Speech

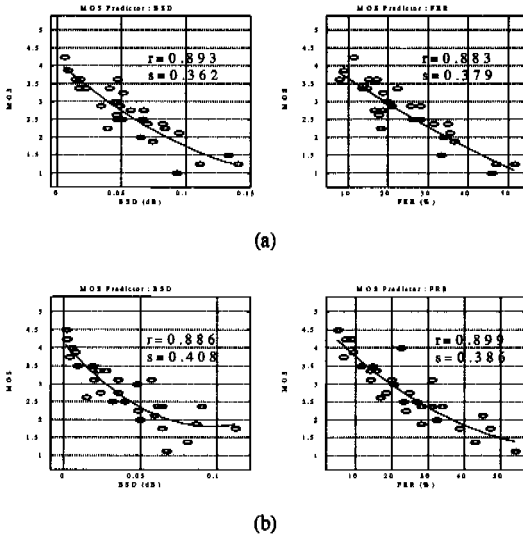


그림 3. 객관적 평가 척도와 MOS와의 상관관계. (a) 여성 화자. (b) 남성 화자.

FER에 대한 주관적 평가 결과와 MOS와의 상관관계를 그림 3에 나타내었다. 그림 3 (a)는 여성 화자의 경우이고 그림 3 (b)는 남성 화자의 경우이다. 각 그림에서 'o'는 실제 평가된 BSD와 FER과 MOS값이고 실선은 식 (4)에서 구한 MOS 예측 함수이다. MOS 예측 함수의 성능을 나타내는 상관계수는 'o'이 실선에 가까울수록 높은 값을 가진다. MOS와 높은 상관관계를 갖는 BSD의 경우처럼 FER도 MOS와 높은 상관관계를 가지고 있음을 알 수 있다. FER은 사용되는 음성에 상관없이 디지털 셀룰러 시스템의 전송 채널 상태에 따라 정해지므로 BSD와는 달리 유일한 MOS 예측함수 계수를 가진다. 그리고 복호화 과정을 거치지 않고 수신단의 상위 레벨에서 채널 상태에 따라 FER이 미리 결정되므로 BSD와 같이 복호화 과정 후에 구해지는 다른 객관적 척도에 비해 효과적이다. FER은 디

Service Option 3 for Wideband Spread Spectrum Digital Systems, Official Ballot Version, April 19, 1996.

[2] W. B. Kleijn, P. Kroon, and D. Nahumi, "The RCELP Speech-Coding Algorithm," *European Transactions on Telecomm.*, Vol 5, Number 5. Sept/Oct 1994, pp. 573-582.

[3] ITU-T Rec. P.800, "Method for subjective determination of transmission quality," Geneva, Switzerland, May 1996.

[4] ITU-T Rec. P.830, "Subjective Performance Assessment of Telephone-band and wideband Digital Codecs," 1996.

[5] Sadaoki Furui and M. Mohan Sondhi, *Advance in Speech Signal Processing*, Dekker.

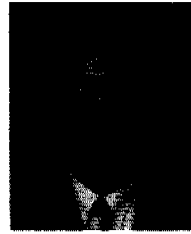
[6] Shihua Wang, et al, "An Objective Measures for Predicting Subjective Quality of Speech," *IEEE J. Select. Areas Commun.*, vol 10, No5, pp 819-829, June 1992.

[7] N.R. Draper, H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, 1981.

[8] R. V. Cox, W. B. Kleijn, and P. Kroon, "Robust CELP Coders for Noisy Backgrounds and Noisy Channels," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, 1989, pp. 739-742.

정성교(Sung-Kyo Jung)

정회원



1997년 2월 : 연세대학교

전자공학과 졸업

1999년 2월 : 연세대학교 대학원

전자공학과 석사

1999년 3월~현재 : 연세대학교

대학원 전기전자공학과

박사과정

<주관심 분야> 음성신호처리, 실시간 구현, VoIP

박삼욱(Sang-Wook Park)

정회원

1996년 2월 : 연세대학교 전자공학과 졸업

1998년 2월 : 연세대학교 대학원 전자공학과 석사

1998년 3월~현재 : 연세대학교 대학원 전기전자공

학과 박사과정

최용수(Yong-Soo Choi)

정회원

1993년 2월 : 연세대학교 전자공학과 졸업

1995년 2월 : 연세대학교 전자공학과 석사

1999년 6월 : 연세대학교 전기컴퓨터공학과 박사

1999년 9월~현재 : LG 정보통신

윤대희(Dae-Hee Youn)

정회원

한국통신학회 논문지 제24권, 제7A호 참조