

EVRC 코덱으로 재생하는 음악의 품질을 개선하기 위한 전처리 기법

정희원 남영한*, 하태균*, 전윤호*, 김재수*, 박섭형**

A Preprocessing Approach to Improving the Quality of the Music Produced by the EVRC

Young-Han Nam*, Tae-Kyoon Ha*, Yun-Ho Jeon*, Jae-Soo Kim*,
Seop-Hyeong Park** *Regular Members*

요 약

이 논문에서는 CDMA(Code Division Multiple Access) 음성 압축 표준의 하나인 EVRC(enhanced variable rate codec) 코덱으로 재생되는 음악의 품질을 개선하기 위한 전처리 알고리즘을 제안한다. EVRC는 음성을 압축하는 목적으로 최적화되었기 때문에, 음악을 압축하여 전송하면 품질을 심하게 저하시키기도 한다. EVRC로 음악을 압축할 때 발생하는 품질 저하 현상들 중에서 가장 심한 것의 하나가 끊김 현상인데, 끊김 현상은 프레임들이 연속적으로 Rate 1/8로 전송될 때 발생한다. EVRC는 장기 예측 이득의 크기를 바탕으로 입력 프레임들의 전송률을 결정하므로, 장기 예측 이득을 증가시켜 대부분의 프레임이 Rate 1 혹은 Rate 1/2로 압축될 수 있도록 하였다. 실험 결과를 보면 이러한 전처리 방법은 음악 신호에 대해 잘 적용되며 끊김 현상이 발생하는 프레임의 개수가 상당히 줄어들었음을 확인할 수 있다.

Key Words : EVRC, preprocessing, music enhancement, pitch harmonics enhancement, time-clipping

ABSTRACT

This paper proposes a preprocessing approach to improving the quality of the music produced by the EVRC(enhanced variable rate codec) which is one of the CDMA(Code Division Multiple Access) voice codecs. Since the EVRC is optimized only for speech signals, it can deteriorate the quality of the music passed through it. One of the problems with the EVRC-coded music is time-clipping, which usually occurs when subsequent frames are encoded at Rate 1/8. Since the EVRC determines the bit rate for an input frame based on the long-term prediction gain, we increase the long-term prediction gain in order for the most of the frames to be encoded at Rate 1 or Rate 1/2. Experimental results show that the approach works well on music signals and the number of time-clipped frames is considerably reduced.

1. 서론

이동 통신 시스템의 음성 채널 대역폭은 유선 통신 시스템의 64 kbps에 비해서 매우 작기 때문에 음성 신호를 압축하여 전송한다. 현재 국내의 이동 통신에 사용되는 음성 압축 방식에는 EVRC(enhanced variable rate codec)와 QCELP

(Qualcomm code excited linear prediction) 등이 있다. EVRC는 선형 예측 부호화(LPC: linear predictive coding) 분석 기반의 압축 방식을 사용하는데, 사람의 발성 구조에 최적화된 모델을 사용하고 있어서 사람의 음성을 중, 저 전송률로 압축하는데 매우 효율적이다^{[1][2]}.

*와이더댄컴(주) 기술연구소({namhan,taekyoon,yhjeon,jaekim}@widerthan.com), **한림대학교 정보통신공학부(spark@hallym.ac.kr)
논문번호 : 020494-1114, 접수일자 : 2002년 11월 14일

최근에 통화 대기 상태에서 신호음 대신에 가입자가 지정한 오디오 파일을 재생해 주는 서비스 등과 같이 이동 통신 단말기에 음악을 전송하는 경우가 점점 늘고 있다. 그러나 LPC 계열의 음성 압축 방법은 사람의 음성이 아닌 음악이나 주파수 스펙트럼이 복잡한 신호를 압축하기에는 적당하지 않기 때문에 이동통신 단말기로 듣는 음악의 품질이 심하게 저하되는 경우가 많다. 예를 들면, 음악 중간 부분이 일시적으로 단절되거나 일정 시간이 경과하면 음악이 들리지 않는 현상이 발생하는데, 이런 현상이 이동통신 시스템에서 음악 서비스의 가장 큰 장애 요인으로 작용한다.

이동 통신 시스템의 단말기에서 재생되는 오디오 신호의 품질을 향상시키는 방법에 관한 선행 연구들은 코덱이 부호화할 오디오 신호가 음악인지 음성인지를 확인하여 확인 결과에 따라 음악과 음성을 다른 방법으로 코딩하거나^[3], 음악의 경우 모두 Rate 1로 전송하는 방법을 제안하였다^[4]. 그러나 이런 방법들을 사용하기 위해서는 단말기나 교환기에 포함되어 있는 음성 코덱을 교체해야 하는 현실적인 어려움이 있다. 또한, 음악과 음성의 구별 과정에서 에러가 발생할 수도 있다.

본 논문에서는 단말기에서 재생되는 오디오 신호의 품질을 향상시키기 위한 현실적인 접근 방법으로 음악 서비스에 사용되는 원천 음악 파일을 전처리 하는 방법을 제안한다. 구체적으로, 오디오 신호를 EVRC를 통해 전송할 때 발생하는 끊김 현상을 완화할 수 있는 전처리 기법을 제안한다.

II. EVRC로 음악을 압축/복원할 때 끊김 현상이 발생하는 원인

CDMA 단말기에서 음악의 끊김 현상(time-clipping)이 발생하는 원인은 EVRC의 가변 전송률 때문이다. EVRC는 채널의 효율을 높이고 시스템의 소비 전력을 줄이기 위해 사람이 말을 할 때만 음성을 압축하여 전달하고 사람이 말을 하지 않을 때는 음성을 전달하지 않는 방법을 사용한다.

프레임 전송률은 추정 잡음(background noise estimate)에 비례하는 전송률 결정 임계치(RDT; rate determination threshold)와 대역 에너지(band energy)에 의해서 결정되는데, 추정 잡음이 커지면 같은 대역 에너지를 갖는 프레임이 낮은 전송률로 부호화될 확률이 높아진다. EVRC의 추정 잡음은 주 피치(dominant pitch)의 검출 여부에 따라 갱신되는데, 주 피치가 검출되지 않으면 추정 잡음이 점차 증가한다. 사람의 목소리와는 달리 여러 화음이 포함된 음악에서는 주 피치가 검출되지 않는 경우가 많다. 이런 이유 때문에 음악의 경우에는 끊김 현상이 심각하며, 특히 시간이 경과하면서 끊김 현상이 점차 빈번하게 발생한다.

그림 1은 EVRC 부호기^[5]의 블록도이다. 입력 $s[n]$ 은 8kHz, 16bit로 디지털화된 PCM 신호이며, 20ms에 해당하는 160 샘플로 구성된 프레임 단위로 처리된다. 입력 신호는 차단 주파수가 120Hz인 고역통과 필터를 통과한 후에 잡음 억제기(noise suppressor)를 통과한다. 모델 파라미터 예측기(model parameter estimator)는 $s'[n]$ 으로부터 LPC 계수, 단기 예측 오차(short term prediction residual), 피치(pitch), 그리고 장기 예측 이득(long-term prediction gain) 등을 계산한다. 전송률 결정기(rate determinator)는 이 모델 파라미터들을 사용하여 전송률을 결정한다.

EVRC의 전송률은 각각 'Rate 1', 'Rate 1/2',

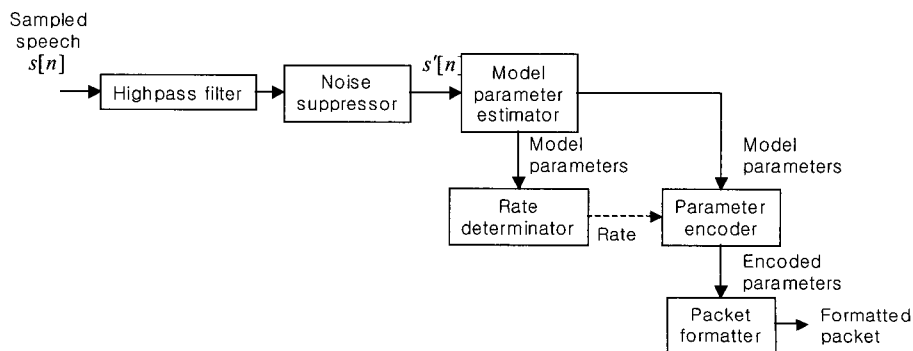


그림 1. EVRC 부호기의 고수준 블록도
Fig. 1. A high-level block diagram of the EVRC encoder

'Rate 1/8'의 세 가지이다. 파라미터 부호기(parameter encoder)는 전송률에 따라 파라미터를 부호화하여 비트열을 출력하며, 패킷 포맷터(packet formatter)는 비트열을 출력 패킷 포맷으로 변환한다. EVRC가 프레임의 음성 구간으로 인식하면 그 프레임은 Rate 1로 전송되며, 잡음 구간으로 인식하면 Rate 1/8로 전송된다. Rate 1/2은 전송률이 Rate 1에서 Rate 1/8로 변화할 때 거쳐 가는 전송률이다.

EVRC 부호기는 통과 대역이 각각 0.3~2.0 kHz와 2.0~4.0 kHz인 두 개의 대역통과 필터를 사용하여 잡음 억제 블록을 통과한 신호 $s'[n]$ 을 두 개의 대역 신호로 분리한다. 여기서 두 대역을 각각 대역 1과 대역 2라고 하자. EVRC 부호기는 각 대역 신호의 대역 에너지(band energy)와 전송률 결정 임계치(RDT: rate decision threshold)들을 비교하여 전송률을 결정한다.

대역 에너지는 $s'[n]$ 의 자기상관 계수들의 합에

의해 결정되며, 전송률 결정 임계치는 추정 잡음 레벨에 EVRC가 정의한 배율이 곱해져 계산된다. 이에 관한 자세한 내용은 [5]에 정의되어 있다. EVRC는 m 번째 프레임의 대역 i 의 추정 잡음 레벨 $B_{m,i}$ 를 그림 2의 순서도와 같이 갱신한다.

그림 2의 순서도에서 β 는 장기 예측 이득, $B1, B2$ 는 각각 대역 1, 2의 추정 잡음, $E1, E2$ 는 각각 대역 1, 2의 평활화 대역 에너지(smoothed band energy), 그리고 $SNR1, SNR2$ 는 각각 대역 1, 2의 신호 대 잡음비를 각각 의미한다.

그림 2의 순서도에 따르면 추정 잡음은 주로 β 가 작은 값으로 지속될 때(연속 8 프레임 이상 동안 β 가 0.3 미만일 때) 증가하게 되고, 평활화 대역 에너지가 기존의 임계치보다 작을 때, 평활화 대역 에너지에 맞추어 감소하게 된다. 여기에서 β 는 오차의 최대 자기상관과 오차 에너지의 비로 결정되며, 프레임 내에 주 피치가 존재하면 크고, 그렇지 않은 경우에는 작다. 이와 같은 RDT 증감 메커니

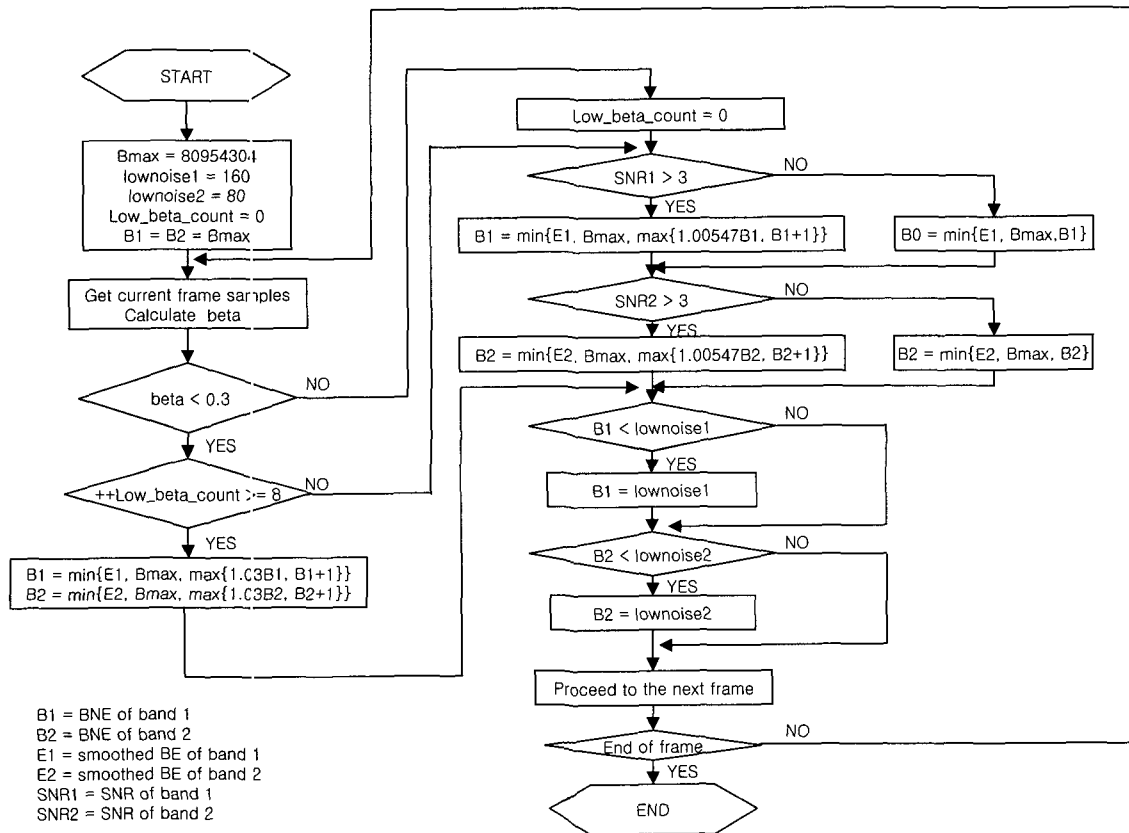


그림 2. 추정 잡음 갱신 알고리즘
 Fig. 2. The updating algorithm of the background noise estimate

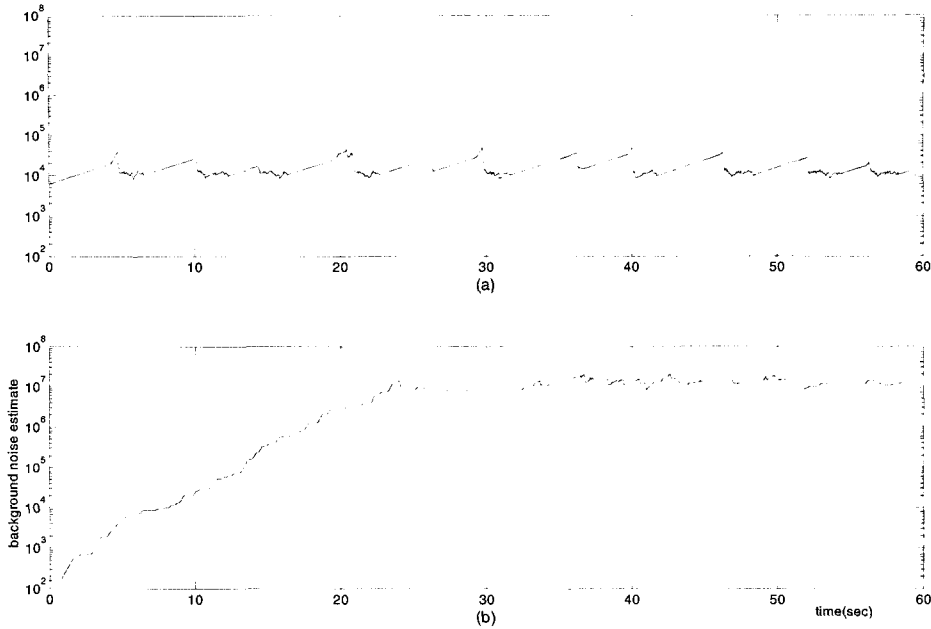


그림 3. 통화 음성과 음악을 각각 EVRC 부호기로 압축할 때 대역 1 추정 잡음의 변화 추이 : (a) 통화 음성 (b) 음악
 Fig. 3. Plots of Background noise estimate of Band 1 of the signals being compressed by an EVRC encoder: (a) a speech signal (b) a music signal

좁은 이동통신에서 음성 신호를 주로 전송하는 경우에는 적절하다. 음성 통화 상황에서 통화자의 음성 신호에는 반드시 말을 쉬는 구간이 포함되어 있으므로 추정잡음의 갱신 알고리즘에 의하여 통화자가 말을 쉬는 구간에서 추정 잡음이 줄어들고, 주변 잡음에 따라 적절하게 추정 잡음이 갱신된다.

반면에 음악 신호에는 일반적으로 묵음 구간이 전혀 없으며 주 피치가 검출되지 않는 경우가 많으므로 추정 잡음은 시간이 경과하면서 처음에 설정된 값보다 1.03배 혹은 1.00547배 등으로 점차 증가하고, 어느 순간 추정 잡음의 크기가 평활화 대역 에너지보다 커지게 되는 경우에만 추정 잡음의 다음 갱신값이 감소한다. 전송률 결정 임계치는 추정 잡음에 직접 비례하므로, 평활화 대역 에너지가 비교적 일정한 범위를 유지한다면, 시간의 경과에 따라 전송률 결정 임계치의 값이 점점 증가하고, 그에 따라 같은 대역 에너지를 갖는 프레임에 대해 그 프레임이 'Rate 1/8'로 부호화될 확률이 높아진다. 즉, 음악을 재생하는 시간이 길어질수록 끊김 현상의 발생 빈도가 증가할 수 있다.

그림 3과 4는 1분 길이의 통화 음성과 음악을 각각 EVRC 부호기로 압축할 때 대역 1의 추정 잡음

과 RDT의 변화 추이를 나타낸 것이다. 음성 통화의 경우에는 수화자가 말하는 동안 통화자는 말을 쉬게 되므로 묵음 구간에서 추정 잡음이 감소하지만, 음악의 경우에는 묵음 구간이 전혀 없으므로 RDT는 계속적으로 증가하는 것을 확인할 수 있다. 이제 음악의 끊김 현상에 가장 큰 영향을 미치는 장기 예측 이득 β 에 대해서 살펴보기로 한다. 나머지 파라미터들을 계산하는 방법은 [5]를 참고한다.

β 는 식 (1)과 같이 예측 오차(prediction residual)의 자기상관(autocorrelation)에 의해 정의된다.

$$\beta = \max \left\{ 0, \min \left\{ 1, \frac{R_{\epsilon}[0]}{R_{\epsilon}[0]} \right\} \right\} \quad (1)$$

여기서, ϵ 은 예측 오차 신호를, $R_{\epsilon}[0]$ 는 예측 오차 신호의 에너지를 각각 의미한다. 그리고 R_{\max} 는 양의 자기 상관 첨두치의 최대값으로 식 (2)와 같이 정의된다.

$$R_{\max} = \max_{20 \leq d \leq 120} R_{\epsilon}[d] \quad (2)$$

β 의 크기는 주 피치 성분의 존재 여부에 따라 달라진다. 그림 5는 주 피치 성분이 존재하는 음악 신호와 그렇지 않은 음악 신호의 $R_e[0]$ 에 대해 정규화된 오차의 자기 상관을 나타낸 것이다. 그림 5에서 원으로 표시된 부분이 β 의 위치이다. 식 (1)과 (2)에서 살펴본 바와 같이 오차의 정규화된 침두 자기 상관인 β 를 의미하는데, 주 피치가 존재하는 신호인 (a)에서는 β 가 0.6792로 상대적으로 크고, 주 피치가 존재하지 않는 신호인 (b)에서는 β 가 0.2616으로 상대적으로 작다.

III. 제안하는 전처리 알고리즘

이 논문에서 제안하는 음악 전처리 과정은 음폭 압축(dynamic range compression)과 피치 강화(pitch enhancement)로 이루어진다.

1. 음폭 압축 - CDMA 단말기에서의 음질 최적화와 대역 에너지 증가

특정 음악의 경우 입력 오디오 신호의 음폭이 이동 통신 시스템이 감당할 수 있는 음폭보다 넓어서

소리가 작은 부분에서 끊김 현상이 발생하거나 큰 부분에서 음의 포화현상이 나타날 수 있다. 음성 통화 상황이라면 에너지가 작은 프레임은 잡음 프레임으로 간주할 수 있지만, 음악의 경우에는 에너지가 작다고 해서 잡음 프레임으로 생각할 수는 없다. 또, 이동 통신 시스템이 감당할 수 있는 음폭은 매우 좁고, EVRC의 전송률 결정 알고리즘은 에너지가 작은 프레임을 잡음으로 취급하는 경향이 있으므로, 음폭이 넓은 CD 품질 오디오에 대해 음폭 압축을 수행할 필요가 있다. 음폭 압축을 위해서 AGC^[6]를 사용하였다.

음폭 압축은 CDMA 단말기에서 재생되는 음악의 품질을 최적화할 뿐만 아니라 전송률 결정 알고리즘에서 대역 에너지를 증가시킨다. 대역 i 의 대역 에너지는 대역 통과 입력 신호의 0에서 16번째 자기 상관 계수의 합에 의해 정해지는데^[5], 음폭 압축을 수행하면 음폭의 변화가 작아져서 한 프레임 안에서 낮은 차수의 자기 상관 계수가 커지므로 대역 에너지가 증가된다.

2. 피치 강화 - RDT를 작게 유지하는 방법

RDT를 작게 유지하려면 추정 잡음 갱신 과정에서 RDT가 증가되는 요인을 제거해야 하는데, 이를

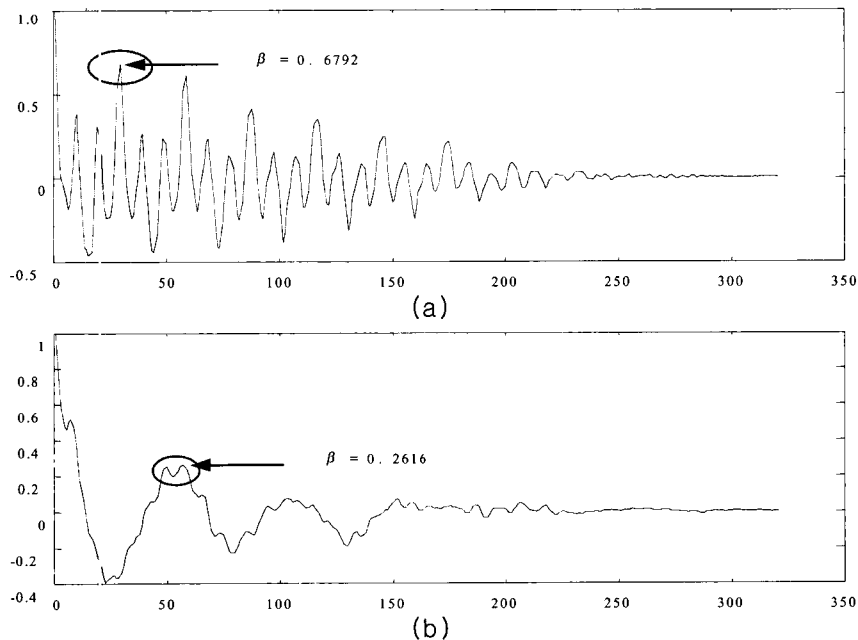


그림 5. 음악 신호의 프레임 오차 자기 상관 : (a) 주 피치 성분이 존재하는 경우 (b) 주 피치 성분이 존재하지 않는 경우
 Fig. 5. Autocorrelation of the frame residual of music signals : (a) when a dominant pitch exists (b) when a dominant pitch does not exist

위해 각 프레임에 대해 프레임의 장기 예측 이득 β 가 증가 되도록 하는 피치 강화 전처리를 수행하였다. 피치 강화 전처리는 편의상 주파수 영역에서 이루어지며, 기본 주파수(fundamental frequency)와 그의 배음 성분들(harmonics)의 크기를 증가시키는 전처리이다.

전체 오디오 신호 $s[n]$ 를 160 샘플 단위의 프레임 신호 $s_m[n]$ 으로 분할한다. 여기서 m 은 프레임 번호이고 n 은 프레임 내 샘플 번호로 $0 \leq n < 160$ 의 값을 갖는다. 각 프레임 신호 $s_m[n]$ 을 EVRC의 피치 검출 윈도우와 일치하도록 아래 식 (3), (4)와 같이 두 개의 신호 $s_{m,1}[n]$ 과 $s_{m,2}[n]$ 으로 분리하고, 각 신호에서 자기상관 피치 검출법 (Autocorrelation Pitch Estimation)^[7]으로 피치 p_1 과 p_2 를 구한다.

$$s_{m,1}[n] = \begin{cases} s_{m-1}[n+80], & 0 \leq n < 80 \\ s_m[n-80], & 80 \leq n < 160 \end{cases} \quad (3)$$

$$s_{m,2}[n] = s_m[n], 0 \leq n < 160 \quad (4)$$

각 신호 $s_{m,1}[n]$ 과 $s_{m,2}[n]$ 에 대해 160의 크기를 갖는 식 (6)의 해닝 윈도우(Hanning Window) $w[n]$ 을 아래 식 (5)와 같이 곱하고, 이를 256-point DFT하여 $S_{m,i}[k]$ ($i=1, 2, 0 \leq k < 256$)를 얻는다.

$$\tilde{s}_{m,i}[n] = w[n] \times s_{m,i}[n], i=1, 2, 0 \leq n < 160 \quad (5)$$

$$w[n] = 0.5 + 0.5 \cos(2\pi n/160), 0 \leq n < 160 \quad (6)$$

$S_{m,i}[k]$ 의 각 성분들에 대해 아래의 식 (7)과 같이 배음 강화 필터 $H_{m,i}[k]$ 를 곱한다.

$$S'_{m,i}[k] = H_{m,i}[k] \times S_{m,i}[k], 0 \leq k < N \quad (7)$$

여기서 N 은 DFT 크기로 256이며, 배음 강화 필터 $H_{m,i}[k]$ 는 아래의 식 (8)과 같이 N/p_i 간격으로 최대값을 갖는 주기함수로 정의된다.

$$H_{m,i}[k] = \begin{cases} 1, & 0 \leq k < \lfloor N/p_i \rfloor \\ \frac{-(\eta-1)}{2} \cos\left(\frac{2\pi k}{N/p_i}\right) + \frac{(\eta+1)}{2}, & \lfloor N/p_i \rfloor \leq k < N \end{cases} \quad (8)$$

여기서 $\lfloor \cdot \rfloor$ 는 입력을 넘지 않는 최대의 정수를 출력하는 함수이다. η 는 억제 계수(suppressing coefficient)로서 0과 1 사이의 값을 가지며, 0에 가까울수록 배음 이외의 성분들이 더 억제되기 때문에 배음성분들이 상대적으로 뚜렷해지고 1에 가까울수록 처리된 신호가 원본 신호와 가깝게 되므로 배음성분들이 강화되는 효과가 줄어든다. 작은 η 를 사용하면 끊김 현상은 매우 줄어들지만 주파수 스펙트럼상에서 많은 성분들이 제거되어서 음질이 저하될 수 있다. 많은 곡을 대상으로 시험한 결과 체감 음질의 저하 없이 끊김 현상을 완화하는 η 의 값은 음악에 따라 다르지만 일반적으로 0.5에서 0.9 사이라는 것을 확인할 수 있었다.

주파수 영역에서 필터링된 신호 $S'_{m,i}[k]$ 를 IDFT하여 시간 영역 신호 $s'_{m,i}[n]$ 을 얻고, 이를 아래 식 (9)와 같이 오버랩-에드(overlap-add)하여 최종 전처리 출력 $\mathfrak{z}[n]$ 을 얻는다.

$$\mathfrak{z}[n] = \begin{cases} s'_{m-1,i}[n+80] + s'_{m,i}[n], & 0 \leq n < 80 \\ s'_{m,i}[n+80] + s'_{m+1,i}[n], & 80 \leq n < 160 \end{cases} \quad (9)$$

이런 방법으로 주파수 영역에서 N/p_i 간격의 배음 성분들을 증가시키면 시간 영역에서 p_i 간격의 성분들이 증가하여 예측 오차의 자기상관의 최대값을 증가시킬 수 있으며, 그에 따라 장기 예측 이득 β 가 증가하여 끊김 현상을 줄일 수 있다. 이상의 과정은 배음 성분만을 더 강화시키는 과정이며 포먼트(Formant)에는 변화를 주지 않으므로 체감 음질상의 저하는 거의 없다.

IV. 실험 결과 및 검토

이 장에서는 제안한 전처리 알고리즘의 성능을 분석하기 위한 실험 과정을 설명하고, 실험 결과를 검토하기로 한다.

실험에 사용된 곡은 모두 10 곡이며, 현재 이동통신의 음악 서비스에 사용되는 곡 가운데 다양한 특성을 갖는 곡을 선정하였다. 원 음악은 모두 CD 품질(44.1 kHz, 16bit, 스테레오)을 가지고 있는데, 이를 EVRC가 입력으로 사용하는 포맷인 8kHz, 16bit 모노로 변환하였다. EVRC에 입력하는 음원의 품질을 실제 전화상황과 같게 하기 위하여 다운 샘플링 전에 사용하는 엘리머싱 방지 필터는 ITU-T 권고안^[8]에 정의된 대로 통과 대역이 200

~ 3,400 Hz가 되도록 하였다. 전처리에 사용한 억제계수는 100 여곡을 대상으로 실험을 하여 장르별로 최적값을 설정하였다.

성능 평가 방법으로는 원곡과 전처리된 음악을 ARS 서버에 저장한 다음에 CDMA 단말기로 ARS에 전화를 건 피험자들에게 들려주고 1 ~ 5 점 사이의 점수를 부여하는 방법인 MOS(mean opinion score) 평가^[6]를 사용하였다. 건강한 20대와 30대 연령의 11 명이 피험자들이 평가에 참여하였다. 단말기 특성에 따라 평가 결과가 달라질 수 있으므로 피험자들은 모두 동일 단말기를 사용하도록 하였다. 재생되는 곡의 순서는 원곡과 전처리된 곡을 짝으로 하였으며, 각 곡마다 원곡과 전처리된 곡의 순서를 임의로 변경하여 피험자들은 원곡과 전처리된 곡을 모르는 상태에서 평가하였다.

MOS 테스트는 많은 수의 피험자들을 대상으로 음성 코덱에서 부/복호된 음성 신호에 대해 주관적 음성 명료도를 평가하게 하는 방법으로 ITU-T 권고안 P.800^[9]에서 제안되었으며, 음성 코덱의 성능을 평가하는 정량적인 방법으로 널리 사용되어 왔다. 피험자들은 음성의 명료도에 대해 5점 척도로 절대 평가를 하게 되는데, 이 5점 척도의 평가 기준은 표 1과 같다.

표 1. MOS 기준 점수
Table 1. MOS scores

점수	1	2	3	4	5
품질	bad	poor	fair	good	excellent

이 논문에서는 음악의 품질을 평가하는 용도로 MOS 테스트를 사용하므로, 원래 ITU-T가 제안한 MOS 테스트의 과정을 그대로 이용할 수는 없다. 왜냐하면, 5점 척도의 절대 평가를 하더라도 개개인마다 품질의 최적 정도를 판단하는 기준이 다를 수 있기 때문이다. 단적으로, CDMA 단말기의 음질을 평가할 때 CD 음질을 그 판단 기준으로 한다면 곤란할 것이다. 평가 실험을 설계할 때 이러한 점을 고려하여 피험자들이 유선 전화상에서의 음질을 최적 품질이라고 생각하였을 때를 기준으로 5점 척도 평가를 할 수 있도록 하였다.

표 2는 MOS 평가에 사용된 10 곡에 대한 원곡과 전처리된 곡의 MOS를 비교한 것이다. 등분산 가정 T 검정을 수행한 결과, 대부분 p-value가 0.05 미만이므로 95%의 신뢰도로 처리곡이 원곡에 비해 품질이 개선되었다고 말할 수 있다.

표 2의 결과를 보면, 가곡이나 발라드와 같이 뚜렷한 주 피치가 존재하는 음악에서는 피험자들이 원곡과 전처리된 곡의 품질 차이를 크게 느끼지 못하는 반면에, 록 음악이나 교향곡과 같이 다양한 악기가 동시에 연주되는 곡에서는 원곡의 점수가 대단히 나쁘고 전처리된 곡의 점수가 상대적으로 높은 것을 알 수 있다. 특히, 1 점대의 곡들은 곡 중간에 끊김 현상이 많은데 이런 현상이 전처리된 곡에서는 상당 부분 해소된 것이 이런 평가 결과의 원인으로 분석된다.

표 2. MOS 평가 결과
Table 2. MOS evaluation results

번호	장르	원곡 평균	처리곡 평균	p-value
1	Hard rock	1.727	3.000	0.00053
2	Hard rock	1.000	2.545	0.00000
3	Hard rock	1.200	2.091	0.00025
4	Hard rock	1.455	2.455	0.00136
5	Rock	1.545	3.000	0.00159
6	피아노솔로	3.000	3.273	0.16463
7	Ballad	2.727	2.727	0.50000
8	피아노솔로	1.727	2.455	0.02321
9	가곡	2.545	2.273	0.19837
10	교향곡	2.091	2.727	0.06105

그림 6은 표 2의 2번곡에 대하여 제안된 전처리를 수행하였을 때, 전처리에 의한 전송률 결정 임계치와 대역 에너지의 변화를 나타낸 것이다.

그림 6을 살펴보면, 전처리된 곡 (b)에서는 음폭 압축의 결과로 대역에너지가 높은 값을 유지하며, RDT가 상대적으로 천천히 증가할 뿐만 아니라 (a)의 경우에 비해서 작은 값으로 유지되는 것을 볼 수 있다. 이런 이유 때문에 (a)에서는 대략 처음 20 초 이후로 RDT가 대역 에너지보다 높은 값을 유지하여 계속 끊김 현상이 나타나지만, (b)에서는 RDT가 대역 에너지보다 낮은 값을 유지하여 끊김 현상이 거의 나타나지 않는 것을 확인할 수 있다. 즉, 원곡에서 Rate 1/8로 부호화되는 프레임들도 전처리된 곡에서는 대부분 Rate 1로 부호화될 수 있다.

그림 7은 표 2의 두 번째 곡에 대해서, 원곡과 전처리된 곡의 파형, 그리고 각각을 EVRC 코덱으로 압축, 재생한 후의 파형을 비교한 것이다. 그림 7에서 볼 수 있듯이 원곡과 전처리된 곡에는 큰 차이가 없는 반면에 EVRC 코덱을 통과하고 난 후의 파형에는 매우 큰 차이가 있는 것을 알 수 있다. 특

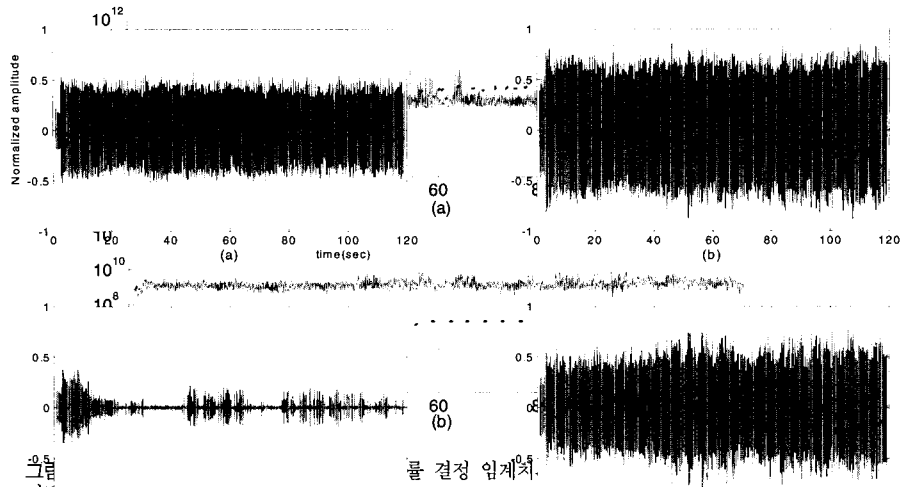


Fig. 6. Plots of the band energy and RDT before and after the preprocessing: (a) the original signal (b) the preprocessed signal (c) the original signal coded by an EVRC codec (d) the processed music signal coded by an EVRC codec

그림 7. 전처리 결과 : (a) 원본 음악의 파형 (b) 전처리된 음악의 파형 (c) 원본 음악의 EVRC 부/복호 출력 파형 (d) 전처리된 음악의 EVRC 부/복호 출력 파형

Fig. 7. The result of preprocessing : (a) the waveform of the original music signal (b) the waveform of the preprocessed music signal (c) the waveform of the original music signal coded by an EVRC codec (d) the waveform of the processed music signal coded by an EVRC codec

히, 원곡이 EVRC 코덱을 통과한 후의 파형인 (c)에서는 끊긴 부분이 대단히 많은 것을 볼 수 있다. 표 3은 실험에 사용된 10 개의 원곡과 전처리된 곡에서 'Rate 1/8'로 부호화된 프레임의 개수를 비교한 것이다. 이 결과는 소프트웨어 EVRC 코덱을 사용하여 시뮬레이션을 한 결과이다. 표 3에서 볼 수 있듯이 전처리 과정을 거치면 'Rate 1/8'로 부호화된 프레임의 개수가 줄어든다.

표 3. 전처리 여부에 따른 "Rate 1/8"로 부호화된 프레임 개수
Table 3. Number of the frames encoded at "Rate 1/8" between the original and the preprocessed music signals

번호	원곡(a)	처리곡(b)	개선률((a-b)/a) (%)
1	43	0	100.0
2	1265	0	100.0
3	34	14	58.8
4	9	0	100.0
5	33	0	100.0
6	3	0	100.0
7	25	13	48.0
8	144	13	91.0
9	43	10	76.7
10	173	28	83.8

이상의 결과를 종합해 보면, 이 논문에서 제안하는 전처리 방법은 EVRC 코덱에서 끊김 현상을 효과적으로 방지하여 EVRC 복호기에서 재생되는 음악의 체감 품질을 효과적으로 높이는 것을 알 수 있다.

V. 결론

이 논문은 CDMA 단말기에서 음악을 청취할 때 EVRC의 구조적인 한계로 인해서 끊김 현상이 발생하는 것을 방지하는 것을 목표로 하였으며, 나아가서 CDMA 단말기에서 듣는 음악 품질의 최적화를 추구하였다. 이를 위해서 먼저 음폭 압축을 수행하였으며 이를 통해 큰 소리의 포화 현상과 작은 소리의 끊김 현상을 방지할 수 있었다. 또한, 배음 성분들을 강화하는 방법을 사용하여 재생되는 음악의 끊김 현상을 획기적으로 개선할 수 있었다.

참고 문헌

[1] Andreas Spanias, "Speech coding: a tutorial review," *Proceedings of the IEEE*, vol. 82, no. 10, pp 1541-1582, October 1994.
[2] M. Zeng, A. Annamalai, and V.K. Bhargava,

- "Recent advances in cellular wireless communications," *IEEE Communications Magazine*, pp.128-138, September 1999.
- [3] Sean A. Ramprasad, "A Two Stage Hybrid Embedded Speech/Audio Coding Structure", *Proc. of International Conference on Acoustics, Speech, and Signal Processing(ICA-SSP '98)*, May 1998, vol. 1, pp. 337-340.
- [4] 3GPP2 C.S0030-0 Version 2.0, *Selectable Mode Vocoder Service Option for Wideband Spread Spectrum Communication Systems*, 3rd Generation Partnership Project 2, December 2001.
- [5] TR45, *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, IS-127 July 19, 1996.
- [6] Frerking, *Digital Signal Processing in Communication Systems*, Van Nostrand Reinhold, 1994.
- [7] A.M.Kondoz, *Digital Speech - Coding for Low Bit Rate Communication Systems*, John Wiley & Sons, 1994.
- [8] ITU-T Recommendation P.830, *Subjective Performance Assessment of Telephone-Band And Wideband Digital Codecs*, International Telecommunication Union, Geneva, Switzerland, February 1996.
- [9] ITU-T Recommendation P.800, *Methods for Subjective Determination of Transmission Quality*, International Telecommunication Union, Geneva, Switzerland, August 1996.

남 영 한(Young-Han Nam) 정회원

1998년 2월 : 서울대학교 전기공학부 졸업
 2002년 2월 : 서울대학교 대학원 협동과정 의용생체공학 전공 석사
 2002년 3월~현재 : 와이더넛닷컴(주) 기술연구소 전임연구원

<주관심분야> 오디오/생체 신호처리, 인공지능

하 태 균(Tae-Kyoon Ha)

정회원



1990년 2월 : 서울대학교 제어계측공학과 졸업
 1992년 2월 : 서울대학교 제어계측공학과 석사
 1992년 2월~1995년 11월 : 삼성 전자 통신연구소 전임연구원
 1995년 11월~1997년 11월 : 삼성 전자 자동화연구소 전임연구원
 1997년 11월~2000년 4월 : (주)컨벡스 실장
 2000년 5월~2001년 5월 : (주)제스텔 부장
 2001년 5월~현재 : 와이더넛닷컴(주) 기술연구소 수석연구원 재직중

<주관심분야> VoiceXML, CTI

전 윤 호(Yun-Ho Jeon)

정회원



1989년 2월 : 서울대학교 제어계측공학과 졸업
 1991년 2월 : 서울대학교 제어계측공학과 석사
 2001년 2월 : 서울대학교 전기컴퓨터공학부 박사
 1991년 2월~1994년 7월 : 한국전자통신연구원(ETRI)
 1994년 8월~1996년 6월 : 한국과학기술연구원(KIST)
 2000년 5월~2001년 5월 : (주) 제스텔
 2001년 6월~현재 : 와이더넛닷컴(주)

<주관심분야> 이동통신, 내장형/실시간 시스템, 인공지능, S/W공학

김 재 수(Jae-Soo Kim)

정회원



1984년 2월 : 한양대학교 전자공학과 졸업
 1987년 : 미국 아이오와 주립대 전기/컴퓨터 공학과 석사
 1995년 : 미국 뉴욕 주립대 전기/컴퓨터 공학과 박사

1996년~2001년 : 미국 Lucent Technology Bell Labs 연구원
 2001년~현재 : 와이더넛닷컴(주) CTO

<주관심분야> 무선 인터넷, Ubiquitous Comuting,
NGN system

박 섭 형(Seop-Hyeong Park)

정회원



1984년 2월 : 서울대학교 제어계
측 공학과 졸업

1986년 2월 : 서울대학교 제어계
측 공학과 석사

1990년 8월 : 서울대학교 제어계
측 공학과 박사

1990년 9월~1992년 10월 : 한국생산기술연구원
HDTV개발단 선임연구원

1992년 9월~1998년 2월 : KT 통신망연구소 선임연
구원

1993년 10월~1994년 9월 : 일본 NTT Human
Interface 연구소 객원연구원

1998년 3월~현재: 한림대학교 정보통신공학부 교수

<주관심분야> 신호처리, 영상 및 음성 신호처리, 멀
티미디어 통신, VoiceXML