

# 가변 잡음 레벨을 이용한 음성신호에 대한 SBR 성능 향상 기술

## Enhancement of SBR for Speech Signal Using Adaptive Noise Floor Level

이 세 원\*, 오 승 준\*, 안 창 범\*, 이 태 진\*\*, 강 경 옥\*\*, 박 호 종\*  
(Sewon Lee\*, Seoung-Jun Oh\*, Chang-Beom Ahn\*, Taejin Lee\*\*, Kyoung-Ok Kang\*\*, Hochong Park\*)

\*광운대학교 전자공학과, \*\*한국전자통신연구원 방통미디어연구부  
(접수일자: 2008년 11월 12일; 수정일자: 2008년 12월 29일; 채택일자: 2008년 12월 30일)

오디오 부호화 기술에서 SBR은 고대역의 시간-주파수 정보를 저대역으로부터 구하고 보정 파라미터를 이용하여 고대역 정보를 보정하여 고대역 신호를 합성하는 기술이다. SBR은 고대역 정보의 부호화를 위하여 보정 파라미터만 전달하므로 매우 적은 비트로 오디오 신호를 압축할 수 있도록 하며, MPEG-4 HE-AAC의 핵심 모듈로 사용되고 있다. SBR은 원래 오디오 신호를 기반으로 개발되었기 때문에 음성 입력에 대하여 성능이 저하되는 문제점을 가지며, 성능 저하의 대표적인 이유는 톤 성질이 부정확하게 계산되어 잡음 레벨이 높게 설정되고 복원된 고대역 정보에 과도한 잡음이 포함되기 때문이다. 본 논문에서는 음성 신호에 대한 SBR 성능 저하 문제를 해결하기 위하여 잡음 레벨을 입력 음성 신호의 특성에 맞게 가변적으로 적용하는 기술을 제안한다. 제안하는 SBR은 기존의 SBR과 호환성을 유지하며, 주관적 평가를 통하여 기존 SBR에 비하여 남성 음성에 대한 성능이 향상된 것을 확인하였다.

**핵심용어:** SBR, HE AAC, 잡음 레벨, 피치 주파수, 선형 예측

**투고분야:** 음성처리 분야 (2)

In audio coding, SBR technology synthesizes the high-bands using patched time-frequency information from low-bands and the correction parameters. Since SBR transmits only correction parameters for high-bands, it provides a low-rate coding of high-bands, and is used as a core module of MPEG-4 HE AAC. SBR was originally designed for audio signal and its performance for speech signal tends to decrease, and the major reason is an excessive noise floor in high-bands which is caused by incorrect tonality computation. In this paper, a new method to determine noise floor level in an adaptive fashion according to the speech characteristics is proposed in order to solve the problem of SBR for speech signal. The proposed method maintains the compatibility with the standard SBR, and the subjective performance evaluation shows that the proposed method improves the SBR performance especially for male speech signal compared with the standard SBR.

**Keywords:** SBR, HE AAC, Noise floor level, Pitch frequency, Linear prediction

**ASK subject classification:** Speech Signal Processing (2)

## I. 서론

음성과 오디오 부호화 기술은 디지털 통신 및 방송의 품질과 용량을 결정하는 매우 중요한 기술로서 1980년대 부터 매우 광범위하게 연구되고 있고, 응용 분야에 따라 많은 표준 음성 부호화기와 오디오 부호화기가 개발되어

방송 및 통신 서비스에 사용되고 있다. 최근, 기존의 음성 중심의 이동통신에서 탈피하여 이동통신에서도 멀티 미디어 정보의 전송이 요구되고, 디지털 방송/통신 융합 시대를 맞이하여 더 이상 음성과 오디오 신호를 별도의 콘텐츠로 분리하는 것이 어렵게 되었다. 이와 같은 서비스의 융합에 따라 하나의 부호화기로 다양한 음성과 오디오 정보를 동시에 처리하려는 통합화 노력이 시작되었다. 그러나 기존의 기술은 음성 또는 오디오 신호의 특징을 기반으로 서로 다른 목표에 따라 설계되어 하나의 방

책임저자: 박 호 종 (hcpark@kw.ac.kr)  
서울시 노원구 원계동 447-1 광운대학교 전자공학과  
(전화: 02-940-5104; 팩스: 02-913-9057)

법으로 음성과 오디오를 통합적으로 부호화하는데 한계를 가진다.

현재 최고 성능의 음성 부호화기와 오디오 부호화기(저대역과 고대역 신호를 각 신호 특성을 고려하여 서로 다른 방식으로 처리하는 구조를 가진다 [1-5]). 저대역 신호는 핵심 영역으로서 신호의 특성에 따라 CELP 또는 변환 부호화 기술로 부호화 하고, 고대역 신호는 저대역 정보를 활용하고 부가적인 보정 과정을 거치는 기법으로 부호화 하며, 대표적으로 MPEG-4 HE-AAC에서 사용하는 SBR (spectral band replication) 기술이 있다 [6]. SBR 기술을 사용하면 고대역 신호를 매우 적은 비트로 부호화 할 수 있으므로 전체 부호화기의 성능이 크게 향상된다. 예로, HE-AAC와 enhanced aacPlus는 AAC-LC로 저대역을 처리하고 SBR로 고대역을 처리하여 64kbps 이하에서 매우 우수한 성능을 제공한다.

SBR은 원래 오디오 신호를 기반으로 개발되었으며, 음성 신호에 기존 SBR을 그대로 적용하면 오디오에 비하여 성능이 크게 저하되는 경우가 나타난다. 즉, 현재의 SBR은 입력 신호의 특성에 따라 성능이 변하는 문제를 가지며, 이에 따라 음성/오디오 통합 부호화에 기존의 SBR을 그대로 사용하는 것은 불가능하다. 따라서 음성 신호에 대한 SBR의 성능을 향상시키는 새로운 기술 개발이 필요하며, 이 기술은 음성/오디오 통합 부호화기에서 고대역을 처리하는 모듈에 적용되어 모든 입력 신호에 대하여 우수한 성능을 가지도록 할 것이다.

본 논문에서는 HE-AAC의 SBR을 기준으로 음성 신호에 대한 SBR 성능을 향상시키는 기술을 제안한다. 먼저, SBR 동작을 분석하고 음성 신호에 대한 성능이 저하되는 원인을 분석하여 잡음 레벨을 추정하는 과정에서 문제가 있는 것을 확인하고, 이를 해결하기 위하여 음성 특성에 따라 가변적인 잡음 레벨을 사용하는 방법을 제안한다. 제안하는 기술은 기존 SBR과 호환성을 유지하므로 기존 SBR에 간단히 적용되어 성능 향상을 제공할 수 있는 장점을 가진다.

본 논문의 구성은 다음과 같다. 2장에서 기존 SBR 동작을 기반으로 음성에 대한 성능 저하 원인을 분석하고, 3장에서는 성능 향상을 위한 새로운 SBR 방법을 제안한다. 마지막으로 4장에서는 주관적 평가를 통하여 기존 SBR과 제안한 SBR의 성능을 비교 평가한다.

## II. SBR 문제점 분석

### 2.1. 개요

기존 SBR의 동작은 다음과 같다 [3-4]. 입력 신호를 64 채널 QMF (quadrature mirror filter)를 사용하여 32~64 시간-주파수 영역으로 분해하고, 저대역 신호로부터 고대역 신호를 생성하는 패치 (patch) 구조에 따라 고대역의 QMF 정보를 결정한다. 다음, 원 고대역 QMF 값과 패치된 QMF 값을 비교하여 패치된 QMF 값이 원래의 QMF 값에 근사하도록 하기 위한 보정 정보를 구한다. SBR이 사용하는 보정 정보는 크게 에너지와 톤 성질 (tonality)로 구분된다. 에너지는 패치된 고대역의 QMF 크기가 원 크기에 맞도록 보정하는 정보이고, 톤 성질은 패치된 고대역의 톤 성질이 원 성질에 맞도록 보정하는 정보이다. 일반적으로 고대역의 톤 성질은 저대역에 비하여 낮으며, 이 경우에는 패치된 고대역의 톤 성질을 약화시키기 위하여 잡음을 추가하거나 역필터 (inverse filtering) 기법을 사용한다.

SBR에서 각 QMF 채널의 톤 성질은 각 채널 신호를 시간 축에서 2차 선형 예측을 하여 예측 오차의 에너지를 이용하여 측정한다. 즉, 예측 오차가 작으면 예측 성능이 우수한 것이고, 이는 톤 성질이 강한 것을 의미한다. 따라서, 예측이 잘 되고 톤 성질이 강한 저대역 채널에 잡음을 추가하여 원 고대역 채널과 동일한 예측 오차를 가지게 하면 원하는 고대역의 톤 성질 보정이 구현 되었다고 할 수 있다. 이에 따라 저대역과 고대역의 각 채널에 대하여 예측 오차를 구하여 톤 성질을 정의하고, 대응하는 두 채널의 톤 성질이 동일하게 되기 위하여 필요한 잡음 레벨을 결정한다.

### 2.2. 문제점 분석

음성 신호에 대하여 SBR 성능이 저하되는 현상을 먼저 실험적으로 확인하기 위하여 다양한 음성 신호에 대한 청취 평가를 진행하였으며, 여성 음성에는 문제가 없지만 남성 음성에 대하여 큰 성능 저하가 발생하는 것을 확인하였다. 그림 1은 남성 음성에 대한 실험 결과를 보여 준다. (a)는 0.5초 길이와 11 kHz 주파수 영역에 대한 원 신호의 스펙트로그램이며, 고대역에서 매우 규칙적인 주기성을 볼 수 있다. (b)는 기존 SBR로 복원된 신호의 스펙트로그램이며, 고대역에 많은 잡음이 추가되어 원 신호에 비하여 주기성이 크게 감소한 것을 볼 수 있으며, 청취 성능이 크게 저하되는 것을 확인하였다. (c)는 SBR 복호



그림 1. 남성 음성에 대한 스펙트로그램 (x-축:시간, y-축:주파수) (a) 원 신호 (b) SBR 출력 신호 (c) SBR 복호화기에서 잡음을 제거한 출력 신호  
 Fig. 1. Spectrogram of male speech (x-axis:time, y-axis:frequency). (a) Original. (b) Standard SBR output. (c) SBR output without noise floor.

화기에서 잡음 레벨을 인위적으로 0으로 하여 추가적인 잡음 추가를 하지 않는 경우의 스펙트로그램이다. (c)를 (b)와 비교하면 고대역 정보가 원음에 더 근접하고 있음을 확인할 수 있고, 잡음을 인위적으로 추가하지 않으면 실제 청취 품질도 크게 개선되는 것도 실험적으로 검증하였다.

다양한 음성 신호에 대한 SBR의 종합적 청취 평가에 의하여 여성 음성보다 남성 음성에 대한 품질이 크게 저하되는 것을 알 수 있고, 이와 같이 남성 신호에 대한 품질 저하의 원인 중의 하나가 과도한 잡음 레벨이라고 결론 내릴 수 있다. 여성과 남성 음성의 가장 큰 특성 차이는 음성의 피치 주파수이다. 따라서 피치 주파수에 따라 잡음 레벨 계산의 정확도에 차이가 있고 남성의 피치 주파수에 대하여 오류가 발생하는 것으로 예상되며, 이와 같은 예상을 이론적으로 검증하면 다음과 같다.

SBR에서 전체 대역이 64개의 QMF 채널로 분해되므로 각 QMF 채널의 주파수 대역폭은 (샘플링 주파수)/2/64이며, 실험에서 사용하는 샘플링 주파수 48 kHz에 대하여 QMF 채널의 대역폭은 375 Hz이다. 만일 100 Hz 기본 주파수를 가지는 신호가 입력되면 100 Hz의 모든 정수배에서 하모닉 피크를 가지게 되며, 그 결과 하나의 QMF 채널에 여러 개의 하모닉 피크 성분이 존재하게 된다. 반면, 만일 기본 주파수가 260 Hz인 경우 260, 520, 780 Hz 등에서 하모닉 피크를 가지므로 이때는 각 QMF 채널에 한 개의 하모닉 피크만 존재한다.

SBR에서 사용하는 예측기의 차수가 2로 매우 낮고 각 하모닉 피크가 실제로는 완벽한 하모닉 정수배에 위치하지 않으므로 각 채널에 여러 개의 하모닉 피크가 존재하면 하나의 피크만 존재하는 경우에 비하여 예측 성능이 저하되고 이에 따라 톤 성질 값이 실제보다 작게 측정된

다. 이와 같은 이유에 따라 피치 주파수가 낮으면 예측 동작에 문제가 발생하고 예측 오차가 정상보다 크게 계산되고 톤 성질이 정상보다 낮게 측정되는 문제점을 가진다.

실제 QMF 신호에서 하모닉 피크가 발생하는 현상을 QMF 출력 신호의 스펙트럼을 통하여 확인하였다. 그림 2는 특정 피치 주파수에 대하여 QMF 채널의 하모닉 성분 분포에 차이가 발생하고 이에 따른 톤 성질에 차이가 발생하는 예를 보여준다. 여기서 48 kHz 샘플링 주파수를 사용하였고 채널의 대역폭은 375 Hz이다. 먼저, (a)는 QMF의 두 번째 채널 (375 Hz-750 Hz)와 세 번째 채널 (750 Hz-1125 Hz)에 대하여 피치 주파수 105 Hz 및 225 Hz 신호의 이론적 하모닉 주파수 위치를 보여준다. (b)는 피치 주파수 225Hz를 가지는 신호에 대하여 두 번째 채널 신호의 스펙트럼을 보여준다. 두 번째 채널에는 450 Hz와 675 Hz를 가지는 두 하모닉 성분이 포함되며, 675 Hz 성분이 첫 번째 피크, 450 Hz 성분이 두 번째 피크로 나타난다. 여기서, QMF 과정의 64배 down-sampling에 의한 주파수 folding이 발생하여 순서가 변경된다. 가운데의 매우 작은 두 개의 피크는 각각 900 Hz와 225 Hz에 해당하는데, 이들은 채널 밖의 하모닉 성분이지만 QMF의 aliasing에 의하여 일부 에너지가 두 번째 채널에 포함되어 나타난 것이다. 이와 같이 2개의 뚜렷한 하모닉 피크가 나타나는 경우에는 2차 예측기로 정확하게 예측되며, 실제 톤 성질은 약 40으로 측정된다.

(c)는 225 Hz를 가지는 신호에 대하여 세 번째 채널 신호의 스펙트럼을 보여준다. (b)와 비교하면 하모닉 피크 수는 감소하였지만 aliasing 성분에 해당하는 675 Hz의 크기가 증가하고 그에 따라 톤 성질이 약 1.0으로 매우 작게 측정된다. 이는 채널 밖에 존재하는 675 Hz 성분이 채널 경계에 매우 근접하여 aliasing 효과가 크게 발생하기 때문이다.

(c)은 피치 주파수 105 Hz를 가지는 신호에 대하여 두 번째 채널 신호의 스펙트럼을 보여준다. 해당 채널에 4개의 정상적인 하모닉 피크 (420 Hz, 525 Hz, 630 Hz, 735 Hz)와 2개의 aliasing에 의한 하모닉 피크가 포함되며, 이에 따라 2차 예측기가 정상적인 예측 동작을 수행하지 못하고, 이때의 톤 성질은 약 1.0 정도로 매우 작게 잘못 측정된다.

그림 3은 기본 피치 주파수 80 Hz부터 400 Hz에 대하여 첫 3개의 QMF 채널의 톤 성질 값을 기존의 SBR 기술에 따라 측정한 결과이다. 모든 피치 주파수에 대하여 동일한 톤 성질을 가지므로 피치 주파수에 관계없이 이론적으로 동일한 톤 성질 값이 계산되어야 한다. 그러나 SBR

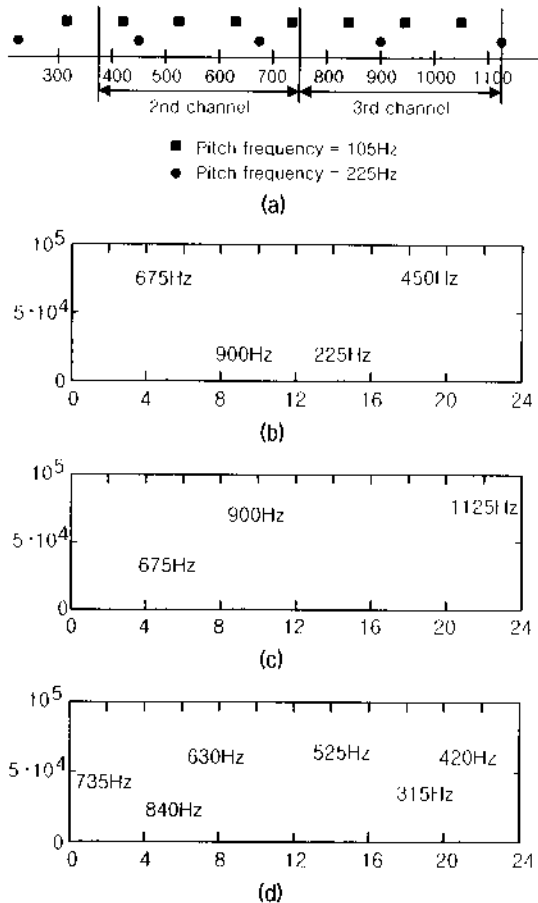


그림 2. QMF 채널 신호의 하모닉 성분 분포. (a) 하모닉 피크의 이론적 위치. (b) 피치 주파수 225 Hz에 대한 두 번째 채널의 스펙트럼. (c) 피치 주파수 225 Hz에 대한 세 번째 채널의 스펙트럼. (d) 피치 주파수 105 Hz에 대한 2 번째 채널의 스펙트럼

Fig. 2. Harmonic-peak distribution of QMF channel signal. (a) Ideal location of harmonic peaks. (b) Spectrum of 2nd channel for pitch frequency of 225 Hz. (c) Spectrum of 3rd channel for pitch frequency of 225 Hz. (d) Spectrum of 2nd channel for pitch frequency of 105 Hz.

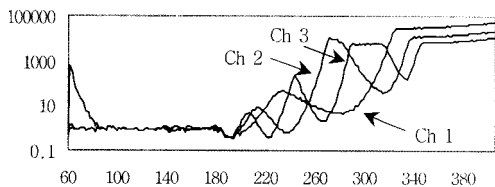


그림 3. 피치 주파수 (x-축)에 따른 QMF 채널별 톤 성질 값 (y-축)

Fig. 3. Pitch frequency (x-axis) vs. tonality (y-axis) for each QMF channel.

에서 측정된 톤 성질 값이 피치 주파수가 감소함에 따라 크게 감소하는 것을 알 수 있다. 결론적으로, 기존 SBR에서 피치 주파수에 따라 톤 성질이 다르게 측정되고, 피치 주파수가 감소하면 톤 성질이 정상보다 낮게 측정되는

문제점을 가진다. 구체적으로 200 Hz 이하에서는 매우 작은 일정값을 가지고, 200 Hz 이상에서 점차 증가하여 정상적인 톤 성질 값을 가진다. 200 Hz 이상에서 톤 성질이 국부적으로 변하는 것은 각 채널 내에 존재하는 aliasing 성분의 크기와 위치에 따라 예측 성능이 변하기 때문이다.

그림 2와 3에서 볼 수 있듯이 피치 주파수와 채널 대역에 대한 하모닉 성분의 상대적 위치에 따라 톤 성질이 다르게 측정된다. 이 중에서, 대역 사이의 aliasing 문제는 근본적으로 해결할 수 없으며, 이 현상은 피치 주파수가 크면 모든 채널에서 발생하는 것이 아니라 채널 대역과 하모닉 주파수와의 상대적 위치에 따라 일부 채널에서만 발생한다. 특히 SBR 동작에서는 톤 성질을 측정할 때 채널 해상도를 줄여 복수의 채널에 대한 평균값을 구하는 경우도 있으며, 이 경우는 작게 측정된 톤 성질 값이 약간 보정될 수도 있다.

그러나 피치 주파수가 작을 경우, 한 채널에 포함되는 정상적인 하모닉 피크 수가 증가하고 이 문제는 모든 채널에 공통적으로 나타나게 되므로 채널 해상도 감소에 의하여 평균값을 계산하여도 그 영향이 감소하지 않는다. 또한, 피치 주파수가 작으면 채널 aliasing 영향이 발생할 확률이 증가하므로 피치 주파수가 작을 때는 잡음 레벨의 문제점이 더 크게 나타난다.

이와 같이 잡음 레벨은 피치 주파수에 따라 정확도가 변경되며, 샘플링 주파수에 따라 각 주파수 채널의 대역 폭과 경계를 계산하고 입력 신호의 피치 주파수와의 상관 관계를 분석하여 각 채널의 잡음 레벨을 수정하여 주는 과정이 필요하다.

### III. 가변 잡음 레벨 결정

2장에서 실험과 이론적 분석에 의하면 피치 주파수가 낮은 음성 신호에 대하여 각 QMF 채널에 복수의 하모닉 피크가 존재하고, 이에 따라 강한 톤 성질을 가지더라도 2차 예측 성능이 저하되어 실제 톤 성질에 비하여 톤 성질 값이 낮게 측정되는 오류가 발생한다. 톤 성질이 낮게 측정된 경우, 이 문제가 최종 잡음 레벨에 미치는 영향은 다음과 같다. 고대역의 원 톤 성질을  $T_{orig}$ , SBR에 의하여 패치된 고대역 톤 성질을  $T_{sh}$  이라 할 때, 잡음 레벨은  $\frac{1}{T_{orig}}$  또는  $G \frac{T_{sh}}{T_{orig}^2}$ 로 계산된다 [3-4]. 만일 톤 성질 측정이 잘못되어 실제 값에 비하여 측정된 값이  $\frac{1}{Q}$  배가 되면 최종 잡음 레벨은  $\frac{Q}{T_{orig}}$  또는  $G \frac{T_{sh}/Q}{(T_{orig}/Q)^2} = QG \frac{T_{sh}}{T_{orig}^2}$

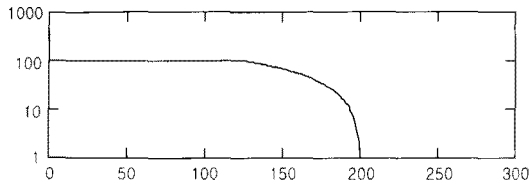


그림 4. 피치 주파수 (x-축)에 따라 적용하는 이득값 (y-축)의 예  
Fig. 4. Example of gain (y-axis) for pitch frequency (y-axis).

가 되어 실제 잡음 레벨에 비하여  $Q$  배 증가한다. 따라서, 낮은 피치 주파수를 가지는 신호에 대하여 톤 성질 측정이 실제보다 작게 되고 그에 따라 잡음 레벨이 증가하여 과도한 잡음이 고대역에 추가되어 SBR 성능이 저하된다.

이상에서 분석한 결과를 토대로 피치 주파수에 따라 톤 성질 값과 잡음 레벨을 보정하여 피치에 가변적인 잡음 레벨을 결정하는 방법을 제안한다. 먼저, 입력 신호의 피치 주파수를 측정하고, 측정된 피치 주파수 값으로부터 잡음 레벨을 보정할 이득값을 계산하여 식 (1)에 따라 새로운 잡음 레벨을 구한다.

$$\text{새로운 잡음 레벨} = \frac{\text{SBR에서 설정한 잡음레벨}}{\text{이득값}} \quad (1)$$

이를 통하여 잘못된 톤 성질의 영향을 최종 잡음 레벨 결정 과정에서 보정하여 SBR의 성능을 향상시킨다.

본 논문에서 성능 측정에 사용한 이득값은 그림 4와 같으며, 피치 주파수만의 함수로 결정되는 값으로서 실험을 통하여 톤 성질 값의 오류 정도를 통계적으로 측정하여 매우 간단하게 구한 것이다. 물론, 피치 주파수에 따라 QMF 채널내의 하모닉 피크 위치와 aliasing 현상 등을 이론적으로 정확히 분석하여 보다 복잡한 이득값을 구하는 것이 가능하다. 그러나 4장에서 볼 수 있듯이 그림 4와 같은 매우 간단한 이득값을 적용하여도 큰 성능 향상을 얻을 수 있으므로 본 논문에서는 그림 4의 이득값을 사용하기로 한다.

#### IV. 성능 분석

본 논문에서는 출력 스펙트로그램과 주관적 청취 평가를 통하여 제안한 방법을 적용한 SBR의 성능을 기존 SBR 성능과 비교하여 평가한다. 평가에는 Korean/English/German Male, English Female, Classic, Jazz 신호를 사용하였고, 각 신호는 10~15초 내외의 길이로 가진다. 부



그림 5. 제안한 기술을 적용한 남성 음성 신호의 SBR 출력 스펙트로그램 (x-축:시간, y-축:주파수)

Fig. 5. Spectrogram of male speech by SBR with the proposed method (x-axis:time, y-axis:frequency).

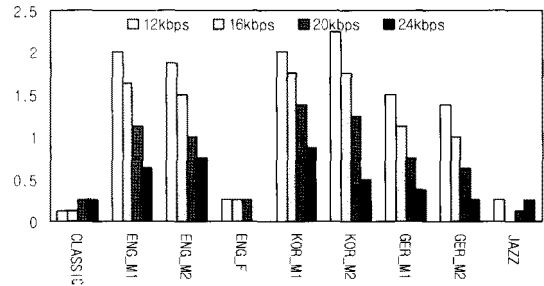


그림 6. 제안한 기술에 대한 주관적 청취 성능 평가 결과  
Fig. 6. Results of subjective performance evaluation of proposed method.

호화기는 3GPP에서 제공하는 enhanced AAC Plus Encoder/Decoder를 사용하고 [7], 모노 12 kbps, 16 kbps, 20 kbps, 24 kbps 모드의 성능을 평가하였다. 평가에 참여한 평가자는 20~30대 남성과 여성을 포함하여 총 8명이고, 조용한 사무실 환경에서 헤드폰을 통하여 청취 하였다.

그림 5는 그림 1에서 사용한 남성 음성 신호에 대하여 제안한 방법을 적용한 출력 스펙트럼을 보여준다. 그림 1(b)에서 보듯이 기존의 SBR을 적용하였을 경우에는 고대역의 잡음이 원 음성 신호보다 많아 성능이 저하되지만, 제안한 방법을 적용하였을 경우에는 잡음 레벨이 보정되면서 보다 원 신호에 가까운 고대역 특성을 갖게 되고, 청취 평가에서 그림 1(b)보다 성능이 크게 향상된 것을 확인할 수 있다.

그림 6은 제안한 기술을 이용한 SBR을 사용하여 복원된 신호의 품질을 -3 ~ +3 scale의 주관적 비교 평가 (comparison scale)로 측정된 결과이다. 기존의 SBR을 적용한 신호와 비교하여 제안한 기술에 의한 신호의 상대적 성능을 측정하며, -3은 'Much Worse', -2는 'Worse', -1은 'Slightly Worse', 0은 'About the Same', 1은 'Slight Better', 2는 'Better', 3은 'Much Better'를 의미한다. 청취 평가에 참가한 모든 평가자가 남성 음성에 대하여 bit rate에 관계없이 제안한 기술을 적용한 SBR의 성능이 우수하다고 평가하였으며, 특히 bit rate가 낮을수록 제안한 기술을 적용한 음성의 음질이 기존의 것보다 월등히 좋다는 것을 알 수 있다.

남성 음성을 제외한 다른 입력에 대해서는 음질 차이가

거의 없는데, 여성 음성 신호는 모든 프레임에서 피치 주파수가 200 Hz 이상이므로 이득값이 "1.0" 로 결정되어 본 논문에서 제안한 방법에 의한 차이점이 발생하지 않으며, Classic와 Jazz는 피치 주파수가 200Hz인 구간이 매우 적게 발생하고 그에 따라 일부 프레임에서만 잡음 레벨의 보정이 이루어지기 때문이다. 따라서 제안하는 기술을 사용하여 기존 SBR의 성능이 크게 저하되었던 콘텐츠에 대하여 성능이 크게 향상되고, 나머지 콘텐츠의 성능에는 차이가 발생하지 않는 것을 확인할 수 있다.

이상의 주관적 성능 평가 결과에 따라 본 논문에서 제안하는 SBR 기술을 사용하면 기존 SBR이 가지고 있는 잘못된 잡음 레벨을 결정하는 문제를 해결하고 최종 복원된 신호의 음질이 크게 향상되는 것을 알 수 있다. 특히 제안하는 기술은 기존 SBR과 호환성을 유지하므로 기존 SBR을 사용하는 모든 부호화기에 쉽게 적용되어 향상된 성능을 제공할 수 있는 장점을 가진다.

## V. 결론

본 논문에서는 음성 신호에 대한 SBR의 성능 향상을 위하여, 톤 성질 보정 과정에서 사용되는 잡음 레벨을 입력 신호의 피치 주파수에 가변적으로 결정하는 방법을 제안하였다. 샘플링 주파수가 높고 남성 음성과 같이 피치 주파수가 낮은 입력 신호가 SBR에 입력 될 경우, QMF 각 채널에 다수의 하모닉 성분이 존재하게 되고, 이에 따라 각 채널 신호의 예측 성능이 저하되어 실제보다 분성질 값이 작게 측정되어 잡음 레벨이 크게 계산되는 오류가 발생한다. 이를 해결하기 위해 입력 신호의 피치 주파수를 측정하고 피치 주파수 값에 따라 톤 성질 값의 오류 정도를 추정하여 이를 잡음 레벨을 계산하는 과정에 적용하여 고대역에 추가되는 잡음 레벨의 값을 보정하는 방법을 사용한다.

제안한 방법을 적용할 경우, 고대역의 잡음 레벨이 원 신호의 잡음 레벨에 좀 더 근접하는 것을 실험적으로 확인하였고, 다양한 특성의 신호에 대한 주관적 청취 실험을 통하여 제안한 방법을 적용하였을 경우 SBR 품질이 향상된 것을 확인하였다. 또한 제안한 방법은 기존 SBR과 호환성을 가지므로 기존 SBR을 사용하는 부호화기에 적용되어 향상된 음질을 제공할 수 있다.

## 감사의 글

본 연구는 지식경제부 및 정보통신연구진흥원의 IT 신성장동력 핵심기술개발사업 [2008-F-011-01 차세대 DTV 핵심기술 개발]과 2008년 광운대학교 교내연구비 지원으로 이루어졌습니다.

## 참고 문헌

1. ISO/IEC 14496-3, "Coding of Audio-Visual Objects - Part 3 : Audio, Subpart 4 : General Audio Coding - AAC, TwinVQ, BSAC," 2001.
2. 3GPP TS 26,290, "Extended AMR Wideband Codec: Transcoding Function," 2002.
3. ISO/IEC 14496-3, AMD, 1, "Bandwidth Extension," 2003.
4. 3GPP TS 26,404, "Enhanced aacPlus General Audio Codec : Encoder Specification SBR Part," 2004.
5. ITU-T G,729,1, "G,729 based Embedded Variable Bit-Rate Coder : An 32kbit/s Scalable Wideband Coder Bitstream Interoperable with G,729," 2006.
6. M. Dietz, L. Liljeryd, K. Kjolring and O. Kunz, "Spectral band replication, a novel approach in audio coding," *proc. 112th AES Convention, May, 2002.*
7. 3GPP TS 26,410, "Enhanced aacPlus General Audio Codec : Floating-point ANSI C Code," 2006.

## 저자 약력

### •이 세 원 (Sewon Lee)

2000년 2월: 광운대학교 전자공학부 (공학사)  
 2002년 2월: 광운대학교 전자공학과 (공학석사)  
 2004년 2월: 광운대학교 전자공학과 (박사수료)  
 2004년 2월~2006년 7월: 다산 네트워크 중앙연구원  
 2006년 7월~2007년 8월: 비로비전 중앙연구원  
 2007년 9월~ 현재: 광운대학교 전자공학과 박사과정  
 ※주관분야: 음성/오디오 신호처리, 통신 신호처리, 영상 신호처리

### •오 승 준 (Seoung-Jun Oh)

1980년 2월: 서울대학교 전자공학과 (공학사)  
 1982년 2월: 서울대학교 전자공학과 (공학석사)  
 1988년 5월: 미국 Syracuse University 전기/컴퓨터공학과 (공학박사)  
 1982년 3월~1992년 8월: 한국전자통신연구원 멀티미디어연구실 실장  
 1992년 9월~ 현재: 광운대학교 전자공학과 교수  
 ※주관분야: 비디오 데이터 처리, 비디오코덱시스템, 비디오 및 영상압축

### •안 창 범 (Chang-Beom Ahn)

1981년 2월: 서울대학교 전자공학과 (공학사)  
 1983년 8월: 한국과학기술원 전기 및 전자공학과 (M.S.)  
 1986년 8월: 한국과학기술원 전기 및 전자공학과 (Ph.D.)  
 1986년 9월~1991년 4월: University of California, Irvine 연구소교수  
 1991년 4월~1992년 8월: 생산기술연구원 전자정보센터 부교수  
 1992년 9월~ 현재: 광운대학교 전기공학과 교수  
 ※주관분야: 생체 신호처리, 의학영상시스템, 영상 신호처리

•이 태 진 (Taejin Lee)

1996년 2월: 전북대학교 전자공학과 (학사)  
1998년 2월: 전북대학교 대학원 전자공학과 (석사)  
1998년 2월 ~ 2000년 5월: Mobens Co., Ltd. 영상음성기술연구팀 연구원  
2002년 10월 ~ 2003년 9월: Tokyo Denki University 방문연구원  
2000년 5월 ~ 현재: 한국전자통신연구원 방통미디어연구부 음향기술연구팀, 선임연구원  
\*주관심분야: 음성/오디오 부호화, 오디오 신호처리

•강 정 옥 (Kyoung-Ok Kang)

1985년 2월: 부산대학교 물리학과 (학사)  
1988년 2월: 부산대학교 대학원 물리학과(석사)  
2004년 2월: 한국항공대학교 대학원 항공전자공학과 (박사)  
1991년 2월 ~ 현재: 한국전자통신연구원 방통미디어연구부 음향기술연구팀장, 책임연구원  
\*주관심분야: 오디오 신호처리, MPEG-7 및 TV Anytime 기반 맞춤형 방송

•박 호 중 (Hochong Park)

1986년 2월: 서울대학교 전자공학과 (공학사)  
1987년 12월: Univ. of Wisconsin-Madison, 전자공학과 (M.S.)  
1993년 5월: Univ. of Wisconsin-Madison, 전자공학과 (Ph.D.)  
1993년 9월 ~ 1997년 8월: 삼성전자 선임연구원  
1997년 9월 ~ 현재: 광운대학교 전자공학과 교수  
\*주관심분야: 음성/오디오 신호처리, 통신 신호처리, 영상 신호처리