

선 스펙트럼 주파수의 청각 적응 부호화

Perceptual and Adaptive Quantization of Line Spectral Frequency Parameters

한 우 진*, 김 은 경*, 오 영 환*

(Woo-Jin Han*, Eun-Kyoung Kim*, Yung-Hwan Oh*)

*한국과학기술원 전산학과

(접수일자: 2000년 10월 17일; 채택일자: 2000년 11월 11일)

선 스펙트럼 주파수를 양자화하기 위한 대부분의 방법들이 가중 유클리드 거리에 기반하고 있는 반면, 본 논문에서는 청각 마스킹 효과에 기반한 에리 척도를 사용하여 선 스펙트럼 주파수를 효과적으로 양자화하는 방법을 제안하였다. 제안한 방법에서는 noise-to-mask ratio (NMR)를 선 스펙트럼 주파수의 양자화에 적합하도록 변형한 새로운 에리 척도를 유도하고, 이를 사용하여 선 스펙트럼 주파수를 양자화한다. 한편, 본 논문에서는 양자화하고자 하는 음성 프레임이 갖는 청각적인 특성을 고려하여 동적으로 비트를 할당하는 적응 양자화 알고리즘을 제안하였다. 성능 평가를 위해서 11948 프레임의 테스트 자료를 기존의 방법과 제안한 방법으로 각각 양자화하고 perceptually transparent frame의 비율 및 이때의 평균 비트율을 비교한 결과, 기존의 방법이 1800 bps의 비트율에서 89.9%의 perceptually transparent frame을 얻은 데 비해, 제안한 방법은 770 bps의 평균 비트율에서 95.5%의 perceptually transparent frame을 얻음으로써 제안한 방법이 효과적임을 보였다.

핵심용어: 음성 코딩, 선 스펙트럼 파라미터, 청각 특성, 마스킹 효과

투고분야: 음성처리 분야 (2.2)

Line spectral frequency (LSF) parameters have been widely used in low bit-rate speech coding due to their efficiency for representing the short-time speech spectrum. In this paper, a new distance measure based on the masking properties of human ear is proposed for quantizing LSF parameters whereas most conventional quantization methods are based on the weighted Euclidean distance measure. The proposed method derives the perceptual distance measure from the definition of noise-to-mask ratio (NMR) which has high correspondence with the actual distortion received in the human ear and uses it for quantizing LSF parameters. In addition, we propose an adaptive bit allocation scheme, which allocates minimal bits to LSF parameters maintaining the perceptual transparency of given speech frame for reducing the average bit-rates. For the performance evaluation, we have shown the ratio of perceptually transparent frames and the corresponding average bit-rates for the conventional and proposed methods. By jointly combining the proposed distance measure and adaptive bit allocation scheme, the proposed system requires only 770 bps for obtaining 95.5% perceptually transparent frames, while the conventional systems produce 89.9% at even 1800 bps.

Key words: Speech coding, Line spectral frequency parameter, Auditory property, Masking effect

Subject classification: Speech signal processing (2.2)

I. 서 론

선형 예측 계수 (linear predictive coefficients; LPC)는 음성의 단구간 스펙트럼 포락을 효과적으로 표현하는 방법으로서 음성 부호화를 비롯한 음성 처리 전반 분야에서 널리 사용된다[1]. 저 전송률 음성 부호화 분야의 경우, LPC 계수를 직접 양자화하기보다는 양자화 특성이 우수한 형태인 LAR (log area ratio)[2], 반사계수 (reflection

coefficient)[3], 혹은 LSF (line spectral frequency)[4]로 변환한 후에 양자화하는 것이 보통이다. 최근의 연구들에서는 여러 실험에서 양자화 특성이 가장 우수한 것으로 나타난 LSF가 주로 사용되고 있다. LSF를 양자화하기 위한 방법은 크게 스칼라 양자화와 벡터 양자화로 나누어 볼 수 있는데 저 전송률 음성 부호화 분야의 경우 더 적은 비트수만으로 같은 성능을 낼 수 있는 벡터 양자화가 주로 사용되는 추세이다. LSF를 벡터 양자화하기 위해서 가장 먼저 고려해야 할 것은 두 벡터간의 거리를 나타내는 에리 척도이다. 일반적으로 LSF 벡터간의 거리를 측정하기 위해서 스펙트럼 왜곡 (spectral distortion; SD)이 사용되지만 LSF로부터 SD를 계산하는 것이 간단하지 않

기 때문에 수많은 벡터들과의 비교가 필요한 벡터 양자화에는 적합하지 않다. 따라서 SD와 높은 상관관계를 가지면서 LSF로부터 쉽게 구할 수 있는 가중 유클리드 거리가 많이 사용된다[5][6][7].

일반적으로 평균 SD가 1dB 이하가 되면 양자화 에러가 지각되지 않는다고 알려져 있다[8]. 그러나, SD는 단순히 두 스펙트럼간의 차이를 의미하는 척도로서 실제로 인간의 청각 기관에서 지각되는 왜곡의 특성을 정확하게 반영하지 못하는 문제점을 가지고 있다. 이러한 점을 개선하기 위해서 실제 지각되는 청각 에러를 측정하기 위한 여러 에러 척도들이 제안되었다[9][10][11]. 이러한 청각 에러 척도들은 주로 인간의 청각 기관이 갖는 특징 중 마스킹 효과를 이용하고 있다. 마스킹 효과란 높은 에너지를 갖는 성분이 존재할 경우 그 주변 주파수 영역에 있는 낮은 에너지를 갖는 성분이 들리지 않게 되는 현상을 말한다. 마스킹 효과로부터 청각 기관에서 지각 가능한 양자화 에러의 최소 허용 임계치를 의미하는 마스킹 임계치 (masking threshold)를 얻을 수 있다. 따라서, 양자화 에러를 마스킹 임계치 이하로 항상 유지함으로써 양자화 에러가 지각되지 않도록 할 수 있다. 마스킹 임계치에 기반한 청각 에러 척도 중에서 정의가 간단하고 널리 사용되는 것으로 잡음 대 마스킹 임계치 비 (noise-to-mask ratio; NMR)가 있다. NMR은 SD와 같은 유클리드 거리 기반 척도에 비해서 청각 기관에서 실제 지각되는 왜곡을 잘 반영하는 것으로 알려져 있다[12].

본 논문에서는 SD 대신 인간의 청각 기관이 가지는 특성에 기반한 에러 척도인 NMR을 사용하여 LSF를 효과적으로 벡터 양자화 하기 위한 방법들을 제안하였다. 먼저, LSF로부터 NMR을 계산하는 것이 매우 복잡하기 때문에, NMR의 정의로부터 몇 가지 가정을 적용하여 새로운 청각 에러 척도인 PWEU (perceptually weighted Euclidean distance)를 유도하였다. 또한, 대부분의 기존 양자화 방법들이 고정된 비트수를 LSF 벡터에 할당하는 반면, 내부에 복수개의 코드북이 포함되어 있는 코드북 집합을 구성한 뒤, 양자화 에러가 청각적으로 지각되지 않도록 하는 조건을 만족하는 최소의 비트만을 할당하는 적응 양자화 기법을 제안하였다. 마지막으로, 학습 자료로부터 코드북 집합을 최적화하기 위한 두 가지 알고리즘을 제안하였다. 실험 및 평가를 위해서 가중 유클리드 거리와 부분 벡터 양자화 (split vector quantization)를 적용한 기존 LSF 양자화 시스템 및 제안한 방법들을 적용한 시스템을 각각 구현하고, 성능 평가를 수행하였다.

본 논문의 구성은 다음과 같다. II장에서 기존의 LSF 벡터 양자화에 대해서 간략하게 설명한 뒤, III장에서는 NMR에 기반한 새로운 에러 척도인 PWEU의 유도 과정을 기술한다. 또한, IV장에서는 LSF를 위한 적응 양자화 기법 및 코드북 집합을 효과적으로 구성하기 위한 알고리즘들을 기술한다. V장에서는 제안한 방법들의 성능 평가 결과를 보인다. 마지막으로 VI장에서는 결론 및 향후 개선 방향에 대해서 기술한다.

II. LSF의 벡터 양자화 기법

벡터 양자화를 위해서는 먼저, 두 벡터 간의 거리를 잘 표현하는 에러 척도가 정의되어야 한다. LSF 벡터 간의 거리를 나타내는 에러 척도로서 가장 잘 알려진 것으로 스펙트럼 왜곡(spectral distortion; SD)[5]이 있다 (식 1).

$$SD(f, \hat{f}) = \frac{1}{N/2} \sum_{k=0}^{N/2-1} \left[10 \log_{10} \frac{|A'(k)|^2}{|A(k)|^2} \right] \quad (1)$$

식 1에서 $A(k)$, $A'(k)$ 는 각각 LSF 벡터 f 와 \hat{f} 에 의해 계산된 LPC inverse filter의 스펙트럼이며 N 은 DFT 크기이다. 일반적으로 벡터 양자화의 경우 최적의 코드워드를 찾기 위해서 수백개의 코드워드에 대한 에러 척도 계산 과정이 필요하다. 그런데, LSF로부터 SD를 계산하는 과정은 DFT (discrete Fourier transform)를 요구하므로 코드워드 탐색 과정에서 SD를 직접 사용하는 것은 너무 많은 연산량이 필요하다. 대신, SD를 잘 반영하면서도 LSF로부터 쉽게 계산될 수 있는 척도인 가중 유클리드 거리 (식 2)가 주로 사용된다.

$$d_w(f, \hat{f}) = \sum_i w_i (f_i - \hat{f}_i)^2 \quad (2)$$

식 2에서 f_i 와 \hat{f}_i 은 각각 f 와 \hat{f} 에서의 i 번째 LSF 계수이며 p 는 LPC 차수이다. w_i 는 i 번째 LSF 계수에 주어지는 가중치를 의미하는데 이 값은 LSF 계수간의 유클리드 거리와 실제 SD와의 불일치를 보정하기 위해 사용된다[5]. 가중치 w_i 를 효과적으로 계산하기 위한 대표적인 방법으로는 weighted LSF distance[5], spectral sensitivity[6], IHM (inverse harmonic mean)[7] 등이 있으며, 세 가지 방법 모두 기존의 유클리드 거리에 비해 같은 비트율에서 평균 SD를 감소시키는 효과가 있다. 이중 IHM은 다른 방법과 대등하거나 더 우수한 성능을 보이면서도 LSF 계수로부터 쉽게 계산될 수 있기 때문에 널리 사용된다. 식 3은 IHM에 의해 계산된 가중치이다.

$$w_i = \frac{1}{f_i - f_{i-1}} + \frac{1}{f_{i+1} - f_i} \quad \text{for } 1 \leq i \leq p \quad (3)$$

단, 식 2에서 $f_0 = 0$, $f_{p+1} = \pi$ 이다. 일반적으로 저 전송률 음성 부호화에서는 보통 20-30ms의 길이의 프레임에 대해 10차 LPC 분석을 수행하여 10개의 LSF 계수를 얻는 것이 일반적이다. 10차 LSF 벡터 전체를 벡터 양자화하는 경우, 만족할 만한 양자화 성능을 얻기 위한 필요 코드북 크기가 너무 커지기 때문에 Paliwal은 10차 LSF 벡터를 앞 뒤 5차씩 두 부분으로 나누고 각각을 벡터 양자화 하는 부분 벡터 양자화를 사용하여 24bit만으로 1.0dB의 평균 SD를 얻었다[5]. 부분 벡터 양자화에서는 보통 10차 LSF 벡터를 각각 5차의 크기를 갖는 두 개의

벡터, 혹은 각각 3, 4, 3차의 크기를 갖는 세 개의 벡터로 나누는 것이 일반적이다. 본 논문에서 사용한 부분 벡터 양자화에서는 각각 3, 4, 3차의 크기를 갖는 세 개의 벡터로 나누어 양자화 하는 방법을 사용하였다.

III. 청각 특성에 기반한 에러 척도

3.1. NMR (noise-to-mask ratio)

그 동안 인간의 청각 시스템이 갖는 특성들을 모델링 하기 위한 연구들이 많이 진행되어 왔다[13][14][15]. 그 중에서 가장 잘 알려져 있는 특성 중 하나인 마스킹(masking) 효과는 하나의 신호가 주변 주파수 영역에 있는 다른 신호에 의해서 들리지 않게 되는 현상을 말한다. 이는 강한 신호가 basilar membrane의 대응되는 주파수 위치에 강한 영향을 주게 되고 이에 의해서 주위의 약한 신호들이 감지되지 못하게 됨으로써 일어나게 된다.

마스킹 임계치(masking threshold)는 인간의 청각 기관이 갖는 여러 가지 특성들을 고려하여 계산되는 값으로서 왜곡이 인지되기 위해서 필요한 최소 크기라고 생각할 수 있다. 최근 개발된 대부분의 오디오 부호기들은 마스킹 효과를 적극 활용한 알고리즘으로 오디오 신호를 효과적으로 압축하고 있다[16][17][18][19]. 이들 부호기들은 일단 입력 신호를 분석하여 각 critical band에 대한 마스킹 임계치를 얻은 후에, 각 주파수 대역의 양자화 에러가 이 마스킹 임계치를 넘지 않도록 적절한 비트를 할당하는 방법을 사용하고 있다. 마스킹 임계치를 계산하는 방법은 어떤 청각 모델을 사용하느냐에 따라서 달라진다. 본 논문에서는 MPEG-I 표준에서 사용된 psychoacoustic model I[16]을 사용하여 마스킹 임계치를 계산하였다.

이론적으로는 모든 주파수 대역에서 양자화 에러가 마스킹 임계치보다 작기만 하면 청각적으로 양자화 에러를 전혀 지각할 수 없게 된다. 여기에 착안하여 고안한 에러 척도로서, 잡음 대 마스킹 임계치 비(noise-to-mask ratio; NMR)[12]가 있다. NMR(w)를 주파수 w에 대한 잡음 대 마스킹 임계치 비 라고 하자. 이때, NMR(w)가 음수이면 잡음 성분이 마스킹 임계치 이하의 크기를 갖는 것이므로 모든 w에 대해서 NMR(w)의 값을 0보다 작도록 만들면 양자화 에러가 지각되지 않게 된다. 식 4는 이 조건을 기술한 것이다.

$$\text{Max}_{w \in C_k} [NMR(w)] \leq 0 \quad (4)$$

3.2. LSF의 벡터 양자화를 위한 청각 에러 척도

NMR은 LSF로부터 직접 계산하기가 쉽지 않기 때문에 NMR을 LSF의 벡터 양자화에 사용하기 위해서는 NMR의 특징을 잘 반영하면서도 계산이 간편한 에러 척도가 꼭 필요하다. 본 논문에서는 평균 NMR의 정의를 변형함으로써 새로운 청각 에러 척도를 유도하였으며 이 척도가 LSF로부터 간편하게 계산될 수 있음을 보였다. 식 5는 평균 NMR의 정의이다.

$$\overline{NMR} = 10 \log_{10} \left\{ \frac{2}{N} \sum_{w=0}^{N/2-1} 10^{0.1 NMR(w)} \right\} \quad (\text{dB}) \quad (5)$$

식 5에서 N은 DFT 크기이며, NMR(w)는 주파수 w에 대한 NMR을 의미한다. 식 5를 최소화시키는 것은 괄호 내의 summation 부분을 최소화시키는 것과 같으므로 식 6과 같이 이 부분을 D라는 기호로 정의하자.

$$D = \sum_{w=0}^{N/2-1} 10^{0.1 NMR(w)} \quad (6)$$

한편, 하나의 LSF 계수 값이 변화할 때, 그 영향은 대부분 LSF 계수 주위의 주파수 영역으로 국한된다[6]. 전체 주파수 영역을 어떤 LSF 계수에 의해 주로 영향을 받는 영역인가에 따라서 LPC 분석 차수인 p개로 나누고, 각각의 겹치지 않는 영역을 R_i 라고 하자. 즉, R_i 는 i번째 LSF 계수에 의해 주된 영향을 받는 주파수 영역을 의미한다. 이때, 식 6은 식 7과 같이 바꾸어 쓸 수 있다.

$$D = \sum_{i=1}^p \sum_{w \in R_i} 10^{0.1 NMR(w)} \quad (7)$$

단, R_i 를 정확하게 추정하는 것이 어렵기 때문에, R_i 는 i번째 LSF 계수의 주변 값으로 식 8과 같이 추정하였다.

$$R_i = \left[\frac{f_{i-1} + f_i}{4\pi} N, \frac{f_i + f_{i+1}}{4\pi} N \right) \quad (8)$$

식 8에서 f_i 는 radian으로 나타낸 i번째 LSF 계수 값이며, $f_0 = -f_1$, $f_{p+1} = 2\pi - f_p$ 이다. 이때, $R_1 \dots R_p$ 가 모든 주파수 영역 $[0, N/2)$ 을 포함하게 된다.

식 7에서, NMR(w)는 식 9와 같이 두 부분으로 나누어 쓸 수 있다.

$$NMR(w) = SMR(w) - SNR(w) \quad (9)$$

식 9에서 SMR(w)은 주파수 w에서의 신호 대 마스킹 임계치 비이며, SNR(w)은 주파수 w에서의 신호 대 잡음 비를 의미한다. SMR(w)를 계산하기 위해서는 각 주파수 값에 대한 마스킹 임계치를 알아야 하는데, 본 논문에서 마스킹 임계치를 계산하기 위해서 선택한 청각 모델인 MPEG-1 psychoacoustic model I의 경우 각 critical band별 마스킹 임계치 값만을 계산하므로 식 10에 의해 각 주파수 별 마스킹 임계치로 변환하여 구한다.

$$SMR(w) = SMRb_k \quad \text{for } w \in C_k \quad (10)$$

식 10에서 C_k 는 k번째 critical band에 해당하는 주파수 영역이며, SMRb_k는 k번째 critical band에 대한 신호

대 마스크 임계치 비이다.

식 9의 두 번째 부분인 $SNR(w)$ 는 주파수 w 에서의 신호 대 잡음 에너지 비로 정의된다. LSF 양자화의 경우, 신호 에너지는 양자화되지 않은 LPC로부터 구한 스펙트럼 포락이며, 잡음 에너지는 양자화 에러를 의미하므로 식 11과 같이 쓸 수 있다.

$$SNR(w) = -10 \log_{10} |1 - \hat{S}_L(w)/S_L(w)|^2 \quad (11)$$

단, 식 11에서 $S_L(w)$ 은 양자화 되기 전의 LPC로부터 계산된 스펙트럼 포락이며, $\hat{S}_L(w)$ 은 양자화 된 LPC로부터 얻은 스펙트럼 포락이다. 한편, 주파수 w 에서의 스펙트럼 왜곡인 $SD(w)$ 는 식 12와 같이 쓸 수 있다.

$$SD(w) = \sqrt{(10 \log_{10} |\hat{S}_L(w)/S_L(w)|^2)^2} \quad (12)$$

$SD(w)$ 와 $SNR(w)$ 의 관계를 얻기 위해서, 식 12의 $SD(w)$ 를 제곱한 후 상용 로그를 취한다. 그 후 $\hat{S}_L(w)/S_L(w) = r$ 로 치환하고 $r=1$ 에서 테일러 시리즈 전개하면 식 13을 얻을 수 있다.

$$\begin{aligned} \log_{10} SD(w)^2 &= \log_{10} \frac{400}{(\ln 10)^2} + \log_{10} (1-r)^2 \\ &\quad - \frac{r-1}{\ln 10} + \frac{5(r-1)^2}{12 \ln 10} \\ &\quad + O(r-1)^3 \end{aligned} \quad (13)$$

단, 식 13에서 \ln 은 자연 로그 함수, \log_{10} 은 상용 로그 함수를 의미한다. 식 13의 결과를 보면, 두 번째 항을 $SNR(w)$ 로 표현할 수 있음을 알 수 있다. 원음 스펙트럼 포락과 양자화된 스펙트럼 포락 간 오류가 상당히 작다고 가정하면 $r \approx 1$ 이 되므로 식 13에서 두번째 항까지를 취한후, $SNR(w)$ 로 치환하면 식 14을 얻는다.

$$\begin{aligned} \log_{10} SD(w)^2 &\approx \log_{10} \frac{400}{(\ln 10)^2} - 0.1 SNR(w) \\ &\approx c - 0.1 SNR(w) \end{aligned} \quad (14)$$

where $c = \log_{10} \frac{400}{(\ln 10)^2} \approx 18.78$ dB

식 14를 $SNR(w)$ 에 대해 정리하면 식 15와 같이 최종적으로 $SNR(w)$ 와 $SD(w)$ 의 관계를 얻을 수 있다.

$$SNR(w) \approx -10 \log_{10} SD(w)^2 + c \text{ (dB)} \quad (15)$$

그림 1은 식 15의 좌변과 우변에 대해 $\hat{S}_L(w)/S_L(w)$ 의 값을 0.5에서 1.5까지 변화시켜가면서 얻은 값을 그래프로 나타낸 것이다. 그림 1을 보면 $\hat{S}_L(w)/S_L(w)$ 의 값이 1에 가까운 경우 양변의 값이 거의 일치하는 것을 확인할 수 있다. 즉, 원음 스펙트럼 포락과 양자화된 후의

스펙트럼 포락이 충분히 가까운 경우 식 15가 잘 성립한다는 것을 알 수 있다. 식 15와 식 9를 식 7에 대입하고 정리하면 식 16을 얻을 수 있다.

$$D = \sum_{i=1}^L \sum_{w \in R_i} SD(w)^2 10^{0.1(SMR(w)-c)} \quad (16)$$

식 16을 보면 두 번째 summation part에서 i 번째 LSF 계수에 대응되는 영역인 R_i 내에서의 합계를 계산한다. LSF 계수간의 중요도를 결정하는데 가장 중요한 것은 한 주파수 영역 R_i 내에서의 변이가 아니라 서로 다른 LSF 계수에 대응하는 영역간의 상대적 변이이다. 따라서, 본문에서는 같은 R_i 내에서는 $SMR(w)$ 값이 일정하다고 가정하고, 식 16의 $SMR(w)$ 를 R_i 내에서의 평균 $SMR(w)$ 값으로 대치하였다. 식 17은 평균 $SMR(w)$ 의 정의이다.

$$\overline{SMR}_{R_i} = 10 \log_{10} \frac{1}{|R_i|} \sum_{w \in R_i} 10^{0.1 SMR(w)} \quad (17)$$

식 16의 $SMR(w)$ 를 식 17로 대치하면 식 18을 얻는다.

$$D = 10^{-0.1c} \sum_{i=1}^L 10^{0.1 \overline{SMR}_{R_i}} \sum_{w \in R_i} SD(w)^2 \quad (18)$$

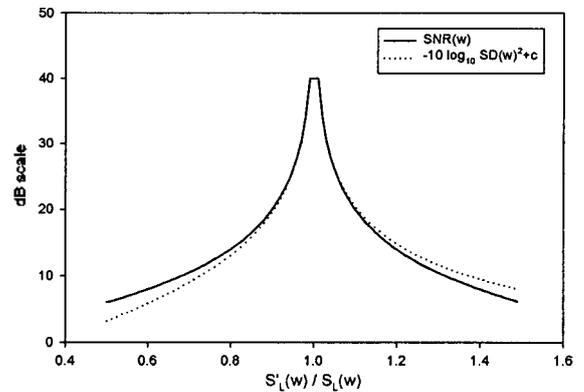


그림 1. $\hat{S}_L(w)/S_L(w)$ 에 대한 $SNR(w)$ 및 $-10 \log_{10} SD(w)^2 + c$ 의 변화

Fig. 1. $SNR(w)$ vs $-10 \log_{10} SD(w)^2 + c$.

식 18에서, 두 번째 summation part는 i 번째 LSF 계수에 대응되는 영역인 R_i 내에서의 총 스펙트럼 왜곡을 의미하므로, 식 19와 같이 가중 유클리드 거리로 근사시킬 수 있다.

$$\sum_{w \in R_i} SD(w)^2 \approx \gamma w_i (f_i - \hat{f}_i)^2 \quad (19)$$

식 19를 식 18에 대입하고, 상수항을 제거하면 최종

청각 에러 척도인 PWEU를 얻을 수 있다.

$$PWEU(f, \hat{f}) = \frac{10^{0.1c}}{\gamma} D = \sum_{i=1}^N w_i 10^{0.1 SMR_k} (f_i - \hat{f}_i)^2 \quad (20)$$

식 20을 보면 기존의 가중 유클리드 거리와 같은 형태이나 i 번째 LSF 계수에 주어지는 가중치가 하나 더 추가된 것을 알 수 있다. 이 가중치는 i 번째 LSF 계수에 대응되는 영역 내에서의 평균 SMR을 의미한다. SMR 값이 큰 경우, 신호 에너지에 비해서 상대적으로 마스크 임계치가 작다는 것을 의미하므로 양자화 에러가 작게 유지되어야 한다. PWEU는 이러한 경우 가중치를 더 많이 주게 되며 반대로 SMR 값이 작은 경우 상대적으로 허용 양자화 에러가 큰 것을 의미하므로 가중치를 더 적게 주게 된다. 이러한 방식으로 마스크 임계치의 상대적 크기에 따라 적절하게 가중치를 분배함으로써 양자화 성능을 개선할 수 있다.

IV. LSF 벡터의 적응 양자화 기법

기존의 LSF 양자화 방법들은 대부분 고정된 비트 수를 LSF 벡터에 할당한다. 이 경우, 항상 일정한 비트율을 얻을 수 있다는 장점이 있지만 양자화 성능을 최대화하기가 힘들다. 가장 효율적인 양자화 방법이 되기 위해서는 양자화 에러가 지각되지 않는 범위에서 최소의 비트 수만을 할당해야 한다. 따라서, 현재 양자화하고자 하는 음성 프레임의 특징에 맞게 가변적으로 비트들을 할당하는 적응 양자화 기법을 사용하면 양자화 성능을 개선할 수 있다. 제 4장에서는 부분 벡터 양자화를 사용한 LSF 벡터 양자화기에서 각 부분 벡터에 할당하는 비트 수를 가변적으로 결정하기 위한 적응 양자화 기법 및 평균 비트율을 최소화 시키기 위한 알고리즘들을 기술한다.

4.1. LSF의 적응 양자화 기법

일반적인 부분 벡터 양자화에서는 각 부분 벡터별 코드북을 1개씩 생성한 뒤, 각 부분 벡터별로 구성된 단일 코드북에서 최적 코드워드를 탐색하는 방법을 사용한다. 이에 비해, 본 논문에서는 그림 2와 같이 여러 가지 크기를 갖는 복수개의 코드북으로 이루어진 코드북 집합을 구성한 후, 이 집합 내에서 양자화 에러가 지각되지 않는 조건을 만족하는 가장 작은 크기의 코드북 및 최적 코드워드를 선택하여 사용하는 방법을 제안하였다. 그림 2는 코드북 집합 내에 코드북이 5개 있는 경우를 나타낸 것이다. 10차 LSF 벡터를 3, 4, 3차의 크기를 갖는 3개의 부분 벡터들로 나누고, 각각에 대응되는 코드북 집합을 C_1, C_2, C_3 라고 하자. 이때, 각 i 번째 코드북 집합은 식 21과 같이 쓸 수 있다.

$$C_i = \{C_{i1}, C_{i2}, C_{i3}, \dots, C_{iN_i}\} \quad (21)$$

식 21에서 N_i 는 i 번째 코드북 집합에 존재하는 코드북의 갯수이며, C_{ij} 는 i 번째 코드북 집합 내에 있는 j 번째 코드북을 의미한다. 각 코드북들은 j 에 따라 크기별로 정렬되어 있다. 제안한 방법에서 코드워드 탐색 과정은 각 부분 벡터별로 두 단계의 과정을 거치게 된다.

첫 번째 단계에서는 먼저, 제 3장에서 제안한 청각 에러 척도에 의해 코드북 집합 내의 한 코드북을 사용하여 최적 코드워드를 찾게 된다. 양자화 되기 전의 LSF 벡터를 f 라고 하고, 각 차수의 LSF 계수들을 f_i 라고 하자. 또한, 세 개의 각 부분 벡터를 s_1, s_2, s_3 이라고 하고, $f = s_1 s_2 s_3$ 와 같이 표기하도록 하자. 이때, 세 개의 코드북 집합에서 세 부분 벡터에 대한 최적 코드워드를 구하기 위해서 각각 j, k, l 번째 코드북을 사용했다고 하면 식 22에 의해 최적 코드워드를 구할 수 있다.

$$\begin{aligned} \hat{s}_1 &= \arg \min_x PWEU(s_1 s_2 s_3, x s_2 s_3) \\ \hat{s}_2 &= \arg \min_y PWEU(s_1 s_2 s_3, s_1 y s_3) \\ \hat{s}_3 &= \arg \min_z PWEU(s_1 s_2 s_3, s_1 s_2 z) \end{aligned} \quad (22)$$

for $\forall x \in C_{1j}, \forall y \in C_{2k}, \forall z \in C_{3l}$

식 22에서 $\hat{s}_1, \hat{s}_2, \hat{s}_3$ 은 각 코드북 집합 내에서 j, k, l 번째 코드북을 사용했을 경우 각 부분 벡터에 대한 최적 코드워드이다.

두 번째 단계에서는 식 22에 의해 얻은 코드워드를 사용할 때 발생하는 양자화 에러의 지각 가능 여부를 검사한다. 그러나, 이를 위해서는 모든 j, k, l 의 조합에 대해

$\hat{s}_1, \hat{s}_2, \hat{s}_3$ 를 계산하고, 이에 대한 양자화 에러의 지각 여부를 검사해야 하는데 j, k, l 의 조합이 너무 많기 때문에 이러한 방법 대신, 각 부분 벡터에 대한 최적 코드워드들을 순차적으로 찾는 방법을 사용한다. 즉, 먼저 식 22의 첫 번째 식을 사용하여 \hat{s}_1 을 계산하고, 나머지 부분 벡터는 양자화되지 않은 $\hat{s}_1 s_2 s_3$ 를 사용하여 양자화 에러의 지각 여부를 검사한다. 그 후, 식 22의 두 번째 식을 사용하여 \hat{s}_2 를 얻은 후, $\hat{s}_1 \hat{s}_2 s_3$ 을 사용하여 양자화 에러의 지각 여부를 검사한다. 최종적으로, 식 22의 마지막 식에 의해 \hat{s}_3 를 얻고, $\hat{s}_1 \hat{s}_2 \hat{s}_3$ 에 대한 양자화 에러 지각 여부를 검사하게 된다. 이러한 순차 탐색 방법은 항상 최적해를 보장하는 것은 아니지만 LSF 벡터 자체가 각 차수별로 독립적인 특성을 가지므로 대부분의 경우 만족스러운 결과를 보여 주었다.

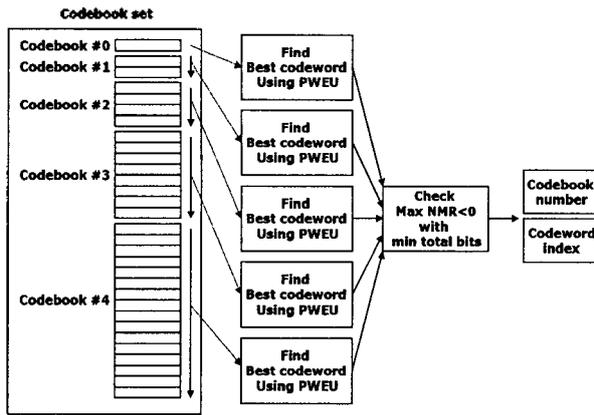


그림 2. LSF 벡터의 적응 양자화 기법
Fig. 2. Adaptive quantization technique for LSF parameters.

양자화 에러의 지각 여부는 식 4와 같이 최대 NMR 값이 음수인지 검사하고, 조건을 만족하면 각 부분 벡터에 대한 검색을 중단하게 된다. 조건을 만족하지 않는 경우 크기가 더 큰 코드북에 대해서 위의 단계를 반복하게 된다. 두 LSF 벡터 f 와 \hat{f} 에 의해 계산된 최대 NMR을 $\text{MaxNMR}(f, \hat{f})$ 이라 하면, 각 부분 벡터에 대한 양자화 에러의 지각 여부 검사는 식 23에 의해 세가지 조건으로 정리된다.

$$\begin{aligned} \text{MaxNMR}(s_1 s_2 s_3, \hat{s}_1 \hat{s}_2 \hat{s}_3) &\leq 0 \\ \text{MaxNMR}(s_1 s_2 s_3, \hat{s}_1 \hat{s}_2 \hat{s}_3) &\leq 0 \\ \text{MaxNMR}(s_1 s_2 s_3, \hat{s}_1 \hat{s}_2 \hat{s}_3) &\leq 0 \end{aligned} \quad (23)$$

제안한 방법에서 PWEU와 최대 NMR의 두 가지 에러 척도를 동시에 사용하는 이유는 PWEU가 평균 NMR로부터 근사된 것이므로 평균적인 청각 에러를 줄이는데는 유효한 반면, 양자화 에러의 지각 여부를 판단하는데에는 적합하지 않기 때문이다. 양자화의 최종 목적은 양자화 에러가 지각되지 않도록 하는 것이므로 첫 번째 단계에서도 최대 NMR을 사용하여 코드워드 탐색을 하는 것이 가장 좋겠지만 NMR의 계산이 상대적으로 많은 연산량을 요구하기 때문에 코드워드 탐색 과정에서는 보다 적은 연산량을 요구하도록 고안된 청각 에러 척도를 사용하고, 일단 선택된 코드워드의 적합성 여부는 최대 NMR을 별도로 계산하여 사용한다. 이 경우, 최대 NMR은 코드북 집합 내의 코드북 수에 해당하는 횟수만큼만 계산하게 되므로 연산량이 크게 증가되지는 않는다. 코드북 탐색이 모두 끝나면 각 부분 벡터마다 몇 비트 크기를 갖는 코드북을 사용했는가에 대한 정보와 그 코드북 내에서의 코드워드 인덱스를 얻게 된다.

4.2. 할당 비트 수의 스칼라/벡터 최적화에 기반한 코드북 집합 구성

제안한 방법에서는 LSF의 각 부분 벡터별 코드워드 인덱스 이외에도 어떤 코드북이 사용되었는가에 대한 정

보를 전송하기 위한 별도 비트수가 필요하다. 만일, 하나의 코드북 집합 내에 0비트에서 12비트사이의 크기를 갖는 코드북들이 모두 존재한다면 하나의 코드북 집합에 대해서 4비트의 별도 비트수가 필요하므로 총 12비트가 추가로 필요한 셈이 된다. 이러한 추가 비트수를 줄이기 위해서는 하나의 코드북 집합 내에 있는 코드북의 수를 최소화할 필요가 있다. 이를 위해, 학습 자료에 대해 4.1절의 알고리즘을 적용시킨 후, 각각의 부분 벡터에 대한 비트수 할당 결과를 살펴 본 결과, 그 분포가 균등하지 않다는 것을 알 수 있었다. 본 논문에서는 이러한 분포를 잘 반영할 수 있는 최소 크기의 코드북 집합을 얻기 위한 두 가지 방법을 제안하였다.

표 1. 스칼라 양자화에 의해 3비트/부분 벡터로 최적화된 코드북 집합내 코드북의 비트 조합

Table 1. Codebook bit combinations optimized by 3 bits/subvector scalar quantization.

부분 벡터 #1	부분 벡터 #2	부분 벡터 #3
0	0	0
1	1	1
2	2	2
3	3	3
4	5	5
6	6	8
9	8	10
12	11	12

표 2. 벡터 양자화에 의해 3비트 최적화된 코드북 집합내 코드북의 비트 조합

Table 2. Codebook bit combinations optimized by 3 bits vector quantization.

부분 벡터 #1	부분 벡터 #2	부분 벡터 #3	총 비트수
0	0	0	0
1	4	1	6
2	7	1	10
6	3	1	10
2	4	6	12
4	10	3	17
10	8	1	19
5	11	11	27

스칼라 양자화에 의한 코드북 최적화 (스칼라 최적화)

첫 번째 방법으로서, 각 부분 벡터들 간의 할당 비트수가 서로 독립적이라고 가정하고 각각을 스칼라 양자화 하는 방법으로 이 경우, 세 개의 할당 비트수 코드북을 얻을 수 있다. 표 1은 각각의 부분 벡터에 할당된 비트수들을 3-bit로 스칼라 양자화한 결과를 보여준다. 단, 코드워드 값들은 실수이므로 반올림 한 후의 결과를 나타내었다. 표 1에서는 각 부분 벡터별로 8개의 코드북이 존재하므로 각 3비트씩 총 9비트의 추가 비트수가 필요하다.

벡터 양자화에 의한 코드북 최적화 (벡터 최적화)

두 번째 방법은 각 부분 벡터들 간의 할당 비트수가 서로 연관성이 있다고 가정하고, 할당 비트수들을 모아 3

비트 벡터를 구성한 후 벡터 양자화를 수행하는 방법이다. 표 2는 할당 비트수들로 구성된 3비트 벡터들을 3비트 벡터 양자화 한 결과를 보여준다. 이 경우, 추가 비트 수는 3비트가 된다. 또한, 세 부분 벡터 전부에 대한 코드북 크기가 한번에 결정이 되므로 식 23의 세 가지 조건을 각각 검사할 필요가 없다. 즉, 크기가 결정된 세 개의 코드북에서 각각 얻은 최적 코드워드들에 대해, 식 23의 마지막 조건만을 사용해서 양자화 에러의 지각 여부를 검사하면 된다. 따라서, 첫 번째 방법에 비해서 더 적은 연산량이 필요한 장점이 있다.

V. 성능 평가

5.1. 실험 환경

본 논문에서 제안한 방법과 기존의 방법을 비교하기 위해 TIMIT 데이터베이스에서 서로 다른 화자가 발성한 문장들 중 일부를 선택하여 학습 및 성능 테스트에 사용하였다. 음성 자료는 8kHz로 다운샘플링하고, 10차 LPC 분석한 후에 LSF 벡터로 변환하였다. 이때, 분석 프레임 길이와 프레임 이동길이는 모두 20ms를 사용하였다. 총 49648개의 음성 프레임을 수집하였으며, 이 중에서 114명의 화자로부터 발생된 37700프레임을 학습 자료로 사용하였고, 33명의 화자로부터 11948프레임을 테스트 자료로 사용하였다. 제안한 방법과의 비교를 위해 구현한 기존 시스템은 가중 유클리드 거리를 사용한 부분 벡터 양자화를 사용하였다. 이때, 10차 LSF 벡터는 3, 4, 3차로 각각 나누어 양자화 하였으며, 에러 척도에서의 가중치는 IHM을 사용하였다. 또한, 제안한 청각 에러 척도에서 SD를 근사하기 위해서 사용된 가중치도 역시 IHM을 사용하였다(식 20의 w_i).

일반적으로 LSF 벡터의 양자화는 평균 SD를 성능 비교의 척도로 사용하지만 본 논문의 경우, SD에 기반한 에러 척도들의 청각적 성능을 개선하고자 한 것이므로 성능 비교를 위해 SD를 사용할 수 없다. 물론 완전한 성능 평가를 위해서는 대규모의 청각 테스트가 필요하지만 이러한 테스트는 짧은 노력과 시간이 필요하다는 문제가 있다. 따라서, 본 논문에서는 성능 비교를 위해서 양자화 한 음성 프레임들 중에서 양자화 에러가 청각적으로 지각되지 않는 프레임, 즉 perceptually transparent frame의 비율을 측정함으로써 성능을 비교하였다. 이러한 방법은 여러 오디오 부호기들과 MPEG-1 표준 오디오 부호기에서 성공적으로 활용되고 있다[16][17][18][19]. 표 3은 제 5장의 여러 실험 결과에 사용된 시스템들의 특징을 간략하게 설명한 것으로서, EU, WEU는 각각 유클리드 거리 및 가중 유클리드 거리를 사용한 현재 널리 사용되고 있는 부분 벡터 양자화 방법이다. 또한, PWEU, PWEU-A, PWEU-AV, PWEU-A는 각각 제안한 방법을 적용하여 구현된 시스템들이다.

5.2. PWEU의 성능 평가

5.2절에서는 유클리드 거리 (EU)와 가중 유클리드 거

리(WEU), 그리고 제 3장에서 제안한 청각 에러 척도 (PWEU)에 대한 성능 평가 결과를 보인다. 그림 3은 유클리드 거리 (EU)와 가중 유클리드 거리 (WEU), 그리고 제안한 청각 에러 척도 (PWEU) 각각에 대해서 perceptually transparent frame의 비율을 나타낸 것이다. 그림 3을 보면 EU와 WEU에 비해서 PWEU에 기반한 시스템이 더 높은 perceptually transparent frame 비율을 가짐으로써 제안한 에러 척도가 청각적 에러를 감소시키는데 효과적임을 알 수 있다.

표 3. 실험에 사용된 시스템들

Table 3. System types used in performance evaluation.

시스템명	특징
EU	유클리드 거리 - 기존 방법
WEU	가중 유클리드 거리 (IHM) - 기존 방법
PWEU	청각 에러 척도 (제 3장)
PWEU-A	PWEU + 적응 양자화 (제 4장) * 0-12비트 코드북 모두 사용
PWEU-AS	PWEU + 적응 양자화 (제 4장) * 스칼라 양자화 코드북 최적화
PWEU-AV	PWEU + 적응 양자화 (제 4장) * 벡터 양자화 코드북 최적화

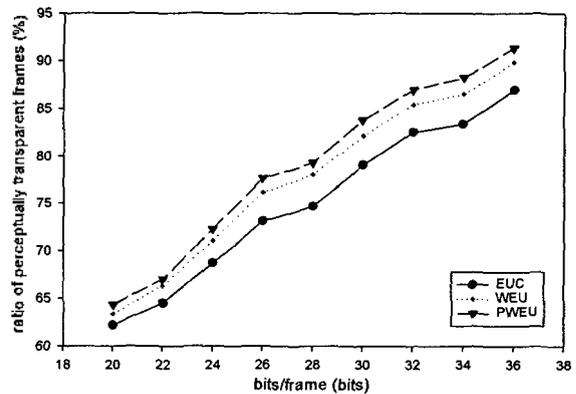


그림 3. EUC, WEU, PWEU에 대한 perceptually transparent frame의 비율

Fig. 3. Ratio of perceptually transparent frames for EUC, WEU, and PWEU.

5.3. 적응 양자화의 성능 평가

제 4장에서 제안한 적응 양자화 방법의 성능을 평가하기 위해서, WEU, PWEU, 그리고 코드북 집합을 최적화 하지 않은 시스템인 PWEU-A의 성능을 비교해 보았다. PWEU-A의 경우, 코드북 집합 마다 0~12비트의 크기를 갖는 13개의 코드북을 사용하였으며 이때 코드북 집합에 할당된 추가 비트수는 12비트이다. 표 4는 WEU, PWEU, PWEU-A의 성능을 비교한 결과이다. 단, WEU와 PWEU는 충분한 비트를 할당했을 때의 결과를 보이기 위해서 36비트 벡터 양자화를 사용하였다. 표 4에서 평균 NMR은

모든 주파수 영역에 대해 구한 NMR 값을 전체 테스트 자료에 대해 평균한 값을 의미하며, 평균 max NMR은 각 프레임 별로 구한 최대 NMR 값을 전체 테스트 자료에 대해서 평균한 것이다. 표 4의 결과에서, 평균 NMR, 평균 max NMR의 값은 WEU와 PWEU가 월등히 낮지만, 실제 perceptually transparent frame의 비율은 적응 양자화를 적용한 PWEU-A가 더 높으며, 이 때의 평균 비트율도 PWEU-A가 다른 방법에 비해서 월등히 낮게 나타났다.

표 4. 적응 양자화 방법(PWEU-A)의 성능 비교
Table 4. Experimental results for adaptive quantization technique (PWEU-A).

항목	36bit WEU	36bit PWEU	PWEU-A
평균 NMR (dB)	-12.17	-13.56	-7.43
평균 max NMR (dB)	-3.52	-3.82	-0.88
perceptually transparent frames 비율 (%)	89.85	91.33	94.57
평균 비트율 (bits)	36	36	20.02

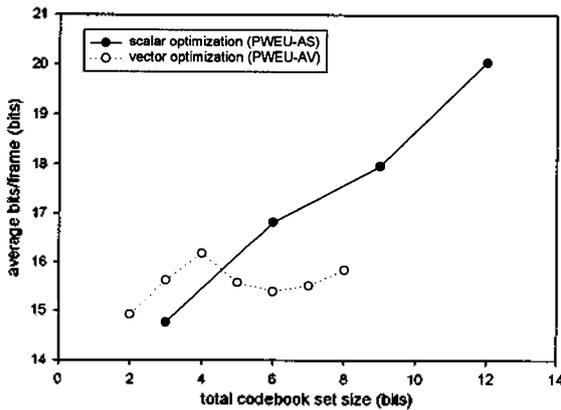


그림 4. 코드북 집합 크기에 대한 평균 비트율 변화
Fig. 4. Average bit-rates with respect to size of codebook set.

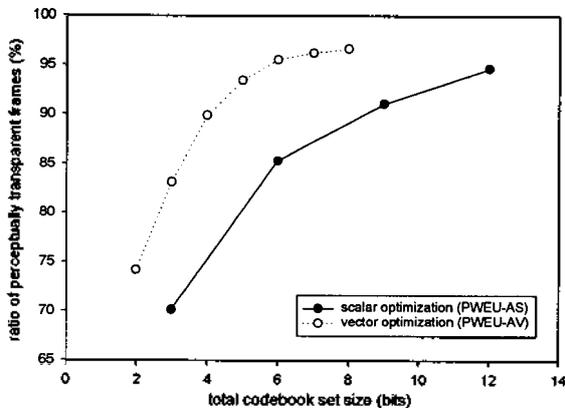


그림 5. 코드북 집합의 크기에 대한 perceptually transparent frame의 비율
Fig. 5. Ratio of perceptually transparent frames with respect to size of codebook set.

다음으로, 코드북 집합 최적화 방법의 성능을 평가하기 위해서 각각 스칼라, 벡터 최적화를 이용한 시스템인 PWEU-AS와 PWEU-AV의 성능을 비교하였다. 그림 4는 코드북 집합의 크기에 대한 평균 비트수 변화를 그래프로 나타낸 것이다. 그림 4의 결과를 보면, 코드북 집합을 위해 할당된 비트수가 5 이상인 경우 벡터 최적화 방식이 더 낮은 평균 비트율을 보임을 알 수 있다. 그림 5는 코드북 집합의 크기에 대한 perceptually transparent frame의 비율을 그래프로 나타낸 것이다. 그림 5를 보면 같은 코드북 집합의 크기에 대해서 벡터 최적화 방식이 월등히 높은 perceptually transparent frame 비율을 갖는 것을 알 수 있다. 그림 4의 결과에서 코드북 집합의 크기가 5 비트 이상인 경우 벡터 최적화 방식이 더 낮은 평균 비트율을 가지므로, 그림 5의 결과와 종합해 볼 때, 벡터 최적화 방식이 더 낮은 비트율에서 더 우수한 성능을 보임을 알 수 있다.

그림 6은 벡터 최적화 방식에 대해서, perceptually transparent frame의 비율에 대한 평균 비트율을 그래프로 나타낸 것이다. 그림 6에서 점선으로 나타낸 점은 코드북 집합의 크기가 6비트인 경우로서, 평균 비트율과 perceptually transparent frame의 비율의 두 가지 측면을 모두 고려하여 선택한 것이다. 이 경우, 평균 15.4비트, 즉 20ms 프레임 이동을 사용하는 경우 770bps의 평균 비트율만으로 95.5%의 perceptually transparent frame을 얻을 수 있었다. 그림 3에서 기존의 방법인 WEU가 36bit 벡터 양자화를 사용하는 경우 89.9% 정도의 perceptually transparent frame을 얻는 것과 비교하면 제안한 방법이 뚜렷하게 성능을 개선할 수 있음을 알 수 있다.

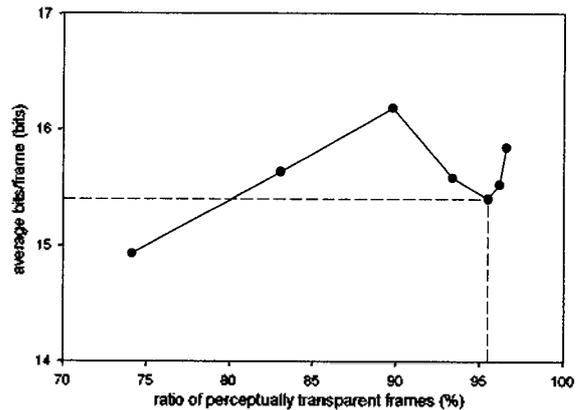


그림 6. 벡터 최적화 방식에서 perceptually transparent frame의 비율에 대한 평균 비트율
Fig. 6. Average bit-rates with respect to ratio of perceptually transparent frames in vector optimization technique.

VI. 결론

본 논문에서는 인간의 청각 시스템에서 실제 지각되는 왜곡과 밀접한 연관관계를 갖는 여러 척도인 NMR을 선

스펙트럼 주파수의 양자화에 적합하도록 변형한 새로운 에러 척도를 유도하고, 이를 사용하여 선 스펙트럼 주파수를 양자화하였다. 또한, 양자화하고자 하는 음성 프레임이 갖는 청각적인 특성을 고려하여 동적으로 비트를 할당하는 적응 양자화 알고리즘을 제안하였다. 제안한 두 가지 방법들을 함께 적용한 결과, 기존의 방법이 1800bps의 비트율에서도 89.9%의 perceptually transparent frame을 얻은 데 비해, 제안한 방법은 770bps의 평균 비트율만으로 95.5%의 perceptually transparent frame을 얻음으로써 제안한 방법이 효과적임을 보였다.

참고 문헌

1. P. Kroon and B. S. Atal, "Predictive coding of speech using analysis-by-synthesis techniques," in *Advances in Speech Signal Processing*, S. Furui and M. M. Sondhi, Eds. New York: Markel-Dekker, pp. 141-164, 1991.
2. R. Viswanathan and J. Makhoul, "Quantization properties of transmission parameters in linear predictive systems," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-23, pp. 309-321, June 1975.
3. A. H. Gray, Jr. and J. D. Markel, "Quantization and bit allocation in speech processing," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-24, pp. 459-473, Dec. 1976.
4. F. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, Vol. 57, PP. S35, Apr. 1975.
5. Kuldip K. Paliwal and Bishnu S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *IEEE Trans. Speech, Audio Processing*, Vol. 1, No. 1, pp. 3-14, Jan. 1993.
6. Frank K. Soong, "Optimal Quantization of LSP Parameters," *IEEE Trans. Speech, Audio Processing*, Vol. 1, No. 1, pp. 15-24, Jan. 1993.
7. Rajiv Laroia, Nam Phamdo and Nariman Farvardin, "Robust and Efficient Quantization of Speech LSP Parameters Using Structured Vector Quantizers," *Proc. of ICASSP*, pp. 641-644, 1991.
8. A. M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, John Wiley & Sons, 1994.
9. J. D. Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, Vol. 6, pp. 314-323, Feb. 1998.
10. M. Schroeder, "Optimizing digital speech coders by exploiting masking properties of the human ear," *J. Acoust. Soc. Am.*, 66(6), Dec. 1979.
11. T. Painter and A. Spanias, "A review of algorithms for perceptual coding of digital audio signals," *Digital Signal Processing Proceedings*, Vol. 1, pp. 179-208, 1997.
12. K. Brandenburg and T. Sporer, "NMR and Masking Flag: Evaluation of Quality Using Perceptual Criteria," *Audio Engineering Society Test and Measurement Conference*, 1992.
13. Oded Ghitza, "Auditory Models and Human Performance in Tasks Related to Speech Coding and Speech Recognition," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 1, Part II, Jan. 1994.
14. Kuansan Wang, Shihab Shamma, "Self-Normalization and Noise-Robustness in Early Auditory Representation," *IEEE Transactions on Speech and Audio Processing*, Vol. 2, No. 3, July 1994.
15. N. Jayant, J. Johnston, R. Safranek, "Signal Compression Based on Models of Human Perception," *Proc. of IEEE*, Vol. 81, No. 10, Oct. 1993.
16. ISO/IEC JTC1/SC29/WG11 MPEG IS11172-3, "Information Technology - Coding of Moving Pictures and Associated Audio for Digital Storage Media at up to About 1.5 Mbit/s, Part 3: Audio," 1992.
17. K. Brandenburg, "OCF - A New Coding Algorithm for High Quality Sound Signals," *Proc. of ICASSP*, pp. 1109-1112, May. 1990.
18. Y. F. Dehery, "A MUSICAM Source Codec for Digital Audio Broadcasting and Storage," *Proc. of ICASSP*, pp. 3605-3608, May. 1991.
19. K. Brandenburg, "ASPEC: Adaptive Spectral Entropy Coding of High Quality Music Signals," *Proc. of 90th conv. Aud. Eng. Soc.*, preprint #3011, Feb. 1991.

▲ 한 우 진 (Woo-Jin Han)



1995년 2월 : 한국과학기술원 전산학과 (학사)
 1997년 2월 : 한국과학기술원 전산학과 (석사)
 1997년 3월 ~ 현재 : 한국과학기술원 전산학과 박사과정 재학중
 ※ 주관심분야 : 음성코딩, 청각처리, 특징추출

▲ 김 은 경 (Eun-Kyoung Kim)



1995년 2월 : 한국과학기술원 전산학과(학사)
 1997년 2월 : 한국과학기술원 전산학과(석사)
 1997년 3월 ~ 현재 : 한국과학기술원 전산학과 박사과정 재학중
 ※ 주관심분야 : 음성합성, 운율제어, 패턴인식

▲ 오 영 환 (Yung-Hwan Oh)



1972년 : 서울대학교 공과대학(학사)

1974년 : 서울대학교 교육대학원
(석사)

1980년 : Tokyo Institute of Technology
정보공학전공(박사)

1981년 ~ 1985년 : 충북대학교 컴퓨터
공학과 조교수

1983년 ~ 1984년 : University of California, Davis 연구교수

1995년 ~ 1996년 : Carnegie-Mellon University 연구교수

1985년 ~ 현재 : 한국과학기술원 전산학과 교수

※ 주관심분야 : 음성인식, 음성합성, 음성코딩, 화자인식,
대화관리, 신경회로망, 전문가 시스템