

정현파 모델 부호화기를 위한 MP(Matching Pursuit) 알고리즘과 파라미터 양자화기

Matching Pursuit Estimation and Quantizer Design for Sinusoidal Model-based Coder

안 명 옥*, 정 규 혁**, 김 종 학**, 양 용 호**, 이 인 성**
(Yeong-Uk Ahn*, Gyu-Hyeok Jeong**, Jong-Hak Kim**, Yong-Ho Yang**, In-Sung Lee**)

* (주)코아로직, **충북대학교 전자공학과

(접수일자: 2005년 6월 2일; 수정일자: 2005년 7월 22일; 채택일자: 2005년 8월 20일)

본 논문에서는 고대역 (4kHz-8kHz)의 주기적 성분이 강하게 나타나는 신호에 대해서 MP (Matching Pursuit) 알고리즘을 이용한 부호화 방법을 제안한다. 또한 분석된 스펙트럼 크기 파라미터와 위상 파라미터의 효율적인 양자화 방법을 제안한다. MP 알고리즘은 오류 상쇄 원리와 정현파 모델에 바탕을 두고 있기 때문에 정확한 피치 주기 예측이 필요하다. 고대역의 정확한 피치 주기 예측을 위해 저대역 (0kHz-4kHz) 신호에서 검출한 피치 주기를 이용함으로써 부호화와 비트할당의 효율을 높일 수 있다. 스펙트럼 크기 계수의 양자화를 위해 계수들에 대해 고정 차원 이산 코사인 변환 (MDCT : Modified Discrete Cosine Transform) 및 다단계 (multi-stage) 구조를 결합시킨 양자화 기법을 사용하였고, 위상 값들은 스펙트럼 크기에 따른 가중치 필터와 위상의 2π 순환 특성을 이용하여 양자화하였다. 또한 제안한 양자화 기법과 부호화 방법을 음성 분석-합성 (analysis-by-synthesis) 시스템에 적용하여, 목적 신호와의 비교를 통해 검증한다. 향후 대역 분할을 기본 구조로 하는 계층 구조의 광대역 음성부호화기에의 적용 가능성을 제시한다.

핵심용어: MP 알고리즘, 스펙트럼 크기, 위상, 이산 코사인 변환, 양자화기

투고분야: 음성처리 분야 (2,3)

In this paper, we propose a coding method using a matching pursuit algorithm in a strongly periodic highband signal. Also, we propose an efficient quantizer for the estimated parameters: spectral magnitude and phase. Based on the error concealment principle and sinusoidal model, the MP algorithm requires the high-precision pitch period estimation. To estimate more accurate pitch period, the refined pitch obtained from lowband speech is used, which increases the efficiency of bit allocation. The spectral magnitude parameters are quantized by the method which is combined with MDCT (Modified Discrete Cosine Transform) and multi-stage structure. The spectral phase quantizer uses the 2π modular characteristic of phases and the weighted function by spectral magnitudes. To evaluate the efficiency of the proposed method, we applied it to analysis-by-synthesis system. Furthermore we suggest the possibility of scalable wideband speech codecs based on band-split structure.

Keywords: Matching Pursuit

ASK subject classification: Speech Signal Processing (2,3)

I. 서론

최근 이동통신 및 멀티미디어 분야에서 고품질 음성의 요구가 커짐에 따라 ITU-T, ISO/IEC, 3GPP 등에서 광

대역 음성 관련 표준화가 활발하게 진행되어왔다. 현재 까지 표준화된 광대역 음성 부호화기[1-3]는 전대역 부호화 방식을 사용하였거나 매우 높은 전송율을 요구하기 때문에 패킷망을 이용한 VoIP기술에의 응용에 부적절하다. 그래서 ITU-T에서는 EV (Embedded Variable bit rate)란 개념의 접근 방식으로 광대역 음성 부호화기 표준화를 계획하고 표준화 작업을 추진하고 있다[4]. 계층

책임저자: 이 인 성 (inslee@chungbuk.ac.kr)
361-763 충북 청주시 흥덕구 개신동 산48번지
충북대학교 전자공학과
(전화: 043-261-3223; 팩스: 043-261-4647)

적 구조를 가지는 대부분의 광대역 음성부호화기들은 16kHz로 표본화된 광대역 음성을 대역 분할하여 저대역 신호와 고대역 신호로 나누는 것을 기본 구조로 하고, 고대역 신호에 대하여 CELP (Code Excitation Linear Prediction) 부호화 방식이나 transform 방식을 적용한다. 이같은 부호화 방식은 고대역의 추가적인 비트할당이 적은 경우에는 주파수 왜곡에 의한 음질의 저하가 나타나게 된다[5-7]. 따라서 본 논문에서는 정현파 모델 기반 부호화 방식을 사용하여 적은 비트를 추가하는 경우에도 주파수 왜곡이 적고 원본 파형을 충실하게 부호화하는 방법을 제시한다.

정현파 모델을 사용하는 대부분의 오디오/음성부호화기는 [8]에서 제안된 윈도잉과 FFT를 이용한 피크 검출 방법에 바탕을 둔다. 특히 저전송률 음성부호화기에 사용되는 정현파 모델은 분석의 효율성을 최대화하고 계산량을 최소화하기 위해 해밍윈도잉과 512-point FFT를 사용한다. 하지만 주파수 변환시 위상 정보가 왜곡되고, 윈도잉할 때의 부엽 (side lobe)의 영향으로 파라미터 산출에 왜곡이 발생한다[9]. 최근 이런 왜곡을 피하기 위해 주파수 변환이 필요한 영상 및 오디오 부호화와 생체신호처리 분야에서는 MP 알고리즘을 이용한 연구가 진행 중에 있다. MP 알고리즘은 목적신호로부터 에너지가 가장 큰 성분들을 반복적으로 추출하여 보다 적응적으로 목적신호의 정현파 파라미터를 분석할 수 있다. MP 알고리즘의 반복과정은 목적신호와 합성신호의 차이 신호의 에너지가 영이나 어떤 정지 기준에 가까워질 때까지 수행된다[10-14].

주기성이 강한 부분에서 고대역 신호의 하모닉 분포는 저대역 (0kHz-4kHz)의 하모닉 분포와는 달리 비교적 낮은 주파수 대역에 몰려있다[15]. 따라서 4kHz-8kHz의 전대역을 FFT하여 정현파 파라미터를 검출하는 것보다 MP 알고리즘을 적용하여 낮은 대역 (4kHz)부터 정현파 파라미터를 검출하는 것은 계산량이나 파라미터의 정확성에서 개선된 성능을 얻을 수 있다. 본 논문에서는 이러한 가정하에 고대역 신호에서 주기성이 강한 부분에 대해, 정현파 모델을 기반으로 한 수정된 MP 알고리즘을 분석-합성 시스템에 적용한다. 제안한 MP 방법과 FFT 피크 검출 방법의 성능 비교를 통해 개선된 성능을 확인하고, 분석된 스펙트럼 크기 파라미터와 위상 파라미터를 효율적으로 양자화하는 방법을 제안한다.

본 논문의 2장에서 정현파 모델을 기반으로 하는 MP (Matching Pursuit) 알고리즘에 대해 설명하고, 3장에

서는 분석된 스펙트럼 크기 파라미터와 위상 파라미터의 양자화 효율을 높이고자 사용한 방법에 대해 설명한다. 4장에서는 부호화 및 양자화 방법들에 대한 성능 분석의 결과를 설명하고, 마지막으로 5장에서는 결론을 맺는다.

II. MP (Matching Pursuit) 알고리즘

2.1. 정현파 모델[8]

목적신호가 주기성이 강할 때 여기신호는 아래식과 같이 정의된 정현파 모델을 사용하여 합성될 수 있다.

$$s(n) = \sum_{k=1}^{L_k} A_k \cos(w_k n + \phi_k) \quad (1)$$

여기서 $s(n)$ 은 입력신호, L 은 하모닉 개수이고, A_k, w_k, ϕ_k 는 각각 k 번째 스펙트럼 크기, 각주파수, 초기위상이다.

2.2. 정현파 사전을 이용한 하모닉 모델

MP 방법은 오류 상쇄 (error cacealment) 원리에 바탕을 둔다[16]. 오류 함수를 정의하고 이를 구하기 위한 해법이 선형적인 해법이 없거나 많은 계산량이 필요하다면 준 최적화 (sub-optimal) 해결 방법으로 MP 방법이 사용될 수 있다. MP 이론은 주기적인 특성이 강한 음성 신호의 하모닉을 표현하기에 적당한 방법이다. 하모닉 성분을 포함하는 신호는 단구간에서 정현파 성분을 포함하고 있기 때문에 MP 알고리즘의 사전을 정현파로 구성하는 것이 효과적이다.

$$E_k^* = \sum_{n=0}^{N-1} [w(n) \{ R_{l-1}^k(n) - A_k^* \cos(w_k^* n + \phi_k^*) \}]^2 \\ = \sum_{n=0}^{N-1} [w(n) \{ R_{l-1}^k(n) - A_k^* \cos(w_k^* n) \cos(\phi_k^*) + A_k^* \sin(w_k^* n) \sin(\phi_k^*) \}]^2 \quad (2)$$

최소 자승법을 사용한 정현파 분석 모델을 이용하여 정현파 모델[17]에 대한 오류 함수를 식 (2)와 같이 정의한다. 식 (2)에서 가중치 함수 $w(n)$ 은 해밍윈도우이며, 가중치 함수는 분석 프레임에 대한 파라미터 값을 구하도록 유도한다. $R_{l-1}^k(n)$ 은 l 번째 프레임에서 1-1번째 반복 단계의 목적신호이다.

식 (2)에서 삼각함수 합성 방법에 의해 w^k 를 lw_0^k 와 같이 기본 주파수의 배수로 고정시킨다면, 변수는 $A_i^k \cos(\phi_i^k)$ 와 $A_i^k \sin(\phi_i^k)$ 만이 존재한다. 따라서 식 (2)는 2개의 파라미터에 대한 최소 자승 문제가 된다. 이 방정식의 해를 찾기 전에 식을 간단히 하기 위해서 다음과 같은 식을 정의한다.

$$a_i^k = A_i^k \cos(\phi_i^k) \tag{3}$$

$$b_i^k = -A_i^k \sin(\phi_i^k) \tag{4}$$

$$fcc = \sum_{n=0}^{N-1} [w(n) \cos(w^k n) \cos(w^k n)] \tag{5}$$

$$fcs = \sum_{n=0}^{N-1} [w(n) \cos(w^k n) \sin(w^k n)] \tag{6}$$

$$fss = \sum_{n=0}^{N-1} [w(n) \sin(w^k n) \sin(w^k n)] \tag{7}$$

$$frc = \sum_{n=0}^{N-1} [w(n) R_{i-1}^k(n) \cos(w^k n)] \tag{8}$$

$$frs = \sum_{n=0}^{N-1} [w(n) R_{i-1}^k(n) \sin(w^k n)] \tag{9}$$

식 (2)의 E_i^k 를 최소화하는 해를 구하기 위해 (3)~(9)의 식을 이용한다. (3)~(9)의 식을 행렬식으로 표현하면 식 (10)이 되며 해는 식 (11)이 된다.

$$\begin{pmatrix} a_i^k \\ b_i^k \end{pmatrix} = \frac{1}{fccfss - fcsfcs} \begin{pmatrix} fss & -fcs \\ -fcs & fcc \end{pmatrix} \begin{pmatrix} frc \\ frs \end{pmatrix} \tag{10}$$

$$a_i^k = \frac{fssfrc - fcsfrs}{fccfss - fcsfcs}, \quad b_i^k = \frac{fccfrc - fcsfrs}{fccfss - fcsfcs} \tag{11}$$

결론적으로 a_i^k 와 b_i^k 를 구하면 처음에 정의했던 스펙트럼 크기와 위상을 아래 식과 같이 구할 수 있다.

$$A_i^k = \sqrt{(a_i^k)^2 + (b_i^k)^2}, \quad \phi_i^k = -\tan^{-1}\left(\frac{b_i^k}{a_i^k}\right) \tag{12}$$

지금까지의 과정은 목적신호 $R_{i-1}^k(n)$ 와 피치주기로 부터 얻을 수 있는 기본 주파수 w_0^k 가 주어졌을 때 l 번째 단계에서 최적 A_i^k, ϕ_i^k 를 구하는 과정이다. 이것은 l 번째 단계에서 기본 주파수의 배수가 되는 주파수 성분 값들을 추출하는 과정을 뜻한다. 여기서 MP 알고리즘에 의하여 반복과정을 위한 $R_i^k(n)$ 은 식 (13)과 같이 정의

하고, $R_0^k(n)$ 은 초기 목적신호가 된다.

$$R_i^k(n) = R_{i-1}^k(n) - A_i^k \cos(w^k n + \phi_i^k) \tag{13}$$

실제로 [17]과 [18]에서 제안한 모델은 시간 변이 이득 (time-varying gain) 항을 포함하고 있지만 여기에서 제안된 모델은 시간 변이 이득 항을 포함하고 있지 않다. 이것은 시간 변이 이득 항을 구하는 과정은 실용적인 성능을 얻으려면 굉장히 큰 지연 값을 가져야 하고, 그것을 표현하는데도 많은 비트 수를 필요로 하여 저 전송률 부호화에 적합하지 못하기 때문이다.

III. 정현파 모델을 위한 스펙트럼 크기 및 위상 양자화기

3.1. MDCT를 이용한 스펙트럼 크기 양자화기

고대역 (4kHz-8kHz)에서 주기성이 강한 신호에 대해서 앞 절에서 설명한 정현파 사전을 이용한 정현파 모델링 (MP)을 한다. 이때 고대역 신호의 주파수 w_k^H 는 저대역 신호의 기본 주파수로부터 얻을 수 있다. 저대역 신호의 기본 주파수에서 고대역 신호의 기본 주파수를 얻는 이유는 대역 분리 시 저대역의 기본 주파수의 정수 배 되는 주파수가 저대역의 4kHz와 일치하지 않기 때문이다. 이러한 고대역 신호의 주파수와 저대역 신호의 주파수의 관계를 식 (14)와 같이 정의한다.

$$w_k^H = w_0^L \left(k + \left[\frac{T_p^L}{2} \right] \right) - \pi \tag{14}$$

여기서 w_0^L 은 고대역 신호의 기본 각주파수, T_p^L 은 저대역 신호의 피치 주기, $[\]$ 은 가우스 정수를 나타낸다. 식 (14)에서 정의된 고대역 각주파수를 식 (1)의 정현파 모델에 적용하여 합성된 신호는 식 (15)과 같이 정의된다.

$$s(n) = \sum_{k=1}^{L_k} A_k \cos(w_k^H n + \phi_k) \tag{15}$$

정의된 합성신호의 변수는 A_k, ϕ_k, w_k^H 이며, A_k 와 ϕ_k 는 2.2절에서 설명된 MP 정현파 분석 방법에 의하여 검출된다. 또한 w_k^H 는 식 (14)를 이용하여 저대역 신호의 양자화된 피치 주기로부터 얻을 수 있기에 추가적인 비트 할당 없이 디코더에 전송이 가능하다.

그림 1은 2단 벡터 양자화를 하는 스펙트럼 크기 양자화기 구조를 도시하고 있다. 첫 번째 단계에서는 고정차원 MDCT (modified discrete cosine transform) 변환을 이용하여 양자화한다. 이러한 MDCT 변환은 양자화 하기 위한 신호의 에너지를 한 곳으로 집중 시키는 특성이 있으므로 양자화 에러를 최소화 할 수 있다는 장점이 있다.

스펙트럼 크기 양자화를 위해 각각의 스펙트럼 크기 값들에 대한 평균 에너지에 의해서 식 (16)과 같이 정규화 된다. 정규화된 스펙트럼 크기는 양자화를 위해 식 (17)에 정의된 MDCT로 변환되어 4개의 스펙트럼 크기 후보 인덱스를 찾는다. 찾어진 후보인덱스는 두 번째 단계에서의 스펙트럼 크기와 합성하여, 최소 에러를 갖는 인덱스를 검출한다.

$$A_k^{norm} = \frac{A_k}{\sqrt{\sum_{k=0}^{L_k} \frac{A_k \cdot A_k}{L_k}}} \quad (16)$$

$$S(i) = \frac{1}{\sqrt{L_k}} \sum_{k=1}^{L_k} A_k^{norm} \lambda(i) \cos\left(\frac{(2k-1)\pi i}{2L_k}\right) \quad (17)$$

$$\lambda(i) = \begin{cases} 1 & , i = 0 \\ \sqrt{2} & , otherwise \end{cases} \quad (18)$$

여기서 A_k^{norm} 은 정규화된 스펙트럼 크기이다

4개의 후보 MDCT는 식 (17)과 같은 방법으로 IMDCT (inverse MDCT)를 하여 정규화된 스펙트럼 크기 이득과 곱한다. 목적 스펙트럼 크기와와의 차로 생기는 잔여 크기는 두 번째 단계에서 다시 한번 양자화 과정을 거쳐 최적의 잔차 스펙트럼의 크기를 검색한다.

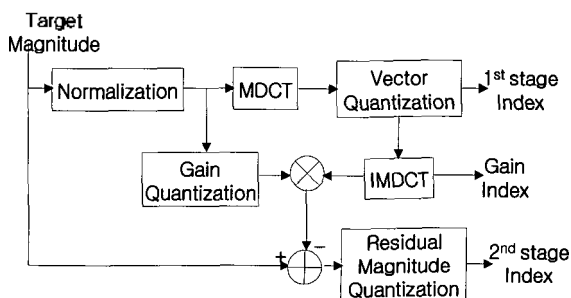


그림 1. 제안하는 스펙트럼 크기 양자화기 구조
Fig. 1. Block diagram of proposed spectral magnitude quantizer.

3.2. 순환 특성을 이용한 위상 양자화기

인간의 청각 시스템은 신호의 위상 정보를 감지할 수 있다는 많은 실험 결과들이 발표되어 왔다[19-20]. 위상 정보는 스펙트럼 크기 정보에 비해 인지 특성의 중요도가 상대적으로 작지만, 최종 출력에서는 저대역 신호와의 합성을 위해 동일한 지연을 가져야 하기 때문에 반드시 필요하다.

분석된 위상 파라미터 중에서 낮은 주파수에 해당하는 고정차원을 갖는 위상은 제안하는 순환 가중치 양자화 방법을 사용하며, 양자화 하지 않은 위상은 랜덤 위상을 사용한다. 순환 가중치 양자화는 위상의 순환 (modular 2π) 하는 특성을 이용한다[21]. 위상을 크기가 1인 단위 원 위의 극 좌표계에서의 표현 방법을 이용하여 각 차원의 목적위상과 양자화된 위상의 차를 고대역 입력신호의 스펙트럼 크기에 따라 가중치를 주어 최적화하는 양자화 방법을 제안한다.

그림 2는 제안하는 위상 양자화기의 구조를 나타낸다. 적은 비트로 많은 수의 위상을 양자화 하기 위해서 벡터 양자화를 실시하였다. 각 차원의 에러는 수직선상에서 양자화를 한다면 최대 에러는 2π 가 된다. 하지만 위상의

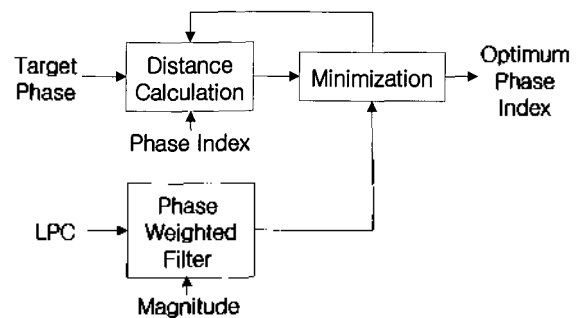


그림 2. 제안하는 위상 양자화기 구조
Fig. 2. Block diagram of proposed phase quantizer.

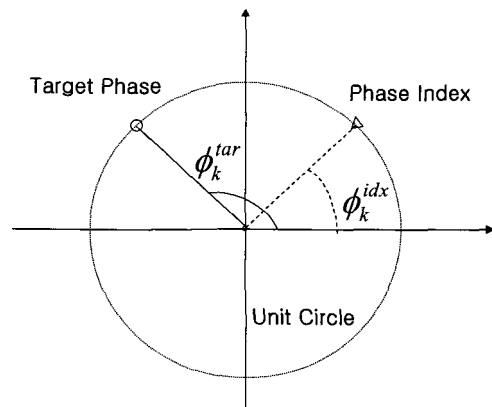


그림 3. 위상 오차 계산의 구조
Fig. 3. Structure of the phase error calculation.

순환하는 특성을 고려하여 그림 3과 같이 극좌표에서 에러를 구하게 된다면 최대 에러는 π 가 된다. 이것은 양자화 오차를 줄일 수 있다.

일반적인 차원별 에러는 식 (19)와 같이 정의된다.

$$e_k = |\phi_k^{tar} - \phi_k^{idx}| \quad (19)$$

여기서 e_k , ϕ_k^{tar} , ϕ_k^{idx} 는 각각 k 번째 차원의 에러, 목적위상, 인덱스 위상이다. 다단계 (multi-stage) 양자화가 아니라면 식 (19)를 이용한 에러로 간단히 계산이 되겠지만 다단계 위상 양자화를 한다면 각각의 차원에 대한 에러의 방향성(±)을 알아야 할 것이다. 이러한 방향성에 대한 고려를 위해서는 식 (19) 대신 표 1을 참고하여 차원별 에러를 계산한다.

순환 구조와 더불어 입력 음성의 스펙트럼 크기에 따라 인덱스 위상에서의 에러 가중치를 다르게 하여 합성된 음성이 시간축에서 원본 음성과 가장 가깝게 표현되기 위한 위상 가중치 필터를 사용하였다. 위상 가중치 필터는 LPC 계수에 의한 포락선과 LPC 잔여신호의 스펙트럼 크기를 이용하여 같은 차원의 위상에 대한 식 (20)의 위상 가중치 함수 $PW(z)$ 를 구한다.

$$PW(z) = S(z) \Big|_{z=w_k^H} \quad (20)$$

여기서 $S(z)$ 는 Z변환된 고대역 입력신호이다. 이렇게 구하여진 위상 가중치 함수는 다음식 식 (21)의 MSE (mean square error)에 의해 최적화를 한다.

$$MSE = PW^2(z) \cdot e_k^2 \quad (21)$$

여기서 최적화 과정을 위한 e_k 는 표 1의 각각의 경우에 대한 k 번째 위상 오차이다.

표 1. 목적위상의 부호에 따른 오차 계산
Table 1. Error calculation according to the sign of target phase.

| Target Index | + | - |
|--------------|--|--|
| + | $e_k = \phi_k^{tar} - \phi_k^{idx}$ | $e_k = 2\pi - \phi_k^{tar} - \phi_k^{idx} $ |
| - | $e_k = \phi_k^{tar} - \phi_k^{idx} - 2\pi$ | $e_k = \phi_k^{tar} - \phi_k^{idx}$ |

IV. 시뮬레이션 및 결과 고찰

4.1. 정현파 모델의 MP 알고리즘 성능

MP 방법의 성능이 FFT 피크 검출 방법보다 뛰어나다는 것은 그림 4를 통해 알 수 있다. 그림 4는 고대역 신호 중 주기적 성분이 강하게 나타나는 100샘플을 분석하고 합성한 결과를 나타낸다. 실험을 위하여 사용된 음성은 NTT DB 중 16 kHz로 샘플링된 "The boss ran the show with the watchfull eyes."라는 광대역 여자 음성의 일부분이다. 실선은 목적신호를 나타내고, 점선은 스펙트럼 크기와 초기 위상을 얻는 방법으로 MP 정현파 사전 방법과 FFT 방법의 주파수 해상도에 따라 분석 및 합성한 신호를 나타낸다. 기존의 정현파 모델에 사용되는 512 point FFT의 피크 검출 방법으로 분석 합성한 결과인 그림 4의 (a)는 SNR과 마찬가지로 MP 정현파 사전을 이용해 분석 및 합성한 그림 4의 (c)보다 현저히 떨

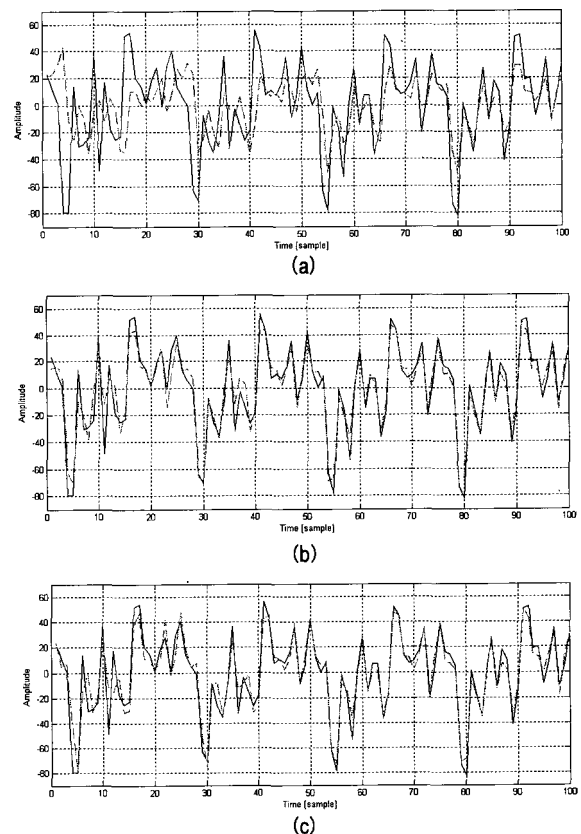


그림 4. 각 방법을 사용하여 얻은 합성된 신호 비교 : 목적신호(실선), 합성된 신호(점선)

- (a) 512-point FFT 피크 검출방법
- (b) 2048-point FFT 피크 검출방법
- (c) MP 방법

Fig. 4. Comparison of the synthesized signal : target signal (solid), synthesized signal(dash).

- (a) using 512-point FFT peak detection method.
- (b) using 2048-point FFT peak detection method.
- (c) using MP method.

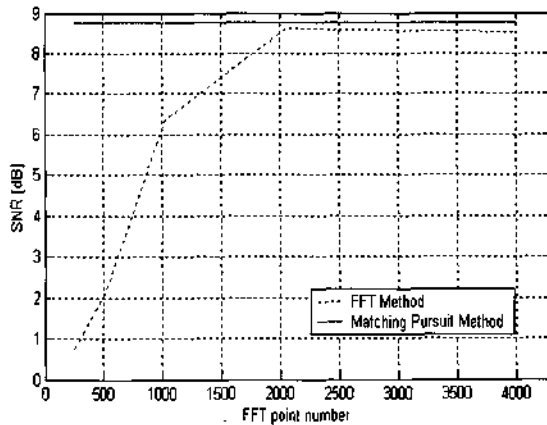


그림 5. 주파수 해상도에 따른 SNR
Fig. 5. SNR according to frequency resolution.

여지는 성능을 나타내고 있다. 시간축 파형에서도 볼 수 있듯이 512 point 피크 검출 분석을 사용하면, 포락선(envelope) 형태는 입력과 흡사함을 보이지만, 5샘플이나 17샘플 부근에서 포락선이 급하게 변하는 부분 등에서는 복원신호와 상대적으로 차이가 있음을 관찰할 수 있다. 하지만 MP 정현파 사전 방법은 시간축에서 분석 및 합성하고 잔여신호를 다시 분석 합성하는 MP의 특성상 시간축의 포락선이 급하게 변화하여도 512 point FFT 방법과 동일한 해상도 분석만으로도 포락선을 잘 표현함을 알 수 있다[13,14,16]. 또한 2048 point FFT 방법과 MP 정현파 사전 방법이 유사한 성능을 나타내는 것을 확인 할 수 있다.

그림 5에서 MP 정현파 사전 방법은 윈도우링 후에 512 point FFT를 취하는 방법에 비해 좋은 성능을 나타냄을 알 수 있다. 즉 MP 방법에 의한 신호 대 잡음비(SNR)에는 근접하려면 2048 point의 해상도를 가져야 한다.

4.2. 파라미터의 양자화기 성능

스펙트럼 크기 양자화 방법의 성능 비교를 위해 고정 차원 벡터 양자화 방법과 비교한다. MDCT를 이용한 2단 벡터 양자화 방법과의 비교를 위해 다음과 같이 형평성을 맞춘다. 8개의 스펙트럼 크기값을 2단 벡터 양자화한다. 그리고 첫 번째 단계에서 4개의 후보값을 저장하고, 두 번째 단계에서의 양자화를 통해 가장 작은 주파수 왜곡을 갖는 값으로 양자화하였다[20]. 합성시에는 8개 이상의 스펙트럼 크기값은 1을 채워주는 방법을 사용한다. 표 2에서 보는 바와 같이 2단 고정 차원 벡터 양자화 방법보다 0.5dB이상의 성능 향상을 나타내었다.

위상 양자화기의 성능 비교를 위해 사용된 위상 개수는 8개로 제한한다. 이는 인간 청각 시스템이 스펙트럼

표 2. 주파수 왜곡 시뮬레이션 결과

Table 2. Result of SD. (단위 : dB)

| 비트수 사용방법 | 10(5/5) | 12(6/6) | 14(7/7) | 16(8/8) | 18(9/9) |
|---------------------|---------|---------|---------|---------|---------|
| 2단 고정 차원 벡터 양자화기 | 3.12432 | 3.10124 | 3.02411 | 2.98345 | 2.78991 |
| 제한한 양자화기 | 2.84125 | 2.64122 | 2.43451 | 2.23445 | 2.14211 |

표 3. 신호대 잡음비 시뮬레이션 결과

Table 3. Result of Segment SNR. (단위 : dB)

| 비트수 사용방법 | 5 | 6 | 7 | 8 | 9 |
|---------------------|-------|-------|-------|-------|-------|
| 벡터 양자화 | 3.816 | 3.871 | 3.925 | 3.945 | 4.123 |
| 가중치 함수 적용 | 4.131 | 4.171 | 4.217 | 4.331 | 4.412 |
| 순환 특성 적용 | 4.034 | 4.048 | 4.078 | 4.113 | 4.178 |
| 가중치 함수와 순환 특성 적용 | 4.315 | 4.371 | 4.401 | 4.439 | 4.467 |

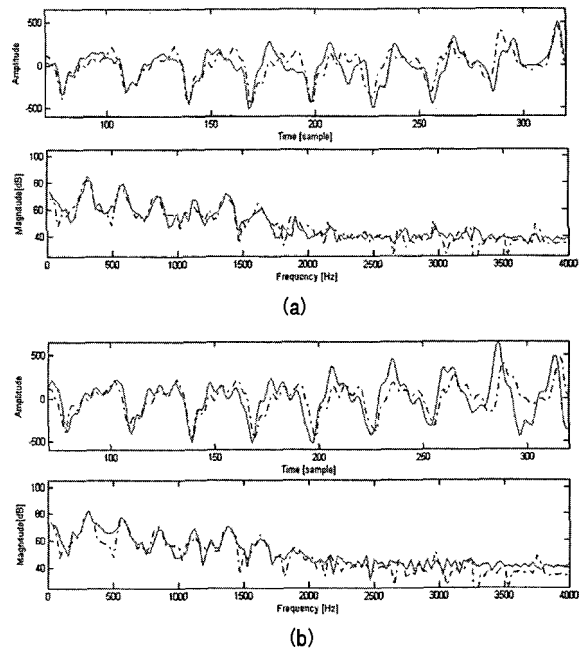


그림 6. 고대역 목적 신호(점선)와 합성신호(실선)의 비교 : 시간축(상) 주파수축(하)

(a) 파라미터 양자화 전

(b) 파라미터 양자화 후

Fig. 6. Comparison of target signal(dashed) and synthesis signal(solid) : time domain(top) and frequency domain(bottom).

(a) before quantization.

(b) after quantization.

크기 정보보다 위상 정보에 덜 민감하기 때문이다[19]. 비교를 위해 제안된 양자화 기법에 따라 8개의 위상을 벡터 양자화한다. 결과는 표 3과 같다. 양자화 비트수에 따라 일반적인 벡터 양자화한 경우보다 제안한 양자화 기법을 적용했을 때가 0.4-0.5dB정도 SNR 성능 개선을 보였다.

양자화 비트는 스펙트럼 크기는 18bit로 위상은 9bit로 양자화하여 시물레이션을 통해 검증해본다. 목적신호와 합성신호와의 비교를 그림 6을 통해 알 수 있다. 목적신호는 고대역 신호 중 피치주기가 30 ($f_0=266\text{Hz}$)인 350 샘플을 나타낸 것이다. 그림 6의 (a)에서의 스펙트럼을 보면 기본주파수의 정수배 되는 곳에서 원본신호와 오차가 적다는 것을 스펙트럼을 확인할 수 있고, (b)에서는 양자화의 오차가 음성신호의 스펙트럼에서 중요한 기본주파수의 배수 부분에서의 왜곡이 적다는 것을 알 수 있다.

V. 결론

광대역 음성부호화기에서의 성능 향상을 위해서는 고대역의 하모닉을 효율적으로 모델링하는 것이 중요하다는 점에 착안하여 고대역 신호에서 정현파 모델을 기반으로 하는 MP 알고리즘과 파라미터를 효율적으로 양자화하기 위한 스펙트럼 크기 양자화기와 위상 양자화기를 제안하였다. 정현파 모델링에서의 중요한 파라미터인 피치정보 전송에 비트할당이 전혀 없도록 저대역의 피치를 이용한 MP (Matching Pursuit) 알고리즘으로 고대역의 스펙트럼 크기와 위상의 파라미터를 분석하여 정확한 파라미터 정보를 얻었다. 스펙트럼 크기 양자화기 성능은 2단 고정 차원 벡터 양자화 방법보다 주파수 왜곡이 0.5dB이상의 성능 향상을 보였다. 위상 양자화기는 벡터 양자화보다 제안한 방법을 적용했을 때 0.4-0.5dB의 높은 SNR값을 보였다. 제안된 방법들을 대역 분할의 고대역 음성부호화기에 적용하므로써 광대역 음성부호화기의 성능 향상을 기대할 수 있다.

감사의 글

본 연구는 2003년도 한국 과학재단의 특정 기초연구 사업(과제번호 R01-2003-000-11620-0)의 지원으로 수행 되었습니다.

참 고 문 헌

1. ITU-T Recommendation. G.722. "7 kHz audio-coding within 64 kbit/s," Nov. 1988
2. ITU-T Recommendation. G.722.1, "Coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," Sep. 1999.
3. ITU-T Recommendation. G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband," Jan. 2001.
4. ITU-T SG16 Q.9, "Report of Q.9/16 meeting," Nov. 2004.
5. A. McCree, "A 14 kb/s wideband speech coder with a parametric highband model," IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 1153-1156, Jun. 2000.
6. K. Koishida, V. Cuperman, A. Gersho, "A 16-kbit/s bandwidth scalable audio coder based on the G.729 standard," IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 1149-1152, Jun. 2000.
7. K. T. Kim, S. K. Jung, Y. C. Park, D. H. Youn, "A new bandwidth scalable wideband speech/audio coder," IEEE International Conference on Acoustics, Speech and Signal Processing, 1, 657-660, May. 2002.
8. R. McAulay, T. Quatieri, "Speech Analysis/Synthesis Based on a Sinusoidal Representation," IEEE Transactions on Signal Processing, 34, 744-754, Aug. 1986.
9. A. M. Kondoz, *Digital Speech*(John Wiley & Sons Ltd., New York, 1994), 36-41.
10. O.K. Al-Shaykh, E. Miloslavsky, "Video compression using matching pursuits," IEEE Transactions on Circuits and Systems for Video Technology, 9, 123-143, Feb. 1999.
11. Yuan Yuan, D.M. Monro, "Improved Matching Pursuits Image Coding," IEEE International Conference on Acoustics, Speech and Signal Processing, 2, 201-204, Mar. 2005.
12. K. Skretting, K. Engan, J.H. Husoy, "ECG compression using signal dependent frames and matching pursuit," IEEE International Conference on Acoustics, Speech and Signal Processing, 4, 585-588, Mar. 2005.
13. P. Vera-Candeas, N. Ruiz-Reyes, "New matching pursuit based sinusoidal modelling method for audio coding," IEE Proceedings on Vision, Image and Signal Processing, vol. 151, 21-28, Feb. 2004.
14. Lajos Hanzo, F. Clare, A. Somerville and Jason P. Woodard, *Voice Compression and Communications* (The Institute of Electrical and Electronic Engineering, Inc., New York, 2001), 531-564.
15. 송재중, 박호중, 김무영, 김도석, 김정수, "광대역 신호 압축기를 위한 주파수 대역 특성에 선택적인 양자화 방법," 음향학회지 20 (7), 76-82, 2001.
16. S. G. Mallet, Zhifeng Zhang, "Matching pursuit with time-frequency dictionaries," IEEE Transactions on Signal Processing, 41, 3397-3415, Dec. 1993.
17. E. B. George, M. J. T. Smith, "Speech analysis/synthesis and modification using an analysis-by-synthesis/overlap-add sinusoidal model," IEEE Transactions on Signal Processing, 5, 389-406, Sep. 1997.
18. E. B. George, M. J. T. Smith, *Audio analysis/synthesis system*, (U.S Patent 5327518, Jul. 1994).
19. F. A. Bilson, "On the influence of the number and

phase of harmonics on the perceptibility of the pitch of complex signals," *Acoustica*, 28, 60-65, Sep. 1973.

20. 김도석, "인지에 중요한 음향신호의 위상에 대해," *음향학회지* 19 (7), 28~33, 2000.

21. Alan V. Oppenheim, Ronald W. Schaffer, *Discrete-time Signal Processing-2nd ed*(Prentice Hall, New Jersey, 1993), 240~339.

22. Peter Lupini, Vladimir Cuperman, "Nonsquare Transform Vector Quantization," in *IEEE Signal Processing Letters*, 3 (1), Jan. 1996.

• 이 인 성 (In Sung Lee)



1983년 2월: 연세대학교 전자공학과 (공학사)
 1985년 2월: 연세대학교 전자공학과 (공학석사)
 1992년 2월: Texas A&M University 전기공학과 (공학박사)
 1993년 2월~1995년 9월: 한국전자 통신연구원 이동통신 기술연구단 산업연구원
 1995년 10월~현재: 충북대학교 전기전자공학부 정교수
 ※주관심분야: 음성/영상 신호 압축, 이동통신, 적응필터

저자 약력

• 안 영 옥 (Yeong Uk Ahn)



2003년 2월: 충북대학교 반도체공학과 (공학사)
 2005년 2월: 충북대학교 전자공학과 (공학석사)
 2005년 2월~현재: (주) 코아로직
 ※주관심분야: 음성/오디오 부호화, 통신신호처리, 적응필터

• 정 규 혁 (Gyu-Hyeok Jeong)



2004년 2월: 충북대학교 전기전자공학(공학사)
 2004년 3월~현재: 충북대학교 전자공학과 (공학석사)
 ※주관심분야: 음성/오디오 부호화, 통신신호처리

• 김 종 학 (Jong Hak Kim)



1998년 2월: 충북대학교 전자공학과 학사
 2000년 2월: 충북대학교 전자공학과 석사
 2000년 3월~현재: 충북대학교 전자공학과 박사과정
 ※주관심분야: 음성/오디오 부호화, 영상압축, 적응필터

• 양 용 호 (Yong Ho Yang)



2002년 2월: 송주대학교 전자공학과 학사
 2004년 8월: 충북대학교 전자공학과 석사
 2005년 3월~현재: 충북대학교 전자공학과 박사과정
 ※주관심분야: 음성/오디오 부호화, DSP 실시간 구현