

웨이블릿 변환을 이용한 잡음제거기 설계

Design of the Noise Suppressor Using Wavelet Transform

원 호 진*, 김 종 학**, 이 인 성**
(Ho-jin Won*, Jonghark Kim**, Insung Lee**)

*LG전자, **충북대학교 전자공학과

(접수일자: 2001년 4월 30일; 수정일자: 2001년 7월 12일; 채택일자: 2001년 7월 30일)

본 논문에서는 웨이블릿 변환을 이용한 주변 잡음제거기를 제안하였다. 기존의 고정된 시간-주파수 해상도를 가지는 단구간 푸리에 분석법 대신 다양한 시간-주파수 해상도를 제공하는 웨이블릿 분석법을 사용함으로써 시간 특성이 변하는 베이블 (Babble) 잡음에 좀더 효율적인 잡음제거 방법을 설계하였다. 본 논문에 제안된 웨이블릿 변환 잡음제거기는 스펙트럴 차감법에 기반하여 구성하였으며, 고주파 영역에서 높은 시간 해상도를 갖는 웨이블릿 마스크 패턴을 사용함으로써 시간 특성이 빠르게 변화하는 고주파 잡음에 더욱 효율적인 동작을 하도록 설계하였다. 성능평가를 위해 차량 잡음, 길거리 잡음, 베이블 잡음과 같은 이동 통신에서 많이 사용하는 주변잡음에서 시험하였으며, 그 주관적 음질 평가 결과 베이블 잡음의 경우 기존의 EVRC (Enhanced Variable Rate Coder) 잡음 제거기보다 MOS (Mean Opinion Score) 0.2의 성능 개선을 이룰 수 있었다. 출력 음성의 스펙트로그램에서도 성능 개선을 확인할 수 있었다.

핵심용어: 음성부호화, 잡음제거, 웨이블릿 변환, 스펙트럼 차감법

투고분야: 음성처리 분야 (2.2)

This paper proposes a new noise suppression method using the Wavelet transform analysis. The noise suppressor using the Wavelet transform shows the more effective advantages in a babble noise than one using the short-time Fourier transform. We designed a new channel structure based on spectral subtraction of Wavelet transform coefficients and used the Wavelet mask pattern with more higher time resolution in high frequency. It showed a good adaptation capability for babble noise with a non-stationary property. To evaluate the performance of proposed noise canceller, the informal subjective listening tests (MOS tests) were performed in background noise environments (car noise, street noise, babble noise) of mobile communication. The proposed noise suppression algorithm showed about MOS 0.2 performance improvements than the suppression algorithm of EVRC in informal listening tests. The noise reduction by the proposed method was shown in spectrogram of speech signal.

Keywords: Vocoder, Noise suppression, Wavelet transform, Spectral subtraction

ASK subject classification: Speech signal processing (2.2)

1. 서론

디지털 이동 통신 시스템에서 통화 품질을 결정하는데 중요한 영향을 미치는 부분 중의 하나인 음성부호화 기술은 디지털 이동 통신 시스템이 발전함에 따라 매우 빠

르게 발전되어 왔으며, 새로운 이동 통신 시스템 개발과 더불어 더좋은 음질 및 서비스를 제공하기 위한 기술들이 새로이 개발되고 있다[1-3]. 특히 2.4 kbps ~16 kbps 전송률대의 음성부호화 방법은 대부분 선형예측모델 (Linear Prediction)이나 하모닉 (Harmonic) 모델을 이용하고 있으며 이러한 음성부호화기는 인간의 성대를 모델화된 구조를 갖고 있어 음성신호와 특성이 다른 주변 잡음이 포함된 음성에서는 음질 저하가 심해지는 단점을

책임저자: 이인성 (inslee@cbucc.chungbuk.ac.kr)
361-763 충청북도 청주시 흥덕구 개신동 산48번지
충북대학교 전자공학과
(전화: 043-261-3223; 팩스: 043-271-4647)

갖고 있다. 따라서 음성 주변잡음이 포함된 환경에서도 좋은 음질을 유지하기 위해 음성부호화 전단에서 주변잡음을 제거 (Noise Suppression)하고 부호화함으로써 음질 개선에 많은 도움을 주고 있다. 실제로 잡음제거기술은 음성부호화기에 포함되고 있으며 EVRC (Enhanced Variable Rate Coder)[4]는 현재 이동통신 시스템에서 상용화되어 사용되고 있으며, 비교적 잡음 환경에서도 만족스러운 음질을 보이고 있다. 또한, 잡음제거 기술은 미국 연방표준인 2.4 kbps MELP (Mixed Excited Linear Prediction) 음성부호화기에서도 사용되고 있다. EVRC 부호화기에서 사용된 잡음 제거 방식은 주파수 영역에서 잡음 섞인 음성 스펙트럼에서 주변잡음 스펙트럼을 차감하는 스펙트럼 차감 방법에 바탕을 두고 있다. 주파수대역의 각 채널별 배경잡음 에너지 레벨을 측정하고 음성신호 대 배경잡음비를 추정해냄으로써, 각 대역별 합성 이득 값을 재조정하는 방식을 사용하고 있다. 각 채널의 에너지 값은 STFT (Short Time Fourier Transform) 기반의 블록 처리에 기반하며, 시간에 대하여 균등한 채널 대역폭을 가지고 동작한다. 이러한 구조는 차량잡음, 길거리잡음 등의 그 통계적 특성이 시간에 대해서 변화가 적은 배경잡음에서 효과적으로 동작하고 있으나, 시간에 따라 통계적 특성 변화가 심한 여러 사람의 음성거림에 의한 베이بل (Babble) 잡음 같은 경우에는 분석 블록 크기에 의한 시간 분해능의 한계로 인해 잡음 제거에 한계를 갖고 있다. 따라서 본 논문에서는 이러한 시간에 따른 대역별 주파수 특성변이가 다른 베이블 잡음에서의 잡음제거 성능을 향상시킨 웨이블릿 변환 (Wavelet Transform)을 기초한 주변잡음 제거 방법을 제안한다. 다해상도 특성을 가지는 웨이블릿 변환의 장점을 이용하여 스펙트럼 차분 방식에 기초한 잡음 제거기를 설계하며, 새로운 채널 설계 및 그에 따른 장구간 평균 파워 스펙트럼 계산방법, 잡음갱신 방법 등을 제안한다.

본 논문의 2장에서는 음성에 섞인 잡음에 대한 특성을 STFT 분석을 이용한 잡음제거기에 기초하여 분석하며, 3장에서는 본 논문에서 제안된 웨이블릿을 사용한 잡음 제거기의 구조 및 각 부분에 대해서 상세히 설명한다. 4장에서는 제안된 잡음제거기의 실험결과에 대해서 논의하며, 마지막으로 5장에서 결론을 맺는다.

II. 주변 잡음의 특성

이동통신 환경에서의 주변잡음은 자동차 잡음, 거리에서의 잡음, 여러 사람의 음성거림 잡음 (Babble Noise) 등이

있다. 대부분의 주변잡음의 특징은 음성신호보다는 시간에 따라 통계적 특성이 변하지 않는 정지적 (Stationary)인 특성을 갖고 있다. 그러나 베이블 잡음은 여러 사람의 음성 신호가 합쳐진 잡음으로 시간에 따라 특성이 변하는 비정지적 (Nonstationary) 특성을 갖게 된다. 본장에서는 이러한 주변 잡음의 특성을 신호대 잡음비 (SNR)의 분산과 신호대 잡음비의 합 (Metric) 등의 요소들을 이용하여 분석한다.

음성신호는 13 ms (샘플링률 8 kHz시 104 샘플) 프레임 단위로 윈도우되어 STFT를 취한다. 푸리에 (Fourier) 변환된 $G(k)$ 값은 15개의 채널로 나누어 채널 에너지를 계산한다. m번째 프레임의 i번째 채널의 에너지 $E_{ch}(m, i)$ 는 식 (1)과 같이 계산되어진다.

$$E_{ch}(m, i) = \max \{ E_{\min}, \alpha_{ch}(m) E_{ch}(m-1, i) + (1 - \alpha_{ch}(m)) \frac{1}{f_H(i) - f_L(i) + 1} \sum_{k=f_L(i)}^{f_H(i)} |G(k)|^2 \} \quad 0 \leq i < N_c \quad (1)$$

여기서 $E_{\min} = 0.0625$ (최소 채널 에너지 값)
 $\alpha_{ch}(m) = 0.45$ (채널 에너지 상수)
 $N_c = 15$ (총 채널 수)
 $f_L(i)$: i 번째 채널의 하위 요소
 $f_H(i)$: i 번째 채널의 상위 요소

각 채널의 대역은 다음과 같이 정의된다.

$f_L = \{2, 4, 6, 8, 10, 12, 14, 17, 20, 23, 27, 31, 36, 42, 49, 56\}$
 $f_H = \{3, 5, 7, 9, 11, 13, 16, 19, 22, 26, 30, 35, 41, 48, 55, 63\}$

각 채널의 신호대잡음비는 식 (2)와 같이 계산되어진다.

$$\sigma_o(i) = \max \{ 0, \min \{ 89, \text{round} \left\{ 10 \log_{10} \left(\frac{E_{ch}(m, i)}{E_n(m, i)} \right) / 0.375 \right\} \} \} \quad (2)$$

여기서 $E_{ch}(m, i)$ 는 현재 프레임의 i번째 채널 에너지값을 나타내며 $E_n(m, i)$ 는 현재 프레임의 채널 잡음 에너지 값을 나타낸다.

각 채널의 신호대 잡음비를 합한 값을 음성 메트릭 (Metric)이라 정의하며 식 (3)과 같이 계산되어진다. 음성 메트릭은 현재 프레임을 잡음인가를 결정하는데 중요한 파라미터로 사용된다. 신호대 잡음비가 크면 음성 메트릭 값이 크게 나타나며 음성 구간일 확률이 높아진다.

$$v(m) = \sum_{i=0}^{N_c-1} V(\sigma_q(i)) \quad (3)$$

여기서 $V(k)$ 는 음성 매트릭 변환표는 다음과 같이 정의된다.

$V = \{2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7, 7, 7, 8, 8, 9, 9, 10, 10, 11, 12, 12, 13, 13, 14, 15, 15, 16, 17, 17, 18, 19, 20, 20, 21, 22, 23, 24, 24, 25, 26, 27, 28, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 50, 50, 50, 50, 50, 50, 50, 50\}$;

채널 에너지의 로그값을 식 (4)를 이용하여 계산하고, 파워 스펙트럼이라 정의한다.

$$E_{dB}(m, i) = 10 \log_{10}(E_{c_k}(m, i)) ; 0 \leq i < N_c \quad (4)$$

현재 프레임의 파워 스펙트럼과 장구간 (long-term) 평균 파워 스펙트럼 값의 차를 분산 값이라고 정의한다. 즉, 분산 값은 식 (5)와 같이 정의된다.

$$\Delta_E(m) = \sum_{i=0}^{N_c-1} |E_{dB}(m, i) - \overline{E_{dB}}(m, i)| \quad (5)$$

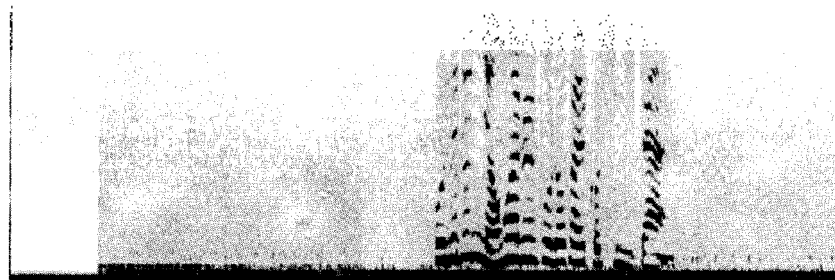
i 번째 채널의 $\overline{E_{dB}}(m, i)$ 시간평균 파워스펙트럼 값이다.

$$\begin{aligned} \overline{E_{dB}}(m+1, i) &= \alpha(m) \overline{E_{dB}}(m, i) \\ &+ (1 - \alpha(m)) E_{dB}(m, i) \end{aligned} \quad (6)$$

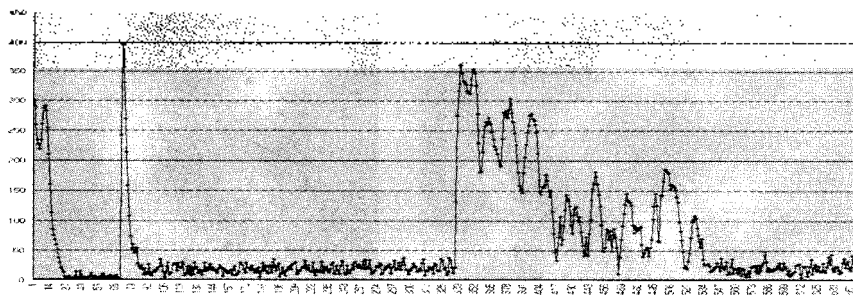
윈도우 인자 $\alpha(m)$ 은 아래의 과정을 통하여 계산되어진다.

$$\alpha(m) = 0.99 - 0.0245(50 - E_{loc}(m)) \quad (7)$$

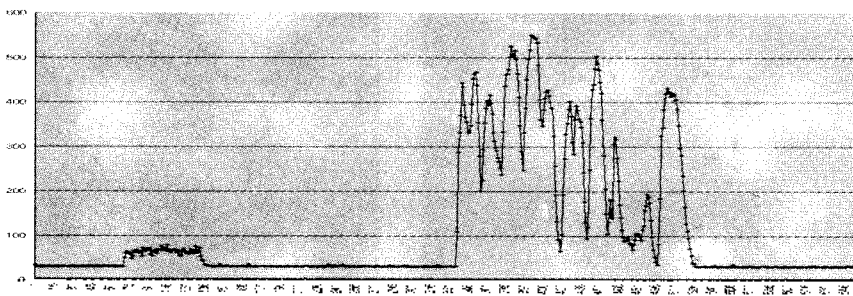
현재 프레임에서 각 채널에너지의 합 $E_{loc}(m)$ 은 다음과 같이 계산되어진다.



(a) 차량 잡음 섞인 음성



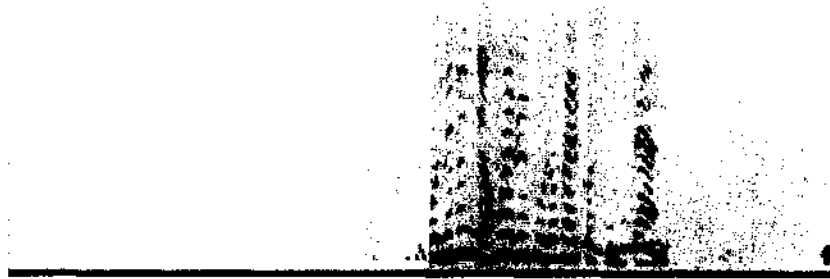
(b) 차량 잡음 섞인 음성 분산 변화도



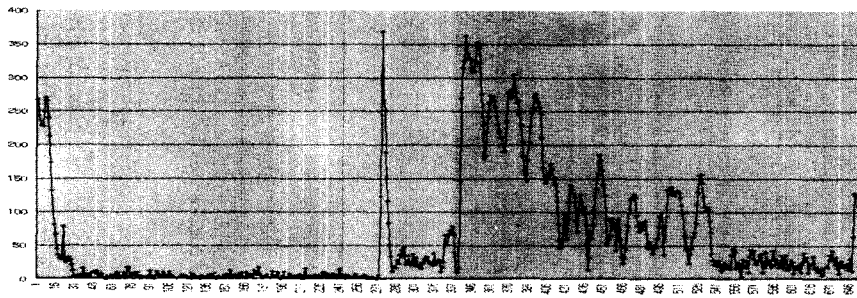
(c) 차량 잡음 섞인 음성 음성 매트릭 합 변화도

그림 1. 차량 잡음 섞인 음성 분석

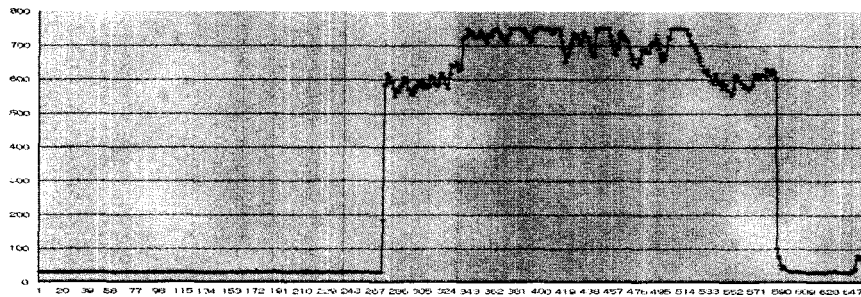
Fig. 1. Analysis of speech with car noise.



(a) Babble 잡음 섞인 음성



(b) Babble 잡음 섞인 음성 분산 변화도



(c) Babble 잡음 섞인 음성 voice metric 변화도

그림 2. Babble 잡음 섞인 음성 분석
Fig. 2. Analysis of speech with babble noise.

$$E_{voc}(m) = 10 \log_{10} \left[\sum_{i=0}^{N_c-1} E_{ch}(m, i) \right] \quad (8)$$

음성 신호잡음의 합, 음성 메트릭은 잡음 결정에 중요한 인자로 사용되며 잡음 제거 성능에 매우 중요한 역할을 한다. 잡음에서는 음성에 비해 에너지 레벨이 낮고 음성 메트릭도 낮은 값을 가진다. 또한 작은 값의 분산이 계속되는 특징을 가지고 있다. 잡음은 음성 미터 값이 일정한 값 약 50 정도 이하이면 음성의 에너지가 약한 것이기 때문에 잡음으로 간주된다. 또한 음성과 잡음이 혼합되어 있을 경우에는 비교적 적은 분산 값이 연속적으로 지속된다면 잡음으로 결정한다. 베이불 잡음은 시간에 따른 변화가 차량 잡음과 길거리 잡음과 비교할 때 변화가 심하였다. 베이불 잡음의 경우 잡음 세기가 커지면, 음성 미터 합은 물론 분산 값으로도 잡음 결정이 잘 되지 않는 경우가 발생한다. 차량 잡음과 베이불 잡음의 음성 메트릭과 신호대 잡음비 분산 값은 그림 1과 그림 2에 나

타나 있다. 차량 잡음의 음성 메트릭은 정확하게 음성의 특징 그대로 따라가고 있지만, 베이불 잡음이 섞인 음성의 경우는 잡음과 음성이 섞인 영역에서 정확하게 잡음 구별을 하지 못함을 볼 수 있다. 따라서 본 논문에서는 시간, 주파수 영역에서 분해능의 선택의 융통성을 갖는 웨이블릿 변환 (Wavelet Transform)을 이용하여 채널을 설계하며 스펙트럼 차감 잡음 제거기를 설계한다.

III. 웨이블릿 변환 (Wavelet Transform)을 이용한 잡음제거기 (Noise Suppressor)

잡음제거기의 입력신호는 고대역 통과필터를 이용한 직류성분이 제거된 신호이며, 입력에 사용된 프레임 크기는 분석프레임 크기 20 ms의 반인 10 ms 프레임 크기 단위로 수행된다. 따라서 20 ms 음성 프레임의 음성 부

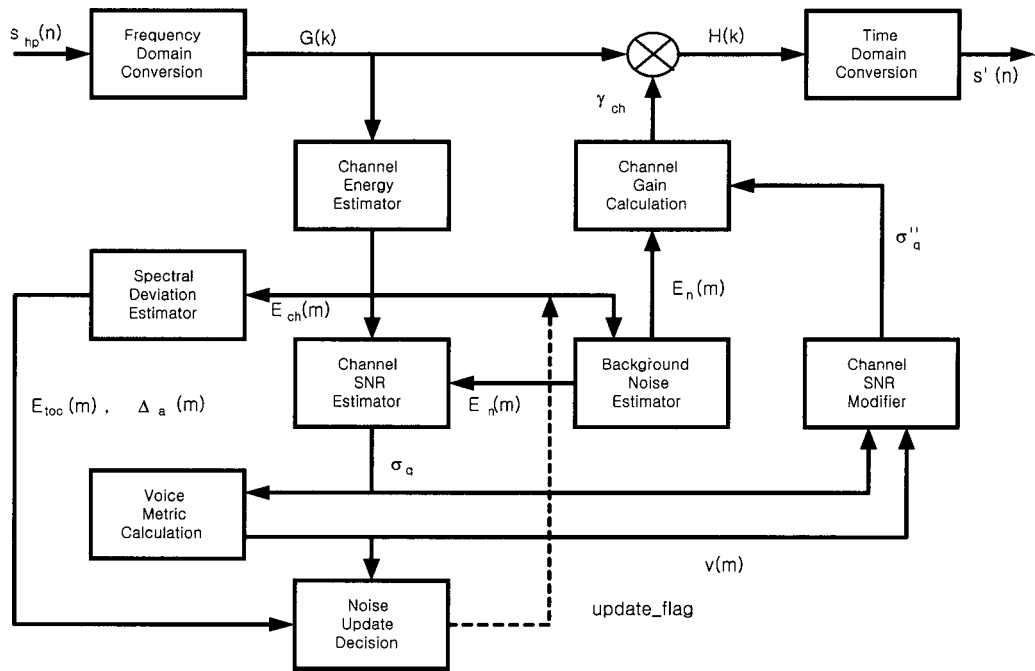


그림 3. 잡음제거기 블록도
Fig. 3. Block diagram of noise canceller.

호화시 잡음제거기는 두 번 실행된다. 잡음제거기의 구성도는 그림 3에 보여진다. 블록다이어그램에서도 볼 수 있듯이 잡음 제거기는 크게 웨이블릿 변환 부분과 채널의 에너지 계산 및 배경잡음 추정부분으로 구성되며 추정된 배경잡음 에너지 값을 사용하여 채널 이득 값을 재조정하게 된다.

3.1. 채널 설계

기존 방식은 푸리에 변환 후 주파수 영역을 임의로 나누어서 채널을 디자인하는 방식을 사용하는 반면에 본 논문에서는 웨이블릿 변환 후 영역을 2대역의 다해상도 특성이 생기도록 채널영역을 디자인한다. 신호를 분석/합성하는 웨이블릿은 길이가 4인 Daubechies 웨이블릿을 모웨이블릿 (mother wavelet)으로 사용하였다[5,6]. 웨이블릿 계수는 반대역 (Halfband) 필터뱅크를 사용하여 계산되어진다. 필터뱅크에서 사용될 임펄스 응답은 다음과 같이 주어진다.

$$h[n] = 0.5 \sin c \left(\left(n - \frac{N-1}{2} \right) \frac{\pi}{2} \right) \quad (9)$$

분석과정에서 저주파대역 필터의 임펄스응답 $h_0[n]$ 과 고주파대역 필터의 임펄스응답 $h_1[n]$ 은 다음과 같이 주어진다.

$$h_0[n] = h[n], \quad (10)$$

$$h_1[n] = (-1)^{(N-1-n)} h[N-1-n] \quad (11)$$

합성과정에서 저주파대역 필터의 임펄스응답 $g_0[n]$ 과 고주파대역 필터의 임펄스응답 $g_1[n]$ 은 다음과 같이 주어진다.

$$g_0[n] = 2h[N-1-n] \quad (12)$$

$$g_1[n] = 2(-1)^n h[n] \quad (13)$$

각 대역의 웨이블릿 계수 d_k 는 필터뱅크의 출력신호에서 얻어진다.

각 채널별 주파수해상도는 저주파성분으로 갈수록 조밀한 구조이며, 정의된 채널영역에 대해서 채널 에너지 및 배경잡음 분석이 이루어진다. 이러한 채널 구조는 웨이블릿 변환 후 계수 값들이 다해상도 특성을 가지도록 배열되는 특성을 이용한 것이다. 기존의 푸리에 변환과 웨이블릿을 변환에서의 다해상도 특성을 비교한 것이 그림 4, 그림 5에 나타나 있다. 그림 6에서도 볼 수 있듯이 고주파로 갈수록 웨이블릿의 채널은 시간해상도를 늘림으로써, 상대적으로 시간변이가 큰 성분들을 잘 표현할 수 있게 하였다.

구체적인 각 웨이블릿 계수에 대한 채널 구성도 일부를

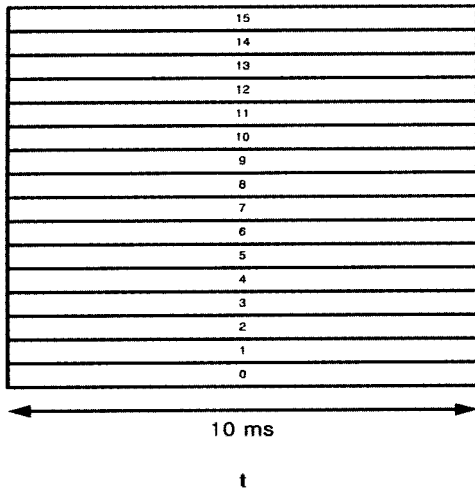


그림 4. STFT의 시간-주파수 해상도
Fig. 4. Time frequency resolution of STFT.

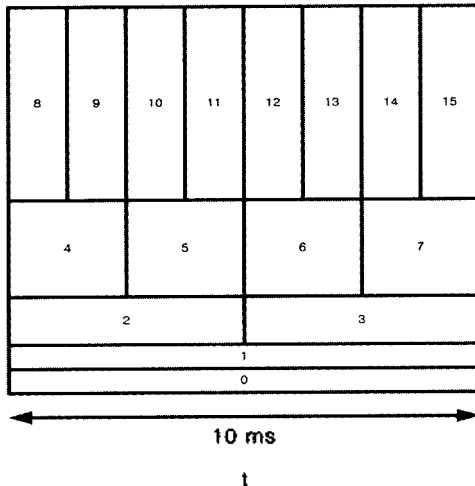


그림 5. 웨이블릿의 시간-주파수 해상도
Fig. 5. Time frequency resolution of wavelet.

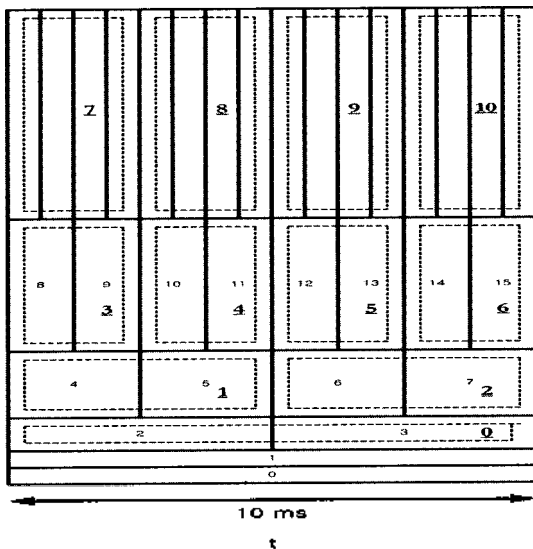


그림 6. 웨이블릿을 이용한 새로운 채널 디자인
Fig. 6. New channel design using wavelet.

그림 6에 나타내었으며, 저주파는 주파수 해상도를 고주파는 시간해상도를 강조한 채널로 설계하였다. 이것은 음성에 포함된 잡음성분들이 고주파로 갈수록 시간지역 특성을 보이기 때문이다. 그림에서 밑줄이 있는 숫자는 채널 밴드를 의미하고 점선은 채널영역을 뜻한다. 밑줄 없는 숫자는 웨이블릿 변환 후의 요소 값을 나타낸다. 채널 1과 2는 5 ms 단위로 분석이 되어지고, 나머지 채널들은 2.5 ms 단위로 분석이 되어진다. 균일한 해상도를 가지는 푸리에 변환과 구별되는 뚜렷한 특징이다. 이렇게 새롭게 디자인된 채널구조는 다음과 같이 주어진다.

$$f_L = \{2, 4, 6, 8, 10, 12, 14, 16, 20, 24, 28, 32, 40, 48, 56\}$$

$$f_H = \{3, 5, 7, 9, 11, 13, 15, 19, 23, 27, 31, 39, 47, 55, 63\}$$

f_L 은 채널의 최소 값, f_H 는 채널의 최대 값이다.

3.2. 연성 문턱값 (Soft Threshold) 적용

본 논문에 제안된 잡음 제거기에서는 스펙트럴 차분방법에 기반한 방법 이외의 연성 문턱값 (Soft Threshold) 방법에 의한 잡음제거 방법과 결합한 구조를 사용한다 [6,7]. 이 방법은 웨이블릿 변환 후 고주파 잡음을 제거하기 위한 것으로 전 단계로부터 추정된 배경잡음 추정 값을 이용하기보다는 현재 웨이블릿 계수 값을 문턱 값을 이용하여 그 크기 값을 재조정하는 원리로 되어 있다. 따라서, 잘못된 배경잡음 추정으로 제거되지 않는 고주파 잔여 잡음을 제거하는데 효과적이다. 변형된 웨이블릿 계수는 식 (14)와 같이 계산되어진다.

$$d_{jk} = \text{sign}(d_{jk})(|d_{jk}| - \lambda) \tag{14}$$

문턱값 λ 는 식 (15)와 같이 정의한다.

$$\lambda = \sqrt{2 \log(128)} \tag{15}$$

3.3. 장구간 파워 스펙트럼의 평균 및 분산 계산

장구간 평균 파워 스펙트럼 예측과정은 스펙트럼 분산 값을 계산하기 위한 전 단계로 장구간에 걸친 스펙트럼의 포곡선 값들을 예측하는 과정이다. 여기서, 장구간 평균 파워 스펙트럼 값이 추정된 후 계산되어지는 분산 값 및 음성 미터 값은 최종적으로 배경잡음 에너지 크기 값의 갱신 여부를 판단하는데 사용된다. 그 과정은 음성 미터 값의 크기가 작으면, 잡음이라 간주하고 스펙트럼 분산

값이 제한 값 이하인 경우 배경 잡음에너지 값을 갱신하는 구조로 되어 있다. 이러한 웨이블릿 계수를 이용한 채널은 주파수 대역 당 각기 다른 시간해상도를 갖기 때문에 대역당 계산되는 평균 파워스펙트럼 값의 계산구조는 변형되어야 하며, 시간해상도를 높이는 결과로 주파수 대역의 개수는 기존의 STFT 방법에 비해 줄어들게 된다. 각 주파수 대역별 장구간 평균 파워 스펙트럼 값의 계산은 고주파로 갈수록 블록내에 인접된 여러 개의 웨이블릿 계수가 많아지므로 이를 고려하여 다음 식 (16)과 같이 변형되어 계산되어진다. 여기서 장구간 평균 파워 스펙트럼 \overline{E}_{dB} 의 첫 번째 매개변수 $m+1$ 은 다음 프레임에서 사용될 값을 의미하고, 두 번째 매개변수는 채널 밴드를 의미한다.

$$\begin{aligned} \overline{E}_{dB}(m+1,0) &= \alpha(m) \overline{E}_{dB}(m,0) \\ &\quad + (1-\alpha(m))E_{dB}(m,0) \\ \overline{E}_{dB}(m+1,1) &= \alpha(m) \overline{E}_{dB}(m,1) \\ &\quad + (1-\alpha(m))(E_{dB}(m,1) + E_{dB}(m,2))/2 \\ \overline{E}_{dB}(m+1,2) &= \alpha(m) \overline{E}_{dB}(m,2) \\ &\quad + (1-\alpha(m))(E_{dB}(m,3) + E_{dB}(m,4) \\ &\quad + E_{dB}(m,5) + E_{dB}(m,6))/4 \\ \overline{E}_{dB}(m+1,3) &= \alpha(m) \overline{E}_{dB}(m,3) \\ &\quad + (1-\alpha(m))(E_{dB}(m,7) + E_{dB}(m,8) \\ &\quad + E_{dB}(m,9) + E_{dB}(m,10))/4 \\ \overline{E}_{dB}(m+1,4) &= \alpha(m) \overline{E}_{dB}(m,4) \\ &\quad + (1-\alpha(m))(E_{dB}(m,11) + E_{dB}(m,12) \\ &\quad + E_{dB}(m,13) + E_{dB}(m,14))/4 \end{aligned} \quad (16)$$

스펙트럼 분산 값은 평균 파워 스펙트럼과 현재 채널 파워 스펙트럼과의 차로 식 (17)과 같이 정의한다. 전체 프레임을 기준으로 평균 파워 스펙트럼 값을 구한 후 분산 값을 구한 이전 방식과 새로운 방식의 차이점이라면, 시간에 따른 분석 해상도를 높였다는 것이다.

$$\begin{aligned} \Delta_E(m) &= |\overline{E}_{db}(0) - E_{db}(0)| + |\overline{E}_{db}(1) - (E_{db}(1) + E_{db}(2))|/2 \\ &\quad + |\overline{E}_{db}(2) - (E_{db}(3) + E_{db}(4) + E_{db}(5) + E_{db}(6))|/4 \\ &\quad + |\overline{E}_{db}(3) - (E_{db}(7) + E_{db}(8) + E_{db}(9) + E_{db}(10))|/4 \\ &\quad + |\overline{E}_{db}(4) - (E_{db}(11) + E_{db}(12) + E_{db}(13) + E_{db}(14))|/4 \end{aligned} \quad (17)$$

3.4. 채널이득 계산 및 주파수 영역 여과

현재 프레임에 대한 전체 이득을 계산한다. 채널 전체 잡음값의 로그 값을 이용하여 이득을 계산한다.

$$\gamma_n = \max \left\{ \gamma_{\min}, 10 \log_{10} \left(\sum_{i=0}^{N_c-1} E_n(m, i) \right) \right\} \quad (18)$$

여기서 $\gamma_{\min} = -13$ (최소이득), $E_n(m, i)$ 잡음 스펙트럼 값을 나타낸다.

각 채널 이득 (channel gains)은 식 (19)를 이용하여 계산한다. σ_q'' 은 수정된 신호대잡음비 색인값이고, σ_{th} 는 기준 신호대잡음비이다.

$$\gamma_{dB}(i) = 0.39(\sigma_q''(i) - \sigma_{th}) + \gamma_n ; 0 \leq i < N_c \quad (19)$$

선형 채널 이득은 식 (20)을 이용하여 계산한다. 채널 이득은 0과 1사이의 값을 가지게 된다. 잡음으로 추정된 부분의 이득은 작은 값을 가진다.

$$\gamma_{ch}(i) = \min \{ 1, 10^{\gamma_{dB}(i)/20} \} ; 0 \leq i < N_c \quad (20)$$

채널 이득을 웨이블릿 변환 계수 d_{jk} 에 적용시킨다. 웨이블릿 변환 계수 d_{jk} 에 식 (21)과 같이 곱해서 잡음성분을 자연스럽게 제거한다.

$$\widehat{d}_{jk} = \gamma_{ch}(i) d_{jk} \quad (21)$$

잡음이 제거된 웨이블릿 변환계수는 역웨이블릿 변환을 이용해서 시간 영역 신호로 변환한다. 역웨이블릿 변환과정은 필터뱅크 방법을 사용하여 수행한다. 합성과정에서 사용되는 필터뱅크의 임펄스 응답은 식 (12), 식 (13)에 주어진다. 최종적으로 시간영역 신호는 중첩과 더하기 (overlap-and-add)를 적용시킴으로서 주파수 영역 여과 처리를 끝낸다. 중첩과 더하기 과정을 위한 프레임 크기는 그림 7에 나타나 있다.

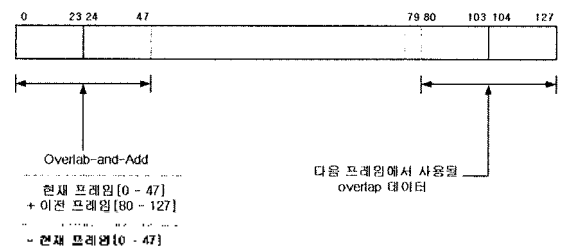


그림 7. 중첩과 더하기 과정을 통한 시간영역 복원
Fig. 7. Time domain reconstruction using overlap and add processing.

3.5. 주변 잡음 갱신

배경잡음의 갱신여부가 결정되면 각 대역당 배경잡음 에너지 크기 값들이 추정된다. 여기서, 각 대역별 추정되는 배경잡음의 크기 값들은 역시 전 단계의 대역별 배경잡음 에너지 값과 채널 에너지를 사용하여 계산되는데, 장구간 평균 파워 스펙트럼 값과 마찬가지로 α_n 에 의해 에너지 레벨의 변화정도가 조정된다. 추정된 대역별 배경잡음 에너지 값들은 채널 신호대잡음비와 채널 이득 값을 조정하는데 사용된다. 자세한 수식은 식 (22)와 같이 계산한다.

$$\begin{aligned}
 E_n(m+1, 0) &= \max \{ E_{\min}, \alpha_n E_n(m, 0) \\
 &\quad + (1 - \alpha_n) E_{ch}(m, 0) \} \\
 E_n(m+1, 1) &= \max \{ E_{\min}, \alpha_n E_n(m, 1) \\
 &\quad + (1 - \alpha_n) (E_{ch}(m, 1) + E_{ch}(m, 2)) / 2 \} \\
 E_n(m+1, 2) &= \max \{ E_{\min}, \alpha_n E_n(m, 2) \\
 &\quad + (1 - \alpha_n) (E_{ch}(m, 3) + E_{ch}(m, 4) \\
 &\quad + E_{ch}(m, 5) + E_{ch}(m, 6)) / 2 \} \\
 E_n(m+1, 3) &= \max \{ E_{\min}, \alpha_n E_n(m, 3) \\
 &\quad + (1 - \alpha_n) (E_{ch}(m, 7) + E_{ch}(m, 8) \\
 &\quad + E_{ch}(m, 9) + E_{ch}(m, 10)) / 2 \} \\
 E_n(m+1, 4) &= \max \{ E_{\min}, \alpha_n E_n(m, 4) \\
 &\quad + (1 - \alpha_n) (E_{ch}(m, 11) + E_{ch}(m, 12) \\
 &\quad + E_{ch}(m, 13) + E_{ch}(m, 14)) / 2 \} \quad (22)
 \end{aligned}$$

여기서, $E_{\min} = 0.0625$ (최소 채널 에너지 값) $\alpha_n = 0.9$ (잡음 상수)를 사용하였다.

IV. 웨이블릿 잡음 제거기의 성능 평가

새로 제안된 웨이블릿 변환 잡음제거기의 성능을 평가하기 위해 EVRC[4] 부호화기내의 잡음제거기와 비교하였다. EVRC 음성부호화기는 20 ms 프레임 단위로 음성 데이터를 처리하며, 잡음제거기는 그의 반인 10 ms 단위로 처리를 한다. EVRC 음성 부호화기에서 사용된 방식은 STFT를 이용한 스펙트럼 차감 잡음 제거 방법이다.

비공식 주관적 음질시험 (MOS 음질 시험)을 수행하였으며 실험에 사용된 음성 데이터는 남자 3명, 여자 3명의 음성 문장을 사용하였으며, 여기에 음성신호대 주변잡음 비가 15 dB, 10 dB, 5 dB의 크기를 갖는 베이بل 잡음, 차량 잡음, 길거리 잡음 등의 주변잡음을 섞은 문장에 대해 음질시험을 하였다. 여기서 사용된 베이블 잡음, 차량 잡음, 길거리 잡음은 이동 통신 환경에서 처리하여야 하는 대표적인 주변 잡음들이다. 음질시험에 훈련되지 않은 20명이 음질시험에 참가하였다. 음질시험 결과는 표 1에 나타내었다. 베이블 잡음에서는 제안된 잡음제거기 경우가 5 dB, 10 dB, 15 dB 모두 전체적으로 약 MOS 0.2 정도의 성능향상을 보였다. 이것은 제안된 웨이블릿 잡음제거기가 시간에 따라 잡음특성이 빠르게 변화하는 Babble 잡음의 경우 주파수 대역에 따라 시간해상도를 조절함으로써, 효과적으로 잡음을 제거할수 있다는 것을 보여준다.

차량잡음 및 길거리 잡음에 대해서는 15 dB, 5 dB의 경우 약 0.1 bB의 향상 또는 유사한 성능을 보였으나, 10 dB의 경우는 오히려 그 성능이 0.1~0.2 dB 떨어지는 경향도 관찰되었다. 결국, 제안된 잡음제거기에서 설계된 웨이블릿 패턴의 주파수 해상도가 STFT를 이용한 주파수 해상도 보다 고주파 대역으로 갈수록 떨어지기 때문에 오히려 성능이 저하된 경우도 나타나며, 이러한 부분을 개선하

표 1. MOS 시험 결과
Table 1. Result of MOS test.

| 음성 종류 | EVRC 방식 | 제안된 웨이블릿 변환 잡음제거기 |
|--------------------|---------|-------------------|
| 원음 | 4.15 | 4.26 |
| Babble 잡음 SNR 15dB | 3.71 | 3.92 |
| Babble 잡음 SNR 10dB | 3.42 | 3.63 |
| Babble 잡음 SNR 5dB | 3.10 | 3.26 |
| 길거리 잡음 SNR 15dB | 3.60 | 3.74 |
| 길거리 잡음 SNR 10dB | 3.47 | 3.35 |
| 길거리 잡음 SNR 5dB | 3.07 | 2.96 |
| 차량 잡음 SNR 15dB | 3.96 | 4.04 |
| 차량 잡음 SNR 10dB | 3.82 | 3.64 |
| 차량 잡음 SNR 5dB | 3.26 | 3.29 |

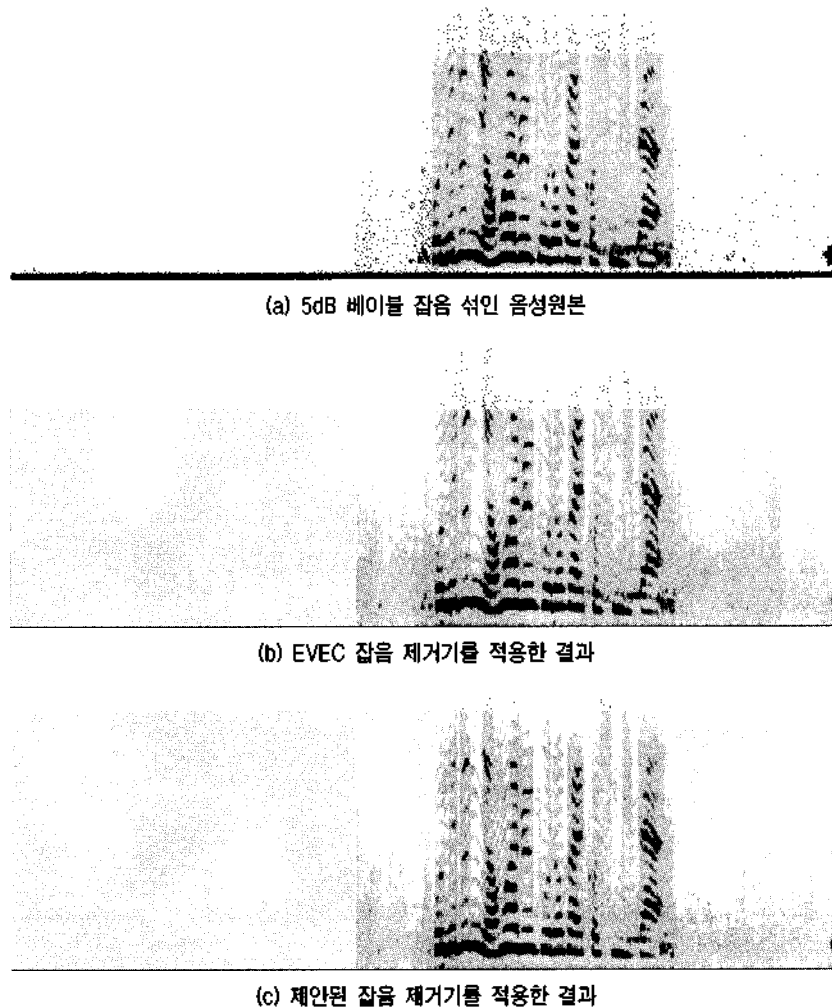


그림 8. 차량 잡음 섞인 음성 합성 결과
Fig. 8. Synthesis results of speech with car noise.

기 위해 웨이블릿의 패턴을 잡음의 각 대역별 에너지에 따라 변화하는 방법이 현재 더 연구되어야 할 부분으로 남아있다.

제안된 잡음제거기에서 설계된 웨이블릿의 패턴이 그림 8은 베이블 잡음에서의 EVEC 방식과 제안된 웨이블릿 변환 잡음제거기 결과 파형을 나타낸다. 그림에서 볼 수 있듯이 하모닉 스펙트럼이 왜곡없이 유지되어 같은 음색을 유지하는 것이 관찰되었으며, 묵음 부분 및 음성부분 하모닉 스펙트럼 라인 (Harmonic Spectrum Line) 사이의 잡음 스펙트럼이 효과적으로 제거되었음을 볼 수 있었다.

V. 결론

본 논문에서는 신호를 분석/합성하는 도구인 웨이블릿 변환을 이용해서 새로운 잡음 제거 알고리즘을 제안하였

다. 웨이블릿 변환을 이용한 스펙트럼 차감 잡음 제거기를 설계하기 위해 새로운 채널 설계, 주변잡음 갱신 방법, 웨이블릿 계수 차감법 등을 제안하였다. 시간에 특성이 빨리 변하는 베이블 잡음의 경우 일정한 시간/주파수 분해능을 갖는 STFT 분석을 이용한 기존의 EVEC 방식은 잡음 결정의 정확성이 떨어지는 것을 볼 수 있었다. 새롭게 제안된 잡음제거기는 시간/주파수 분해능을 융통성 있게 선택할 수 있는 웨이블릿 변환을 이용함으로써 잡음 결정의 정확성을 증가시킬 수 있었다. 제안된 웨이블릿 변환 잡음제거기는 음질시험을 통해 거리잡음, 자동차잡음에서는 EVEC 방식과 비슷한 성능을 보였고 베이블 잡음에서 MOS 0.2 정도의 음질 개선을 보였다. 음성의 스펙트로그램을 통해서도 잡음제거 성능 개선을 확인할 수 있었다.

감사의 글

본 연구는 2000년 정보통신부의 대학기초연구지원사업 연구비에 의하여 연구되었음.

참고 문헌

1. J. E. Natvig, "Evaluation of six medium bit-rate codecs for pan-european digital mobile radio system," *IEEE Journal on Selected Areas in Commun.*, Vol. 6, pp. 346-352, June 1998.
2. I. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP) speech coding at a 8 kbps," *Proc. IEEE Int. Conf. Acoust., speech, Sig. Processing*, pp. 461-464, Albuquerque, New Mexico, USA, April 1990.
3. R. Salami, L. Adoul, S. Hayasi, P. Kroon, and Y. Shoham, "Description of the proposed ITU-T 8 KB/S speech coding standard," *Proc. IEEE Workshop on Speech Coding for Telecommunication*, pp. 3-4, Annapolis, Maryland, USA, September 1995.
4. PN-3292 (IS-127), "Enhanced Variable Rate Codec, Speech service option 3 for wideband spread spectrum communication systems," April 1996.
5. C. Sidney Burrus, Ramesh A., "Introduction to wavelets and wavelet transforms," *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, pp. 98-99, 1998.

6. Wim Sweldens, "Wavelets: what next?," *Proceedings of the IEEE*, 84(4), pp. 680-685, April 1996.
7. Donoho, D., Johnstone, I., Kerkyacharian, G, and Picard, D., "Density estimation by wavelet thresholding," *Technical Report, Department of Statistics, Stanford University*, 1993.

저자 약력

● 원 호 진 (Ho-jin Won)

1997년 2월: 충북대학교 전자공학과 학사
2000년 2월: 충북대학교 전파공학과 석사
2000년 2월~2001년 5월: 현대 전자
2001년 6월~현재: LG 전자
※ 주관심분야: 음성 오디오 부호화, 음성인식, 적응필터

● 김 종 학 (Jonghark Kim)

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

● 이 인 성 (Insung Lee)

1983년 2월: 연세대학교 전자공학과 학사
1985년 2월: 연세대학교 전자공학과 석사
1992년 12월: Texas A&M University 전기공학과 박사
1986년 5월~1987년 7월: 한국통신 연구개발단 전임연구원
1993년 2월~1995년 9월: 한국전자통신연구원 이동통신기술연구단 선임연구원
1995년~현재: 충북대 전기전자공학부 부교수
※ 주관심분야: 음성 및 영상신호압축, 이동통신, 적응필터