

신호 에너지의 지각적인 간섭현상에 의한 컨벌루션 노이즈의 해석에 관한 연구

김헌중, 한헌수, 홍민철, 차형태
숭실대학교 전자공학과

The Study for Perceptual Convolutional Noise by the Masking Effect of Signal

Hun-joong Kim, Hern-soo Hahn, Min-cheoul Hong, Hyung-tai Cha

Dept. Of Electronics Engineering, Soongsil Univ.

Email : hjkim@mms.ssu.ac.kr

요 약

본 논문에서는 잡음 환경에서의 사람의 청각 시스템에서의 신호 에너지에 대한 지각(perception) 패턴을 나타내는 임계 대역에서의 여기(excitation)에너지와 이러한 신호 에너지 사이의 지각적인 간섭 효과를 나타내는 마스킹 특성을 이용하여 지각적인 컨벌루션 잡음(convolutional noise)해석을 통한 잡음 제어 기법을 연구 하였다. 이러한 지각적 컨벌루션 잡음은 지각 영역에서 잡음 에너지에 의한 간섭 현상의 영향을 나타내며, 신호의 지각적인 특성 변화 및 부가 잡음(additional noise) 제어 후에 신호 내에 잔존하는 잔여 노이즈를 반영하는 것이다.

실험 결과 부가 잡음 제어 필터와 컨벌루션 잡음 제어 필터가 대등한 노이즈 제어 성능을 나타내었으며, 다양한 잡음 환경에서 이러한 지각 영역에서의 부가 잡음 제어와 컨벌루션 잡음 제어를 통해 높은 SNR비 개선과 더불어 MOS 측정을 통해서도 높은 음질을 보장할 수 있음을 확인 할 수 있었다.

1. 서 론

최근 시스템 환경의 발전과 다양한 고품질의 음성 및 오디오 신호의 디지털 표현기술의 발전으로 인해 정확한 의미 전달의 관점에서의 신호의 명료도(intelligibility) 개선만으로는 음질 측면에서 늘어나는 사용자의 요구 조건을 충족 시키기에는 부족하게 되었다. 특히 오디오 신호의 경우 다양한 요인으로 인한 신호 품질의 열화에 대한 SNR 측면에서의 개선뿐만 아니라 이러한 음질 측면에서의 신호 개선은 중요한 관건이 된다.

그 결과 최근 잡음에 오염된 신호의 개선을 위한 알고리즘에 있어 높은 SNR비 개선과 더불어 신호에 대한 지각적인 측면에서도 무결성을 보장하기 위해, 인간의 청각적 지각 특성을 이용하여 잡음에 의한 영향을 최소화 하기 위한 알고리즘들이 제안되었다[1][2][3][4].

본 논문에서는 단일 채널 기반의 환경에서 신호 개선을 잡음 신호에 대한 사람의 청각 시스템에서의 지각 패턴을 이용한 주파수 차감 기법 기반의 필터링을 통해 다양한 환경에서 음질 개선을 위한 연구를 수행 하였다.

이것은 사람의 청각 시스템에서의 신호에 대한 지각 패턴을 나타내는 임계 대역에서의 여기 에너지에 기반을 둔 것으로 묵음 구간을 통해 추정된 잡음 신호의 임계 대역에서의 여기 에너지를 이용한 Perceptual Filtering을 통한 부가 잡음(additional noise) 에너지의 제어와 noise unpredictability measure를 통한 임계 대역에서의 추정 잡음의 오차에 의한 신호 특성 변화 및 신호 내에 잔존하는 잔여 잡음 에너지를 반영하는 지각적인 컨벌루션 잡음(convolutional noise) 에너지의 제어를 통해 신호를 개선하는 방법이다.

2. Psychoacoustical Representation of Signals

사람의 청각 시스템에서의 주파수 변별력(frequency selectivity)을 나타내는 임계 대역(critical band)에서의 임의의 신호의 파워 스펙트럼 $F_s(\omega; t)$ 에 대한 critical band intensity는 다음과 같이 계산 할 수 있다.

$$F_s^l(z; t) = a_0(z) \sum_{\omega=\omega_l}^{\omega_k} F_s(\omega; t) \quad 0 \leq z \leq Z-1$$

where, z : critical band index (in Bark)

ω_l, ω_k : the lower and upper bounds

of the critical band z

Z : the total number of critical band

(1)

이때, t 는 길이 N 을 갖는 임의의 윈도우를 통해 구분되어지는 프레임 인덱스를 나타내며, $a_0(z)$ 는 외이(outer ear)에서부터 중이(middle ear)까지의 다양한 전달 특성 요소(transmission factor)들에 의한 주파수

의존 감쇄특성을 나타내는 감쇄 성분이다. 이때, 이러한 $F^l(z, t)$ 에 대한 기저막(basilar membrane)에서의 신호의 에너지 확산현상을 나타내는 에너지 확산함수(spreading function)을 통해 컨벌루션을 취함으로써, 다음과 같이, 기저막에서의 신호에너지에 대한 주파수 응답을 나타내는 임계 대역에서의 여기 에너지를 표현할 수 있다

$$F_s^e(z; t) = \sum_{v=0}^{z-1} s_2(v; z-v) F_s^l(v; t) + \sum_{v=z}^{Z-1} s_1(v-z) F_s^l(v; t) \quad (2)$$

이때 backward spreading을 나타내는 $s_1(v-z)$ 과 forward spreading을 나타내는 $s_2(v, z-v)$ 을 다음과 같은 단일 함수를 이용하여 표현하면

$$ss(v, z) = \begin{cases} s_2(v, z-v), & v < z \\ s_1(v, v-z), & v \geq z \end{cases} \quad (3)$$

임계대역 내에서의 신호에 대한 excitation에너지는 다음과 같이 표현 할 수가 있다.

$$F_s^e(z; t) = \sum_{v=0}^{z-1} [ss(v; z) F_s^l(v; t)] \quad (4)$$

또한, 이러한 임계대역 내에서의 excitation은 시간 영역에서의 에너지 확산의 영향 또는 post masking spreading을 고려 하여 이전 프레임과 결합되어 지게 되는데, 이 과정을 다음과 같이 표현할 수 있다.

$$F^{et}(z; t) = F^e(z; t) + \sum_{m=0}^{t-1} [T_f^{t-m}(z) F^{et}(z; m)] \quad (5)$$

where, $0 \leq z \leq Z-1$

여기에서 $T_f(z)$ 는 신호 에너지의 시간에 따른 마스크 확산의 영향을 표현하기 위한 임의의 함수이다[5][6].

3. Perceptual Additional Noise

이때 이러한 임의의 프레임 인덱스 t 에서의 잡음에 오염된 신호의 전력 스펙트럼을 $F_s(\omega; t)$, 순수 신호의 전력 스펙트럼을 $F_y(\omega; t)$ 라고 하고, 각각의 신호에 대한 청각 시스템에서의 지각 패턴에 대한 신호의 표현(psychoacoustic representation)을 각각 $F_y^e(z; t)$, $F_s^e(z; t)$ 라고 한다면, 이러한 두개의 신호 표현 사이의 차이는 지각적으로 인식하게 되는 가청 영역에서의 잡음의 양을 나타내게 된다. 여기서 개선된 신호의 전력 스펙트럼이 $\hat{F}_s(\omega; t)$ 이고, 그에 따른 지각 영역에서의 신호 표현을 $\hat{F}_s^e(z; t)$ 라고 하고, 임의의 필터 $H(z; t)$ 가 가청 가능한 잡음의 영향을 제거하여 순수 신호와 지각적으로 일치 하는 개선된 신호를 생성한다면 다음과 같은

관계식이 성립하게 되며,

$$F_s^{et}(z; t) \approx \hat{F}_s^e(z; t) = H(z, t) F_y^e(z; t) \quad (6)$$

다음과 같이 time/frequency perception-dependent filter 가 존재 하게 된다.

$$H(z; t) = \frac{F_s^{et}(z; t)}{F_y^e(z; t)}, \quad 0 \leq z \leq Z-1 \quad (7)$$

이러한 단일 채널에서의 잡음에 오염된 신호와 목음 구간동안 추정된 잡음, $\hat{F}_d^e(z; t)$ 을 통한 지각 적인 영역에서의 부가잡음(additional noise)을 제어하기 위한 주파수 영역에서의 필터는 다음과 같이 나타낼 수 있다. 단 이때, $T_f(z) = 0$ ($0 \leq z \leq Z-1$), no-time memory situation이라고 가정한다[4].

$$H^e(z; t) = 1 - \frac{\hat{F}_d^e(z; t)}{F_y^e(z; t)} \quad (8)$$

4. Perceptual Convolutional Noise

잡음의 추정에 있어, 해당 프레임에 첨가된 잡음 신호를 정확히 추정할 수 있다면, 개선된 신호의 경우 원 신호와 같은 결과를 얻을 수 있을 것이다. 그러나, 일반적으로 잡음 신호를 정확하게 추정하는 일은 어려운 일 이므로, 추정 오차가 존재하게 되는데, 이러한 추정 오차에 의한 잡음에 오염된 신호 내에서의 영향을 보상해 줄 필요가 있다.

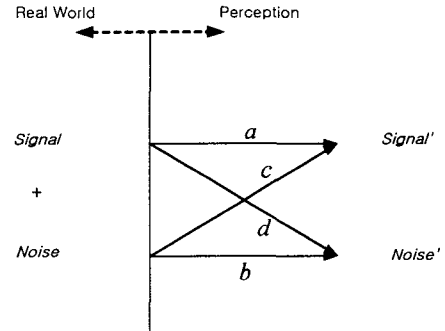


그림 1. 잡음의 지각 패턴 해석을 위한 개념도

그림 1.은 이러한 지각적인 영역에서의 신호 상호간의 영향을 해석하기 위한 개념도 이다. 그림에서 a와 b 는 지각 적인 영역에서의 잡음과 신호에 대한 부가 에너지를 나타내고, c와 d를 신호 상호간의 간섭 현상의 영향으로 해석 할 수 있다. 이때 d는 신호 에너지에 의한 잡음에 대한 간섭, 즉 마스크 현상으로 해석할 수 있으며, c의 경우는 잡음 에너지에 의한 신호 에너지에 대한 영향을 반영하는 것으로, 부가 잡음 에너지의 제거 후 존재하게 되는 잔여 잡음과 잡음 에너지에 의한

신호의 순음(tonal)성분과 비순음(non-tonal)성분의 변화 등을 나타내는 컨벌루션 잡음의 영향을 나타낸다.

본 논문에서는 현재 프레임에 대한 잡음 추정 오차의 영향을 묶음 구간의 잡음 신호 프레임을 기준으로 현재 잡음에 오염된 프레임에 대한 unpredictability measure를 계산하여 지각적인 영역에서의 잡음 에너지의 영향을 계산하였다.

$$C(\omega; t) = \frac{|S_y(\omega; t)e^{j\hat{\varphi}_y(\omega; t)} - \hat{S}_d(\omega; t)e^{j\hat{\varphi}_d(\omega; t)}|}{|S_y(\omega; t)| + |\hat{S}_d(\omega; t)|}, \quad (9)$$

where, $0 \leq C(\omega; t) \leq 1$

이때 $S_y(\omega; t)$ 과 $\hat{S}_d(\omega; t)$ 는 잡음에 오염된 신호와 추정 잡음의 진폭 스펙트럼을, $\varphi_y(\omega; t)$ 과 $\hat{\varphi}_d(\omega; t)$ 각각에 대한 위상 스펙트럼을 나타낸다.

이렇게 계산 되어진 noise unpredictability measure값은 다음과 같이 지각적인 영역에서의 그 영향을 반영하기 위해 잡음에 오염된 신호에 대해 가중치로써 적용 되어진 후, 임계 대역에서의 여기 에너지를 계산 하게 된다.

$$F_y^{cl}(z; t) = a_0(z) \cdot \sum_{\omega=\omega_{lc}}^{\omega_{hc}} C(\omega; t) F_y(\omega; t) \quad (10)$$

where, $0 \leq z \leq Z-1$

$$F_y^{ce}(z; t) = \sum_{v=0}^{z-1} [ss(v, z) F^{cl}(v; t)] \quad (11)$$

이와 같이 noise unpredictability measure를 통해 계산된 잡음 추정 오차와 잡음 에너지에 의한 신호 에너지의 특성 변화에 따른 잡음에 오염된 신호에 대한 임계 대역 내에서 여기 에너지의 변화에 대한 응답은 다음과 같이 잡음에 오염된 신호의 여기 에너지로 정규화를 통해 구해지게 된다.

$$G(z; t) = \frac{F_y^{ce}(z; t)}{F_y^e(z; t)} \quad \text{where, } 0 \leq G(z; t) \leq 1 \quad (12)$$

5. 실험 및 결론

실험에 사용된 신호들은 일반 오디오 CD로부터 획득한 44.1kHz의 샘플링율을 갖고, 16bit/sample의 정밀도를 갖는 다양한 장르의 오디오 신호들을 이용하였다. 이때 사용된 프레임은 1024 크기의 overlap-add hanning window를 적용한 프레임이며, 각각의 필터링의 개선 성능을 비교 하였으며, 이때 잡음에 오염된 신호에 대한 사람의 청각 시스템에서의 지각 패턴 제어를 통한 부가 잡음과 컨벌루션 잡음에 대한 음질 개선 시스템을 다음과 같이 구성하여 다양한 잡음 환경에서 성능 분석을 하였다.

$$\hat{F}_y^l(\omega; t) = G(z; t) \cdot H^p(z; t) \cdot F_y^l(\omega; t), \quad (13)$$

where, $\omega_{lc} \leq \omega \leq \omega_{hc}, \quad 0 \leq z \leq Z-1$

이때 musical tone noise 제어는, 각각의 임계 대역에 대한 필터 응답을 통한 extrapolation을 취하는 simplest intra-frame prediction을 적용하였다.

$$H^p(z; t) = H^p[SNR_{pmt}^p(z; t)], \quad (0 \leq H^p(z; t) \leq 1)$$

$$= \begin{cases} \left(1 - \frac{\hat{F}_d^e(z; t)}{F_y^e(z; t)}\right), & (0 \leq z \leq Z-1) \\ 2 * H^p(z; t-1) - H^p(z; t-2), & \text{if } \hat{F}_d^e(z; t) > F_y^e(z; t). \end{cases} \quad (14)$$

이와 같이 오디오 신호의 경우에서 처럼 인접 프레임 간 변이(variance)가 높고, 상대적으로 긴 신호 구간으로 인한 에너지 변화율이 높은 경우, 직접적인 임계 대역 내의 신호 에너지를 통한 제어 보다, $SNR_{pmt}^p(z; t)$ 를 통한 신호 내 잡음 에너지의 영향의 변화를 반영하는 것이 효율적이다.

이때 식. 13은 파라미터 설정을 위해 다음과 같이 구성할 수 있으며, 본 실험에서는 $\alpha = 5$ 를 사용하였다.

$$\hat{S}_y(\omega; t) = S_y(\omega; t) \cdot \left(\frac{F_y^{ce}(z; t)}{F_y^e(z; t)}\right)^{\frac{\alpha}{2}} \cdot \left(1 - \frac{\hat{F}_d^e(z; t)}{F_y^e(z; t)}\right)^{\frac{\alpha}{2}} \quad (15)$$

where, $\omega_{lc} \leq \omega \leq \omega_{hc}, \quad 0 \leq z \leq Z-1$

실험은 잡음에 의해 SNR=20dB로 열화 시킨 오디오 신호를 사용하여 Segmental SNR과 MOS 테스트 결과를 통해 비교하였다.

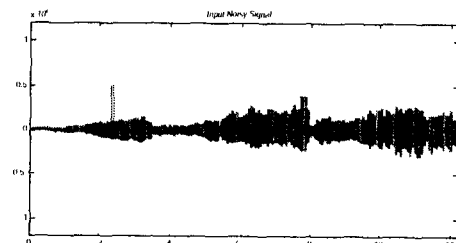


그림 2. 잡음에 오염된 신호 (SNR=20 dB)

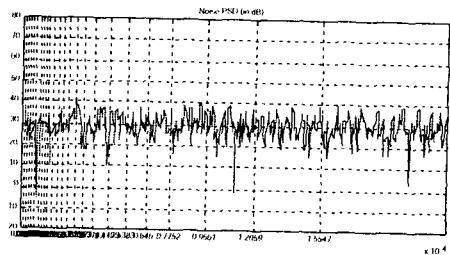


그림 3. 화이트 가우시안 잡음

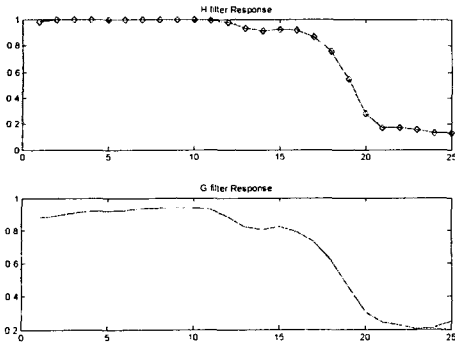


그림 4. 필터 응답

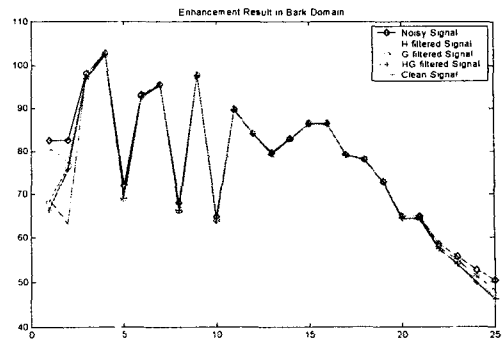


그림 7. 개선 결과(engine noise)

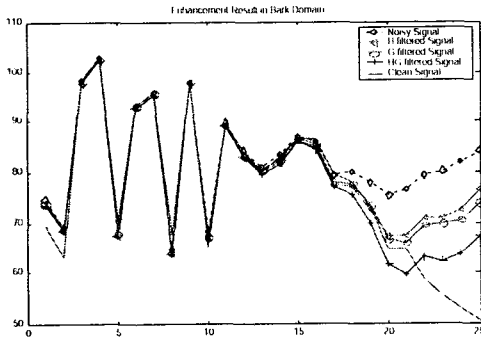


그림 5. 개선 결과 (AWGN)

그림 2.는 잡음에 오염된 오디오 신호를 나타내며, 표시된 구간에서 그림 3.에서와 같은 잡음에 오염되었을 경우에 대한 각각의 임계 대역에서의 H 필터와 G 필터의 응답은 그림 4.에서와 같다. 그림에서 살펴 볼 수 있듯이 지각적인 영역에서의 부가 잡음 제어 필터(H -filter)와 컨벌루션 잡음 제어 필터(G -filter)가 대등한 성능을 나타냄을 알 수 있다. 또한 그림 5.는 이러한 필터들을 이용한 Bark Domain에서의 개선 결과를 나타내고 있는데, 그림 2.의 표시된 구간의 SNR이 17.58 dB일 경우, 이러한 HG -filter를 통해 SNR을 23.37 dB로 개선 시킴으로써 5.79 dB의 개선 효과를 볼 수 있었으며, 전체 프레임 평균 6.5 dB의 개선 결과를 나타내었다. 이때 H -filter와 G -filter의 개선 성능은 평균적으로 동등한 결과를 나타내었다.

다음 그림 6.은 다양한 잡음 환경에서의 결과를 비교하기 위해 첨가된 engine 잡음의 주파수 특성을 나타내며, 그림 7.은 개선 결과를 나타낸다. 실험 결과 앞서 실험한 AWGN의 경우와 같은 개선 성능을 나타내어 다양한 잡음 환경에 적용할 수 있음을 알 수 있었다.

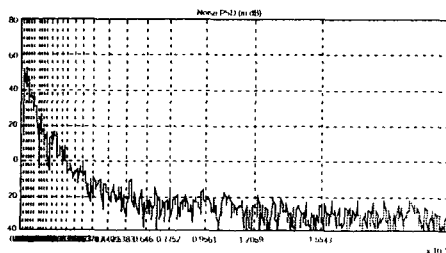


그림 6. engine noise

또한 MOS실험 결과 지각적인 컨벌루션 잡음 처리를 추가 수행함에 따라 지각적인 부가 잡음에 대한 단일 처리일 때보다 평균 0.8의 높은 개선 효과를 보임으로써 지각적인 컨벌루션 잡음 제어를 통해 SNR개선뿐만 아니라 잡음 에너지에 의한 지각적인 영역에서의 신호의 지각 특성 변화를 반영함으로써 높은 음질 개선 효과를 얻을 수 있음을 확인 할 수 있었다. 또한 잡음 추정 알고리즘의 개선과 최적 파라미터 설정을 통한 성능 개선을 기대 할 수 있다.

참고 문헌

1. Johnston J., "Transform Coding of Audio Signals Using Perceptual Noise Criteria," IEEE Journal of Selected Areas in Communications, Vol. 6, No. 2, Feb., 1988.
2. A. Czyzewski, R. Królikowski, "Noise Reduction in Audio Signals Based on the Perceptual Coding Approach," Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 17-20.10, pp. 147-150, USA, NY, Oct., 1999.
3. Nathalie Virag "Single Channel Speech Enhancement Based on Masking Properties of the Human Auditory System," IEEE Trans. Acoustic., Speech, Signal Processing, Vol. 7 N.2, Mar., 1999.
4. Dionysis E. Tsoukalas, John Mourjopoulos, George Kokkinakis, "Perceptual Filter for Audio Signal Enhancement," J. Audio Eng. Soc. Vol. 45 No.1/2, Jan., 1997.
5. E. Zwicker, H. Fastl, *Psychoacoustics : Facts and Models*, Springer 2nd Edition, 1999.
6. C, J. Moore, *Hearing*, Academic Press, 1995