

마스킹 특성을 이용한 Inter-band Smoothing 기법을 통한 음질 개선에 관한 연구

엄혜영, 김현중, 차형태
 숭실대학교 전자공학과

The Study for Quality Enhancement by Inter-Band Smoothing Using Masking Effect Analysis

Hye-yeong Eom, Hun-joong Kim, Hyung-tai Cha
 Dept. Of Electronics Engineering, Soongsil Univ.
 Email : hye0@mms.ssu.ac.kr

요약

본 논문에서는 신호의 마스킹 특성을 이용한 Inter-Band Smoothing 기법을 통한 음질 개선 연구를 하였다. 이것은 임계 대역 신호 처리 후 신호의 전체 신호 대 마스킹 비를 보상해 주는 방식으로, 각 대역의 여기(Excitation) 에너지로부터 근사화 시킨 전역 마스킹 임계치를 통한 보상 기법이다. 잡음 에너지 제어를 통한 음질 개선 알고리즘을 이용한 실험 결과 보다 자연스러운 음질의 제공을 통해 MOS의 개선을 확인 할 수 있었다.

1. 서론

최근 사람의 청각 특성 모델을 이용한 다양한 신호 처리 기법의 연구가 활발히 진행 되어지고 있다. [1][2][3][4]. 이것은 시스템 환경의 발전과 다양한 고품질의 오디오 신호의 디지털 표현 기술의 발전에 따라 사용자의 요구 조건을 충족시키기에는 단순한 의미 전달의 관점에서의 명료도 개선만으로는 부족하게 되었음을 의미한다. 그러나 일반적으로 이와 같은 청각 모델을 기반의 Bark Domain에서의 신호 처리들은 신호 처리 후 Frequency Domain으로의 확장이 요구 되어지게 된다. 그러나 이때 임계 대역의 신호 에너지는 각 Bark ratio에 대응하는 주파수 영역의 에너지 특성을 제한 하게 되므로 각 대역의 경계 면에서 Discontinuity 문제가 발생 하게 되고 대역 내에 에너지가 평균에너지 레벨을 기준으로 분포하게 되는 문제가 발생한다.

본 논문에서는 이러한 음질 영화의 원인을 제공하는 Inter-band discontinuity 문제를 해결하기 위해, 각 임계대역 신호 에너지의 에너지 확산영향을 고려한 전역 마스킹 특성 기반의 Inter-band smoothing 기법을 제안한다.

2. Excitation Energy

사람의 청각 시스템에서의 주파수 변별력을 나타내는 임계대역에서의 임의의 신호의 파워 스펙트럼(Power Spectrum), $F(\omega; t)$ 에 대한 Critical Band Intensity를 다음

과 같이 계산할 수 있다.[5][6]

$$F'(z; t) = a_0(z) \sum_{\omega=k_{lz}}^{k_{uz}} F(\omega; t) \quad 0 \leq z \leq Z-1$$

where, z : critical band index (in Bark)
 k_{lz}, k_{uz} : the lower and upper bounds of the critical band z
 Z : the total number of critical band

이때, t 는 N 의 길이를 갖는 임의의 윈도우를 통해 구분되어지는 프레임 인덱스를 나타내며, a_0 는 위이에서 중이까지의 다양한 전달 특성 요소들에 의한 주파수 의존 감쇄 성분이다. 또한, 이러한 $F'(z; t)$ 에 대한 기저막에서의 신호의 에너지 확산현상을 나타내는 확산함수(Spreading Function)을 통해 컨벌루션(Convolution)을 취함으로써 기저막에서의 신호 에너지에 대한 주파수 응답을 나타내는 임계대역에서의 여기 에너지를 표현할 수 있다.

$$F^e(z; t) = \sum_{v=0}^{z-1} s_2(v; z-v) F'(v; t) + \sum_{v=z}^{Z-1} s_1(v; v-z) F'(v; t)$$

이때, $s_1(v, v-z)$ 와 $s_2(v, z-v)$ 는 각각 역 방향 확산함수(Backward Spreading)와 전방향 확산함수(Forward Spreading)를 나타내며 다음과 같은 단일 함수를 이용하여 표현하도록 하겠다.

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

이와 같은 단일함수를 이용하여 임계대역에서의 여기 에너지를 다시 표현하면 다음과 같다.

$$F^e(z; t) = \sum_{v=0}^{Z-1} [ss(v; z) F'(v; t)]$$

이와 같이 Bark Domain에서의 Excitation Energy, 식 (4) 또는, Critical Band Intensity, 식 (1)를 이용한 임의의 신호 처리 과정 후에는 그 결과를 Frequency Domain의 각각의 주파수 성분들에 대해 적용을 해주어야 한다. 이때 Bark Domain에서 Frequency Domain으로의 Spreading 과정을 다음과 같이 표현할 수 있다. 이때, $\langle \omega_{hz} \leftrightarrow \omega_{lz} \rangle_{bins}$ 은 각 임계대역 내에서의 주파수 성분들의 수를 나타낸다.

$$\tilde{F}(\omega; t) = F'(z; t) \frac{1}{\langle \omega_{hz} \leftrightarrow \omega_{lz} \rangle_{bins}} \quad (5)$$

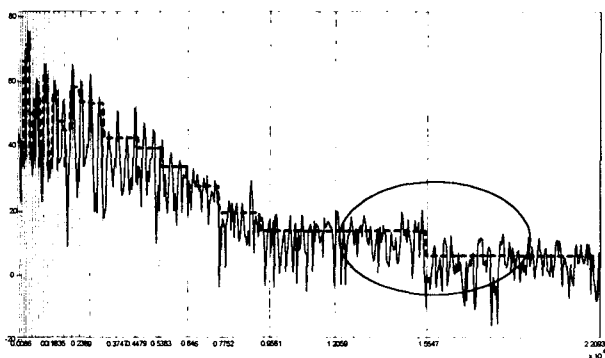


그림 1. 주파수 영역으로 펼쳐진 여기에너지

그러나 이와 같이 Bark Domain에서 Frequency Domain으로의 확장 과정을 통해 다음 그림 1.에서와 같이 각 임계대역 경계 부근에서 Discontinuity 문제가 발생하게 되어 음질 열화를 가져오게 된다. 이때 이러한 Inter-Band Discontinuity 문제를 해결하기 위한 Smoothing 처리 방법으로는 다양한 방법들이 적용되어질 수 있다.

본 논문에서는 이와 같은 Inter-Band Smoothing 처리 후에 신호의 SMR(Signal-to-Mask Ratio)등을 최소화하고, 신호의 자연스러운 음질을 보상하기 위해 신호의 Global Masking Threshold 기반의 Inter-Band Smoothing 기법을 제안한다.

3. Excitation Energy to Masking Threshold

이때 각각의 임계대역에서의 여기에너지로부터 전역 마스크 임계치(Global Masking Threshold)의 계산은 다음과 같이 이루어지며[1][2]

$$T_G(z; t) = F^c(z; t) - O(z; t) \quad (\text{in dB}) \quad (6)$$

이때, Excitation에서 Masking Threshold로의 Distance를 나타내는 Offset 계산은 다음과 같이 계산되어진다.

$$O(z) = t_{index}(z) \cdot (14.5 + z) + (1 - t_{index}(z)) \cdot 5.5 \quad [\text{dB}] \quad (7)$$

여기서, t_{index} 는 Tonality Index를 나타내며, 다음과 같

이 계산되어진다.

$$t_{index}(z) = -0.299 - 0.43 \cdot \ln(cb(z)) \quad (8)$$

where, $0 \leq t_i \leq 1$

이때, cb 는 Normalized Unpredictability Measure를 나타내며 다음과 같이 구한다. [7]

$$cb(z; t) = \frac{F^{ce}(z; t)}{F^c(z; t)} \quad (9)$$

where, $0 \leq cb \leq 1$

여기서, F^c 는 임계대역의 여기에너지를 나타내며, F^{ce} 는 Frame Unpredictability Measure가 적용된 임계대역의 여기에너지를 나타내며 다음과 같이 구해진다.

$$F^{cl}(z; t) = \sum_{\omega=k_z}^{k_{z+1}} C(\omega; t) F(\omega; t) \quad (10)$$

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

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

이때, $F(\omega; t)$ 는 Critical Band Intensity, $F'(\omega; t)$ 를 식 (5)에 의해 주파수 영역으로 확대한 후에 Frame Unpredictability Measure의 가중치를 적용하여야 한다.[7]

$$C(\omega; t) = \frac{|S(\omega; t)e^{j\varphi(\omega; t)} - S(\omega; t-1)e^{j\varphi(\omega; t-1)}|}{|S(\omega; t)| + |S(\omega; t-1)|}, \quad (12)$$

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

$S(\omega; t), S(\omega; t-1)$ 는 이전 Frame과 현재 Frame의 진폭 스펙트럼을 나타내고, $\varphi(\omega; t), \varphi(\omega; t-1)$ 는 각각의 위상 스펙트럼을 나타낸다.

4. Inter-band Smoothing 기법

4-1. Spreading Function Expansion 기법

이러한 Inter-Band Discontinuity 문제는 신호의 특성을 각각의 임계대역마다 제한하게 되므로 음질 열화의 원인을 제공하게 된다. 이러한 각 Band간 Discontinuity를 포함한 신호에너지를 각 대역의 Intensity, $F'(z; t)$ 의 에너지 확산의 영향을 고려하여 신호의 특성을 자연스럽게 반영한 Smoothing 처리를 할 수 있다.

다음 그림 2.는 각각의 임의의 Critical Band Intensity, $F'(z; t)$ 에 대한 에너지 확산 특성을 나타내고 있다.

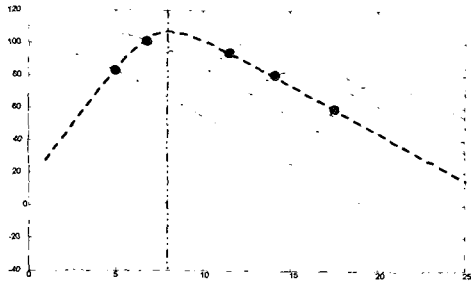


그림 2. 임계대역에서 확산함수가 적용된 여기에너지

이때, 그림에서 살펴볼 수 있듯이 임의의 임계대역에서의 여기에너지에 대한 Spreading 함수의 적용은 다른 임계대역으로의 에너지 확산의 영향을 표현하고 있는데, 각 임계대역에서의 확산에너지 레벨을 통해 다음과 같이 주파수 영역으로 확장할 수 있다.

이와 같이 주파수 영역으로 확장된 대역간 불연속성을 갖는 신호에너지의 불연속성을 제거하기 위한 각 임계대역에서의 Intensity에 의한 에너지 Spreading은 각 임계대역 간 주파수 성분들에 대해 Interpolation을 적용하여 Frequency Domain으로 확장할 수 있다.

$$\tilde{F}_m'(\omega; t) = \tilde{F}_m(\omega_{kz}; t-1) + \left[\frac{\tilde{F}_s(\omega_{kz}; t) - \tilde{F}_s(\omega_{kz}; t-1)}{\langle \omega_{kz} \leftrightarrow \omega_{kz} \rangle_{bins}} \right] k \quad (13)$$

where, $1 \leq k \leq \langle \omega_{kz} \leftrightarrow \omega_{kz} \rangle_{bins}$

$\omega_{kz} \leq \omega \leq \omega_{kz}$

m : critical band index

이때, $\langle \omega_{kz} \leftrightarrow \omega_{kz} \rangle_{bins}$ 는 각 임계대역 내에서의 주파수 성분들의 수를 나타내고, k 는 각각의 주파수 성분에 차례로 값을 주도록 하는 Indexing 역할을 한다.

다음 그림 3.은 이러한 확장된 주파수 영역에서의 여기에너지의 Spreading이다. [6]

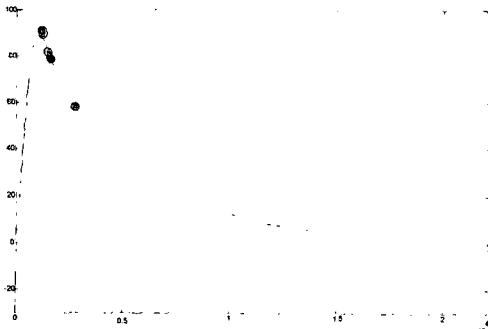


그림 3. 800Hz 여기에너지의 Spreading

이러한 임의의 임계대역에서의 Excitation 에너지 확산의 주파수 영역으로의 확장은 다음과 같이 수행되어 질 수 있다.

$$T_g(\omega; t) = 10 * \log_{10} \left(\sum_{m=1}^M 10^{\frac{\tilde{F}_s'(m,t)}{10}} \right) \quad (14)$$

where, $M = \text{MaxBark}$

이때, 이러한 임의의 Critical Band 각각의 여기에너지의 주파수 영역에서의 확산에너지는 다음과 같이 더해져서 전역 Excitation 에너지 Spreading, $T_g(\omega; t)$ 을 가할 수가 있다.

식 (14)와 같이 계산되어진 전역 Excitation 응답, $T_g(\omega; t)$ 를 이용하여 3장에서 과정들을 통해 전역 마스킹 임계치 정도의 에너지 레벨을 갖는 전역 Smoothing 응답(Global Smoothing Response)을 구할 수 있게 된다.

4-2. Simple Excitation Energy Interpolation 기법

위와 같은 Spreading Function Expansion 기법을 적용한 전역 Smoothing 응답을 이용할 경우 전역 마스킹 임계치와 같은 응답을 얻을 수 있다. 그러나 각 임계대역에서의 에너지 확산 현상을 모두 고려해야 하므로 계산량의 증가가 요구되어진다. 이때 Bark Domain에서의 Critical Band Intensity와 이에 따른 에너지 확산의 영향을 간략히 표현한 Bark Domain에서의 Excitation Energy, $F'(z; t)$ 을 식 (13)을 이용하여 간략화 된 전역 Excitation 응답을 구할 수 있다.

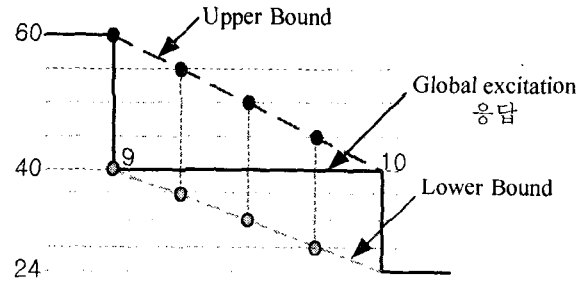


그림 4. Simple Excitation Energy Interpolation 기법

그림 4.는 두 임계대역의 차이를 각 성분이 나누어짐으로써 Discontinuity한 Global Excitation 응답을 Smooth하게 하는 방법을 대략적인 그림으로 나타내었다. 그림에서 알 수 있듯이 Simple Excitation Energy Interpolation 기법으로 두 가지를 생각할 수가 있다. 임계대역 간의 Lower Bound를 사용하거나 Upper Bound를 사용할 수가 있다. Lower Bound를 사용하면 높은 Excitation Energy보다 레벨이 많이 낮아지기 때문에 본 논문에서는 Upper Bound를 사용한 Interpolation 기법을 사용하였다.

임계대역별로 얻어진 Discontinuous Masking 곡선의 여기 에너지에 식 (13)에서와 같은 Interpolation 기법을 적용함으로써 Discontinuity가 사라진 간략화 된 전역 Smoothing 응답 곡선을 얻을 수 있었다.

5. 실험 및 결론

실험에서 사용한 신호들은 일반 오디오 CD로부터 확

특한 신호를 1024의 Frame 크기를 사용하고, Overlap-add Hanning Window를 적용 하였으며, Perceptual Filtering[4]을 이용하여 처리된 신호를 통해 Inter-Band Smoothing 처리를 수행하여 비교하였다. 이 때, Smoothing 과정은 다음과 같이 수행 되어 진다.

$$\hat{S}(\omega;t) = S(\omega;t) \cdot 10^{\frac{10 \log_{10} \left(\frac{\tilde{F}(\omega;t)}{T_g(\omega;t)} \right)}{20}} \quad (15)$$

$$= S(\omega;t) \sqrt{\frac{\tilde{F}(\omega;t)}{T_g(\omega;t)}}$$

이때, $S(\omega;t)$ 는 Filtering된 신호의 진폭 스펙트럼을, $\tilde{F}(\omega;t)$ 는 임계 대역간 불연속성을 포함한 Global Excitation 응답을 나타낸다. 다음 그림 6.은 이러한 $\tilde{F}(\omega;t)$ 와 Global Smoothing Response, $T_g(\omega;t)$ 을 나타낸다.

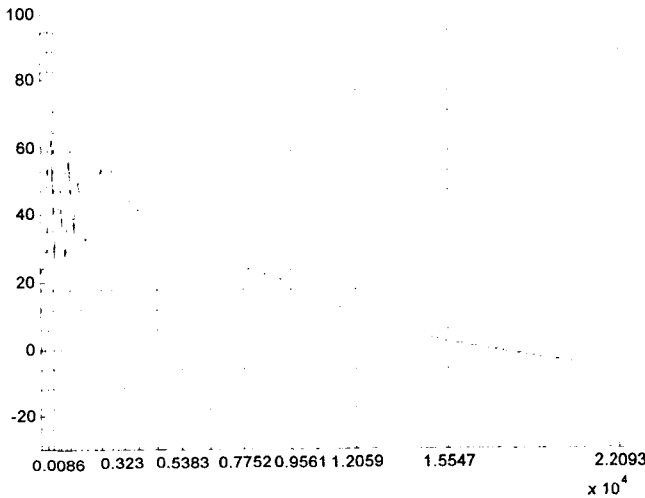


그림 6. Excitation Energy vs Smoothing Energy

다음 그림 7.은 각 임계 대역에서의 여기 에너지의 Discontinuity와 Smoothing기법을 적용한 후의 에너지 분포 결과를 나타낸다. 이때 점선은 SNR를 화이트 가우시안 노이즈(White Gaussian Noise)를 이용하여 20dB로 열화 시킨 오디오신호에 Perceptual Filtering[4]을 처리한 신호 스펙트럼을 나타내고, 실선은 이 결과 신호에 Smoothing 기법을 적용한 신호의 스펙트럼을 나타낸다. 그림에서 확인 할 수 있듯이 Inter-Band Smoothing 기법을 적용한 후 대역간 불연속성이 최소화 됨을 확인 할 수 있다.

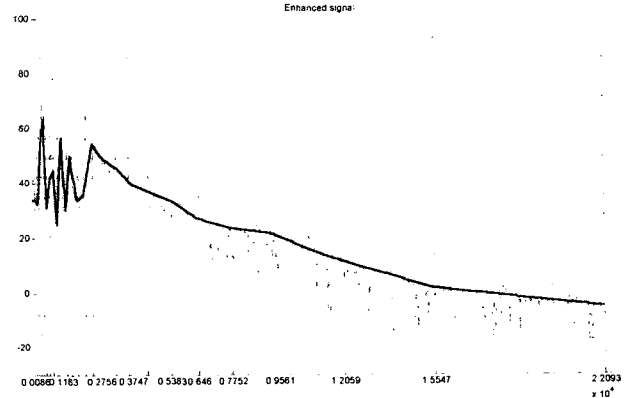


그림 7. Smoothing 기법을 적용한 신호

다음 그림 8. SNR를 화이트 가우시안 노이즈를 이용하여 20dB로 열화 시킨 다양한 장르의 오디오 신호에 대한 Perceptual Filtering[4] 처리 후 Inter-Band Smoothing 처리 결과에 대한 MOS 실험 결과를 나타낸다. 실험 결과 Inter-Band Smoothing처리를 함으로써 약 평균 0.3의 음질 개선을 확인 할 수 있었으며, 보다 자연스러운 음질개선 효과를 나타냄을 알 수 있었다.

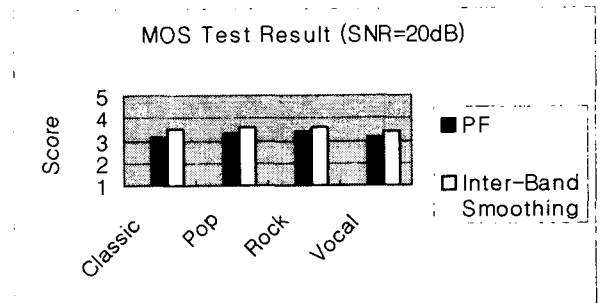


그림 8. MOS Test Result

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. Królkowski, "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
7. ISO/IEC, JTC1/SC29/WG11 MPEG, "Information technology - Generic coding of moving pictures and associated audio - Part 3 : Audio," IS13818-3 1994 ("MPEG-2").