

논문 2013-50-12-25

라우드니스 복원에 기반한 잡음 환경에서의 오디오 청취 향상

(Audio Listening Enhancement in Adverse Environment based on Loudness Restoration)

박 준 형*, 신 중 원**

(Junhyeong Pak and Jong Won Shin[Ⓢ])

요 약

잡음이 있는 환경에서 음악을 들을 경우 잡음의 영향으로 인해 명료한 음악의 청취가 힘들다. 본 논문에서는 잡음 환경에서 오디오 신호를 자동으로 변화시킴으로써 잡음 환경에서의 오디오 청취 경험을 향상시킬 수 있는 방법을 제안한다. 구체적으로, 잡음이 있는 환경에서 밴드 별 오디오 신호의 지각적 크기가 잡음이 없는 경우와 비슷해지도록 오디오 신호를 변화시키는 방법을 제안하였다. 이를 위해 Moore의 라우드니스 지각 모델^[1]을 도입하였으며, 기존의 음성 강화를 목적으로 한 논문을 더욱 발전시켜 48kHz로 샘플링된 전대역 신호를 증폭시키는 기법을 제안하였다. 잡음이 심할 때에는 라우드니스를 복원시켜도 명료성이 떨어지므로 이를 위해 라우드니스를 복원하는 데에 그치지 않고 일부러 고주파의 라우드니스를 약간 더 증폭하는 방법도 제안하였다. 실험 결과를 통해 우리는 제안된 알고리즘이 잡음 환경에서의 오디오 청취 경험을 얼마나 향상시킬 수 있는지를 알 수 있었다.

Abstract

It is hard to listen to the music clearly in the presence of background noise. In this paper, a method that modifies the audio signal automatically to enhance the audio listening experience in adverse environment is proposed. Specifically, the method that amplifies the audio signal so that the perceived loudness of audio signal in each band becomes similar to that of the noiseless signal. The loudness perception model proposed by Moore et. al is utilized. Extending the previous work that is applied to speech reinforcement, the full band signal sampled at 48kHz is manipulated based on the loudness restoration principle. Moreover, based on the observation that the audio clarity is compromised even with loudness restored signal, a modification that intentionally boosts high frequency loudness more than lower band is also proposed. Experimental results showed that the proposed algorithm can enhance the audio listening experience in adverse environment.

Keywords : listening enhancement, loudness perception model, partial specific loudness, loudness restoration

I. 서 론

* 학생회원, ** 정회원, 광주과학기술원 정보통신공학부
(School of Information and Communications,
Gwangju Institute of Science and Technology)

Ⓢ Corresponding Author (E-mail: jwshin@gist.ac.kr)

※ 본 연구는 2013년도 정부(미래창조과학부)의 재원으로 한국연구재단의 기초연구사업(NRF- 2013R1A1A1013418)과 미래창조과학부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업(NIPA-2013-H0301-13-4005)의 지원을 받아 수행되었음.

접수일자: 2013년10월22일, 수정완료일: 2013년11월21일

잡음이 커질수록 음성 혹은 음악에 대한 청음은 어려워진다. 차 안에서, 지하철이나 버스 안에서, 혹은 시끄러운 거리에서 음악을 들어본 적이 있는 사람들은 모두 이를 경험해 보았을 것이다. 일일이 음량을 조절해주게 되면 어느 정도 이에 대한 보상이 가능하지만 잡음의 특성에 따라 달리 보상을 해 주는 것은 어렵다. 음성 통

신에서 흔히 쓰이는 음성 향상 방법^[2-3]은 잡음과 원하는 신호가 섞여 있는 상황에서 잡음을 제거하는 방식으로 많은 연구가 이루어져 왔지만, 잡음 환경에서의 청음의 경우 잡음에 대한 직접적인 통제가 어려워 기존의 음성 향상 기법을 활용할 수는 없다.

이러한 문제를 해결하기 위해 본 논문에서 제안하는 방법은 잡음이 섞여있는 상황에서의 오디오 신호를 강화하는 알고리즘이다. 제안된 알고리즘을 이용하면 인간의 청각 지각 모델을 이용하여 각 주파수 밴드의 음악 신호를 적절히 증폭함으로써 음악 청취 경험을 보다 낫게 만들 수 있다. 물론 이를 위해서는 잡음 추정을 위해 마이크가 필요하다. 최근에는 스마트폰으로 음악을 듣는 사람들이 많아졌는데, 이 경우 스마트폰에 내장된 마이크들과 음성 통화를 위해 들어 있는 잡음 추정 알고리즘을 이용하면 추가적인 하드웨어나 알고리즘 없이 잡음을 추정할 수 있다. 보통의 MP3 플레이어나 씨디 플레이어 등에서도 별도로 마이크 및 잡음 추정 모듈을 추가한다면 이러한 기능을 활용할 수 있다. 그림 1은 제안한 알고리즘의 간단한 블록도를 보여준다.

가장 간단한 방식은 물론 오디오 신호의 전체 에너지를 증폭시키는 방식이다. 이 방법은 간단하기는 하지만 잡음의 주파수 특성이 음성과 매우 다른 경우, 특정 밴드의 잡음으로 인해 다른 밴드까지 지나치게 증폭되는 부작용이 있으며, 얼마나 키우는 것이 좋은지에 대한 근거가 없다. 다른 방법으로는 오디오 신호에 잡음이 완벽하게 마스킹되도록 음성의 주파수 밴드별 에너지를 증폭할 수도 있다^[4]. 하지만 잡음이 안 들릴 때까지 오디오 신호를 키우는 것은 잡음이 일정 수준 이상일 경우 소리가 지나치게 커지게 된다. 또 다른 방법으로 각 주파수 밴드에서 동일한 신호 대 잡음비(signal-to-noise ratio, SNR)를 갖도록 오디오 신호의 주파수 성분

을 키우는 방식이다^[5-8]. 하지만 사람의 청각은 비선형이기 때문에 신호 대 잡음비를 맞추는 것이 사람이 듣기에 가장 적합한 방식이라고는 할 수 없다. 또한 언급한 세가지 알고리즘 모두 출력 신호의 크기가 원래 오디오 신호의 크기보다는 잡음의 크기에 더 좌우되는 경향을 보이게 된다.

일반적으로, 우리가 원하는 신호가 아닌 잡음이 존재할 때 오디오 신호의 청각지각적 크기는 줄어든다^[1,9]. 대부분 잡음이 있는 환경에서 통화를 하거나 음악을 듣게될 때 이러한 상황을 경험한 적이 있을 것이다. 마스킹 효과란, 강한 신호(masker)가 약한 신호(maskee)와 비슷한 시간이나 주파수 영역에서 공존할 경우 작은 신호가 들리지 않는 현상을 말한다. 한 오디오 신호의 관점에서 보면, 다른 신호가 이 신호보다 훨씬 작다면 다른 신호는 들리지 않고 이 신호는 원래 이 신호만 존재할 때의 크기로 들릴 것이다. 반대로 다른 신호가 이 신호보다 훨씬 크다면 이 신호는 다른 신호에 마스킹되어 거의 들리지 않을 것이다. 만약 두 신호의 세기 차이가 충분하지 않다면 이 경우는 위의 두 경우의 중간이 되어 그 오디오 신호는 다른 신호의 영향으로 원래 그 신호만 존재할 때에 비해 그 크기가 줄어든 것으로 느껴질 것이다. 이를 부분 마스킹 효과라고 하고, 부분 라우드니스(partial loudness), 또는 부분 마스킹 라우드니스(partial masked loudness)는 다른 신호가 존재할 때 이로 인해 줄어든 것으로 느껴지는 어떤 신호의 라우드니스를 의미한다^[1,9].

[10]에서는 라우드니스 지각 모델을 바탕으로한 음성 강화 방법이 제안되었다. 이 방법은 잡음 환경에서 음성 신호의 주파수 밴드 별 부분 라우드니스가 잡음이 없는 상태의 그것과 같도록 하여 들리는 그 음성의 크기가 각 주파수에서 원래 신호와 비슷하게 만드는 방법이며, 이를 위해 Moore의 라우드니스 지각 모델을 도입하였다^[11].

본 논문에서는 [10]에서 제시된 음성 강화 기법을 확장하여 일반적인 오디오 신호에 적용하였다. 우선, 모델을 48kHz로 샘플링된 전대역 신호를 대상으로 확장하였고, 푸리에 변환의 주파수 해상도 한계로 인해 생기는 청각 모델과 실제 구현의 대역폭의 차이를 보상하였다. 또, 계산된 주파수 대역 별 이득의 시간적 평활화에 있어 오디오 신호가 처음 시작하는 부분(onset)과 그 외의 부분을 구분하여 평활화 계수를 다르게 적용함으

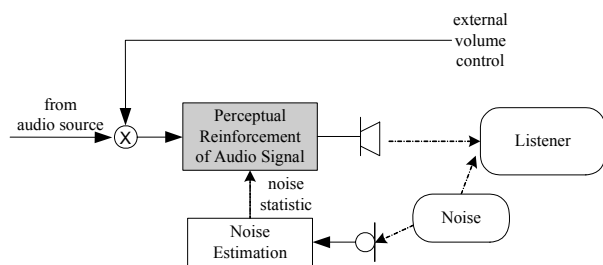


그림 1. 오디오 청취 향상 시스템의 블록도
Fig. 1. Block diagram of an audio listening enhancement system.

로써, 보다 빠르고도 안정적인 증폭이 가능하게 했다. 또한 잡음이 심할 때에는 라우드니스를 복원하여도 오디오의 명료성이 떨어지게 되므로 신호 대 잡음비가 낮은 경우 고주파 신호의 지각적 크기를 저주파에 비해 보다 강조하는 알고리즘을 제안하였다. 음질에 대한 선호도 테스트 결과를 통해 본 알고리즘을 이용하면 오디오 청취 경험의 향상을 도모할 수 있다는 점을 확인할 수 있었다.

II. 라우드니스 복원에 기반한 오디오 청취 향상

본 논문에서 제안하는 오디오 청취 향상 알고리즘에 도입된 라우드니스 지각 모델은 그림 2와 같이 표현된다^[1]. 이 Moore의 라우드니스 지각 모델을 이용하면 원하는 오디오 신호와 잡음 신호의 주파수 스펙트럼으로부터 잡음이 없는 오디오 신호의 상세 라우드니스 (specific loudness, 밴드별 라우드니스)와 잡음이 섞인 오디오 신호의 부분 상세 라우드니스(partial specific loudness, 잡음에 의해 저하된 밴드별 라우드니스)를 수학적으로 얻을 수 있다. 그림 2의 첫 번째 블록은 자유 공간에서 고막까지의 음압에 대한 전달함수를, 두 번째 블록은 중이의 전달함수를 의미한다. 그리고 세 번째 블록은 달팽이관에 도달하는 음압의 스펙트럼에서 흥분 (excitation) 패턴을 추출하고 주파수를 인간의 청각에서 쓰는 등가 사각 대역폭(Equivalent Rectangular Bandwidth, ERB)으로 변환함을 나타낸다. 따라서, 헤드폰이나 이어폰으로 청취가 이루어질 경우 첫 번째 블록의 전달 함수는 생략되어야 한다.

Moore의 라우드니스 지각 모델^[1]은 오디오신호와 잡음에 의한 흥분으로부터 상세 라우드니스와 부분 상세 라우드니스를 구하는 수학적 공식도 포함한다. 여기서 상세 라우드니스는 ERB당 라우드니스를, 부분 상세 라우드니스는 ERB당 부분 라우드니스를 나타낸다. 오디오 신호에 의해 발생하는 흥분 E_{SIG} 이 조용한 환경

에서 들을 수 있는 한계 E_{THRQ} 보다 크거나 같으며 달팽이관 증폭기 포화 한계보다 작다면 잡음이 없을 때의 상세 라우드니스 N'_Q 는 아래와 같이 표현된다.

$$N'_Q = C[(GE_{SIG} + A)^\alpha - A^\alpha] \tag{1}$$

식 (1)에서 C, G, A 와 α 는 실험을 통해 얻는 상수 값으로 [1]에 나와 있다. 앞서 언급한 상황과는 반대로 E_{SIG} 이 E_{THRQ} 에 비해 작거나 달팽이관 증폭기 포화 한계보다 크다면 N'_Q 는 식 (1)과는 다르게 주어지지만, 들리지 않거나 지나치게 큰 신호의 경우는 우리의 관심사가 아니다. 한편, 잡음이 있는 경우의 부분 상세 라우드니스 $N'_{partial}$ 은 E_{SIG} 이 마스킹 한계(masking threshold) E_{THRQ} 보다 크거나 같으며 E_{SIG} 와 E_{NOISE} 의 합이 달팽이관 증폭기 포화 한계보다 작을 경우 아래와 같이 표현된다^[1].

$$N'_{partial} = C[(E_{SIG} + E_{NOISE})G + A]^\alpha - A^\alpha - C[(E_{NOISE}(1 + K) + E_{THRQ})G + A]^\alpha \tag{2}$$

더 정확한 모델은

$$N'_{partial} = C[(E_{SIG} + E_{NOISE})G + A]^\alpha - A^\alpha - C[(E_{NOISE}(1 + K) + E_{THRQ})G + A]^\alpha - (E_{THRQ}G + A)^\alpha \left(\frac{E_{THRQ}}{E_{SIG}}\right)^{0.3} \tag{3}$$

이 된다. 식 (3)에서 $E_{THRQ} = KE_{NOISE} + E_{THRQ}$ 와 같이 모델링할 수 있으며, K 는 실험을 통해 결정되는 주파수에 따라 다른 상수이다. 물론 이 경우에도 E_{SIG} 이 E_{THRQ} 보다 작거나 E_{SIG} 와 E_{NOISE} 의 합이 달팽이관 증폭기 포화한계보다 크면 우리의 관심사를 벗어난다. 식 (2)와 (3)의 차이는 E_{SIG} 이 E_{THRQ} 에 비해 때

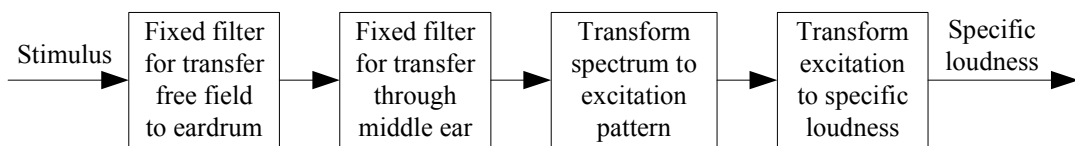


그림 2. 오디오 청취 향상 알고리즘에 도입된 라우드니스 지각 모델의 처리 단계들의 블록도
 Fig. 2. Block diagram of the processing stages in the loudness perception model.

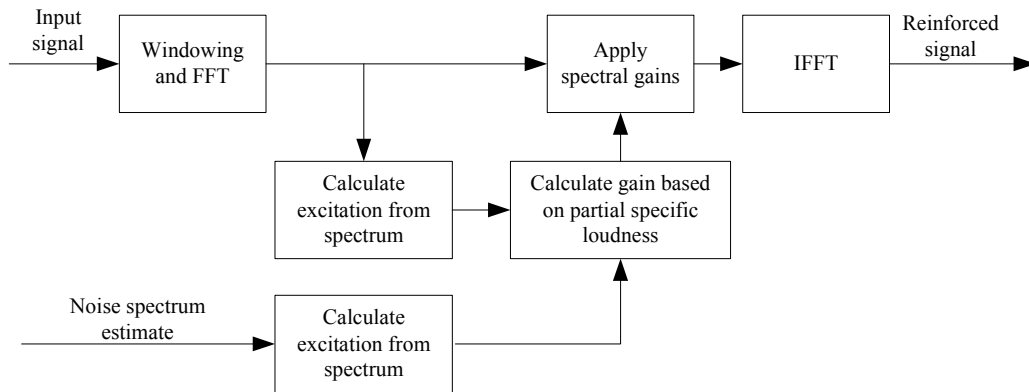


그림 3. 제안된 오디오 청취 향상 알고리즘의 블록도
Fig. 3. Block diagram of the proposed audio listening enhancement algorithm.

우 클 때 음성 신호의 부분 라우드니스가 잡음이 없는 음성 신호의 라우드니스와 동일하게 만들기 위해 추가된 크기 조절 계수 $(E_{THRQ}/E_{SIG})^{0.3}$ 에 있다^[1].

본 논문에서 제안하는 오디오 신호 청취 향상을 위한 알고리즘의 구현에 관련된 모든 상수값과 필터들은 [1]에 언급된 값으로 설정하였다. Moore의 라우드니스 지각 모델은 단음조, 복음조 및 균일홍분잡음^[9] 자극에 관련된 측정된 결과에 대한 모델링을 위하여 만들어진 것이기 때문에 균일하지 않고 광대역의 파워 스펙트럼이 특징인 음성이나 음악 신호에 대해서는 정확하지 않을 수 있다.

제안하는 오디오 신호 청취 향상을 위한 시스템을 그림 3과 같이 표현할 수 있다. 잡음 신호의 파워 스펙트럼은 스마트폰에 포함된 마이크 및 잡음 추정/제거 알고리즘이나 추가로 장착된 마이크 및 잡음 추정 알고리즘으로부터 추정이 가능하다고 가정한다. 우선 그림 2에 언급된 라우드니스 지각 모델을 바탕으로 오디오 신호로 인한 흥분 E_{SIG} 와 잡음으로 인한 흥분 E_{NOISE} 를 구한다. 실제로는 고속 푸리에 변환(fast Fourier transform, FFT)으로 얻어진 스펙트럼의 주파수들을 묶어 밴드를 구성하게 되므로 E_{SIG} 와 E_{NOISE} 를 구할 때, ERB를 실제로 사용된 신호의 대역폭으로 나눈 값을 구해진 신호의 밴드 전력과 곱해서 대역폭 차이를 보상해준다. 그리고 각 밴드에 대하여 원래 신호에 어느 정도의 이득(gain)을 곱하면 원래 잡음이 없는 상태의 음성 신호와 같은 수준의 부분 상세 라우드니스가 도출되는지 그 이득을 계산하도록 한다. 어떠한 대역에 곱해질 이득을 g 라고 한다면, E_{SIG} 대신 gE_{SIG} 를 대

입했을 때 식 (2) 또는 (3)에 언급된 부분 상세 라우드니스 $N'_{Partial}$ 는 식 (1)의 N'_Q 와 동일해야 한다. 부분 상세 라우드니스의 모델을 적용하여 식 (2)를 이용한다면, 이득 g 를 아래와 같이 구할 수 있다^[10].

$$g = \frac{[(GE_{SIG} + A)^\alpha + f(E_{NOISE})]^\frac{1}{\alpha} - A}{E_{SIG}} - E_{NOISE} \quad (4)$$

이때 $f(E_{NOISE})$ 는 아래와 같이 주어진다.

$$f(E_{NOISE}) = [(E_{NOISE}(1+K) + E_{THRQ})G + A]^\alpha - (E_{THRQ}G + A)^\alpha \quad (5)$$

위의 식에 따르면 E_{NOISE} 가 같을 때, 이득은 신호의 레벨이 낮을 경우에 높게 되므로 마스킹이 심한 성분들의 라우드니스의 복원이 이루어진다. 여기서 한 가지 주의할 점은 출력 오디오 신호의 레벨이 잡음 레벨보다 낮을 수 있다는 점이다. 신호 대 잡음비 기반 방법에서는 원래 신호가 그 밴드에서는 아무리 작게 들리던 신호였어도 출력 신호에서는 그것을 잡음 레벨 위로 올려 놓게 된다.

잡음이 심할 경우 밴드별 라우드니스를 복원하여도 오디오의 명료성과 가해성이 저하될 수 있다. 이를 해결하기 위해 신호로 인한 흥분 대 잡음으로 인한 흥분비가 낮을 경우 고주파 밴드의 라우드니스를 조용할 때의 라우드니스보다 조금 더 크게 해 주고, 저주파 밴드

에서는 더 작게 하도록 알고리즘을 변형할 수 있다. 즉,

$$N'_{partial}(gE_{SIG}) = \kappa N'_Q \quad (6)$$

$$\kappa = \gamma \times 1.0 + (1 - \gamma)\xi \quad (7)$$

이 되고, 여기서 ξ 는 밴드별 최악의 잡음 상황에서의 라우드니스 가중치이고 γ 는 그 밴드의 신호로 인한 흥분 대 잡음으로 인한 흥분 비에 비례하는 상수이다. γ 를 구하는 식은 여러 가지를 쓸 수 있지만, 사용된 식은 다음과 같다.

$$\gamma = \frac{2 \cdot \exp\left(-\eta \frac{E_{NOISE}}{E_{SIG}}\right)}{1 + \exp\left(-\eta \frac{E_{NOISE}}{E_{SIG}}\right)} \quad (8)$$

여기서 η 는 신호로 인한 흥분 대 잡음으로 인한 흥분 비에 γ 값이 얼마나 민감하게 변하느냐를 결정하는 상수이다.

식 (6)을 이용한 이득 함수는 식 (2)를 이용할 경우 아래와 같이 표현된다.

$$g = \frac{[\kappa(GE_{SIG} + A)^\alpha + f'(E_{NOISE})]^\frac{1}{\alpha} - A}{G} - E_{NOISE} \quad (9)$$

이때 $f'(E_{NOISE})$ 는 아래와 같이 주어진다.

$$f'(E_{NOISE}) = [(E_{NOISE}(1 + K) + E_{THRQ})G + A]^\alpha - (E_{THRQ}G + A)^\alpha - (\kappa - 1)A^\alpha \quad (10)$$

식 (3)을 쓸 경우 이득은 간단한 식으로 주어지지 않는다. 식 (2)와 식 (3)의 차이는 E_{SIG} 이 E_{NOISE} 보다 매우 클 때 발생하므로, 식 (2)를 이용하여 얻은 이득 함수 식 (9)를 E_{SIG} 이 E_{NOISE} 보다 매우 클 때 줄여 준다면 식 (3)을 사용하는 것과 비슷한 효과를 얻을 수 있다. 이는 여러 가지 방법으로 수행할 수 있는데, 이득을 줄이는 간단한 방법 중 한 가지는 다음과 같다^[10].

$$\begin{aligned} \tilde{g} &= \beta g + (1 - \beta) \times 1.0 \quad \text{if} \\ &gE_{SIG} > E_{THRQ} \times 100, \\ \beta &= \frac{E_{THRQ} \times 100}{gE_{SIG}} \end{aligned} \quad (11)$$

\tilde{g} 는 변형된 이득을 의미한다. 이 경우 E_{SIG} 이 E_{NOISE} 보다 매우 클 때의 심한 신호의 증폭을 방지함으로써 이득을 식 (3)에 따른 이득에 가깝게 한다.

마지막으로 밴드 이득이 시간에 따라 너무 심하게 변할 경우 듣기 싫은 부작용이 생길 수 있으므로 밴드 이득을 다음과 같이 시간에 따라 평활화해 준다.

$$\hat{g}(t) = \lambda \tilde{g} + (1 - \lambda)\hat{g}(t - 1) \quad (12)$$

이 때, 이전 프레임과 현재 프레임에서 E_{SIG} 와 E_{THRQ} 를 비교하여 얻을 수 있는 오디오의 시작 구간에서는 다른 평활화 계수를 사용하여 평활화를 덜 해 주어야 빠르게 반응하면서도 부드럽게 변화하는 오디오 신호를 얻을 수 있다. 또한 최대 이득 값과 평활화에 쓰이는 이전 프레임 이득의 최대값을 다르게 설정하여, 한 프레임에서는 상당히 큰 증폭을 가능하게 하고 그 영향이 지나치게 오래 가는 것을 막았다.

과위 스펙트럼으로부터 흥분 패턴으로의 변환은 보통의 크기의 신호에 대해서는 선형으로 잘 모델링할 수 있으므로, 각각의 ERB에 대하여 계산된 $\sqrt{\hat{g}}$ 을 이에 해당하는 주파수 성분들에 각각 곱하면 원하는 출력 오디오 신호를 얻을 수 있다.

III. 실험 결과

제안하는 음성 신호 강화 알고리즘의 성능을 검증하기 위해, 주관적 선호도 테스트를 실시하였다. 제안된 알고리즘이 인간의 청각 모델에 기반을 둔 것이기 때문에, 신호 대 잡음비 등 객관적인 척도들은 큰 의미를 가지기 힘들다. 단, 테스트의 편의성을 위해 헤드폰을 통한 청음 테스트를 하기 전 출력 오디오 신호에 배경 잡음을 더하여 들려주었다. 실험에 사용한 음악 파일은 8초 길이의 서로 다른 3개의 음악 파일들로 이루어졌다. 첫 번째 파일은 사람의 음성이 없는 연주형태의 음악이며, 두 번째 파일은 빠르고 경쾌한 분위기의 악기연주와 빠른 박자의 사람의 음성이 결합된 형태의 음악이다. 마지막 파일은 두 번째 파일에 비해 상대적으로 조용한 분위기의 악기연주와 사람의 음성이 결합된 형태이다. 3개 음악 파일의 샘플링은 모두 48kHz이다. 잡음의 경우 NOISEX-92의 데이터베이스로부터 배벨 잡음(babble noise)과 차 내부 잡음(volvo noise), 그리고 백

색 잡음(white noise)를 활용하였다. 24세부터 32세까지의 10명의 피시험자(7명의 남성, 3명의 여성)로 실험을 진행했으며, 이중 2명은 음성/오디오 신호 처리가 전공인 학생이다. 이 실험에서는 잡음 스펙트럼은 완벽하게 추정되었다고 가정하였다.

우선 배경 잡음이 있는 경우에 향상된 음악의 느껴지는 음질과 아무런 처리도 되지 않은 음악 신호의 음질을 비교하기 위해 주관적인 선호에 대한 테스트를 실시하였다. 이 테스트는 본 알고리즘이 여러 가지 잡음 환경 하에서 어느 정도로 음질을 개선할 수 있는지를 확인하기 위하여 실시되었다. 이 테스트는 원래 잡음이 없는 상태의 깨끗한 음악 신호가 청자에게 참고로 주어진 점만 다르고 ITU-T P.800의 비교 범주 평가 (comparison category rating) 테스트^[11]와 동일하다. 이때 각 피시험자는 선호도를 -2 ~ 2점의 점수로 매긴다. 그리고 테스트결과로 얻은 점수를 평균 내어 최종 점수를 결정한다. 이 결과들은 표 1에 정리되어 있다. 여기서 양수가 제안된 방식의 출력 음악 신호를 더 좋게 평가했음을 의미한다. 또한 계산된 평균은 주로 낮은 입력 신호 대 잡음비에서 더 높게 나타나는데 그 이유는 아무런 처리도 되지 않은 음악 신호가 잡음에 의해서 심각하게 마스킹되어서이다. 차 내부 잡음의 경우 점수가 다소 낮게 나왔는데, 같은 신호 대 잡음비라고 해도 차 내부 잡음은 저주파 성분이 많아서 지각적으로 큰 영향을 주지 못하기 때문이다. 이 실험 결과로부터 제안된 오디오 청취 향상 방법이 여러 잡음 환경 하에서 청자가 느끼는 음악 신호의 음질을 개선한다고 할 수 있다.

두 번째와 세 번째 실험은 기존의 신호 대 잡음비 기반 알고리즘^[5~6]과 제안된 알고리즘을 비교하기 위한 것이다. 첫 번째 실험과 같은 잡음 환경 하에서 신호 대

표 1. 주관적 선호 테스트 결과 : 제안된 방식의 출력 음악 vs. 처리되지 않은 음악

Table 1. Subjective preference test result: The output of proposed algorithm vs. unprocessed audio signal under various noise conditions.

test set	proposed (4) - unprocessed			
	babble	volvo	white	average
-10 dB	1.70	0.27	1.57	1.18
-5 dB	1.50	0.10	1.47	1.02
0 dB	0.97	-0.13	1.37	0.73
average	1.39	0.08	1.47	0.98

표 2. 주관적 선호 테스트 결과 : 식 (4)에 기반한 제안된 방식의 출력 음악 vs. 신호 대 잡음비 기반 방식의 출력 음악

Table 2. Subjective preference test result: The output of proposed algorithm based on (4) vs. the output of SNR-based method under noise conditions.

test set	proposed (4) - SNR-based			
	babble	volvo	white	average
-10 dB	0.13	0.43	0.17	0.24
-5 dB	0.20	0.53	0.30	0.34
0 dB	-0.23	0.60	0.33	0.23
average	0.03	0.52	0.27	0.27

표 3. 주관적 선호 테스트 결과 : 식 (9)에 기반한 제안된 방식의 출력 음악 vs. 신호 대 잡음비 기반 방식의 출력 음악

Table 3. Subjective preference test result: The output of proposed algorithm based on (9) vs. the output of SNR-based method under noise conditions.

test set	proposed (9) - SNR-based			
	babble	volvo	white	average
-10 dB	0.37	0.60	0.17	0.38
-5 dB	0.20	0.67	0.00	0.29
0 dB	-0.07	0.50	0.27	0.23
average	0.17	0.59	0.14	0.30

잡음비 기반 알고리즘으로 처리된 출력 음악에 잡음을 섞은 것과 각각 식 (4), 식 (9)를 바탕으로 한 제안된 알고리즘으로 처리된 출력 음악에 잡음을 섞은 것의 주관적인 선호도를 마찬가지로 방식으로 비교하였다. 표 2, 표 3에는 이 주관적 선호도 테스트 결과가 나타나 있다.

이 표들은 제안된 알고리즘이 신호 대 잡음비 기반의 방식보다 주관적인 음질을 더 잘 개선함을 보여준다. 또, 식 (9)를 이용하여 잡음이 심할 경우 고주파를 강조해주게 되면 약간의 성능 향상을 얻을 수 있었다. 백색 잡음에서는 오히려 고주파를 강조해주지 않는 편이 더 성능이 나왔는데, 이는 백색 잡음의 경우 고주파 성분이 많아서 고주파 성분을 강조해 주면 보다 거슬리게 느껴지는 것 때문인 것으로 보인다.

IV. 결 론

본 논문을 통해 제안한 알고리즘은 잡음 환경에서 오디오 신호를 대역별로 변형시킴으로써 지각적인 음질을 개선하는 새로운 접근 방법이다. 라우드니스 지각 모델

을 사용하여 잡음 환경 하에서 오디오 신호의 부분 상
세 라우드니스가 잡음이 섞여 있지 않은 경우의 상세
라우드니스와 같은 값을 갖도록 하는 알고리즘을 제안
하였다. 더 나아가 낮은 신호 대 잡음비에서는 라우드
니스가 복원되어도 오디오의 명료성이 저하됨에 착안하
여, 신호 대 잡음 비가 낮은 상황에서는 고주파 대역의
라우드니스를 원래 잡음 없는 신호의 라우드니스보다
더 강조하는 알고리즘을 제안하였다. 보다 빠르고도 부
드러운 반응을 위해 최대 이득과 평활화를 위한 최대
이득값을 별도로 설정하고 오디오의 시작 구간에서는
다른 평활화 계수를 적용하였다. 주관적인 신호도 테스
트를 통해 제안된 알고리즘이 여러가지 잡음 환경하에
서 오디오 신호의 음질을 기존 방법보다 더 잘 개선함
을 알 수 있었다.

REFERENCES

- [1] B. C. J. Moore, B. R. Glasberg, and T. Baer, "A model for the prediction of thresholds, loudness, and partial loudness," *Journal of Audio Engineering Society*, vol. 45, no. 4, pp. 224-240, Apr. 1997.
- [2] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. ASSP-32, no. 6, pp. 1109-1121, Dec. 1984.
- [3] 3GPP2 Document C.S0014-0 v1.0, Enhanced Variable Rate Codec (EVRC), Dec. 1999.
- [4] M. Tzur (Zibulski) and A. A. Goldin, "Sound equalization in a noisy environment," *Audio Engineering Society 110th Convention*, Preprint No. 5364, May 2001.
- [5] B. Sauert and P. Vary, "Near end listening enhancement: Speech intelligibility improvement in noisy environments," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. 1, pp. I-493-I-496, 2006.
- [6] A. A. Goldin, A. Budkin and S. Kib, "Automatic volume and equalization control in mobile devices," *Audio Engineering Society 121th Convention*, Preprint No. 6960, Oct. 2006.
- [7] Jae-Hun Choi and Joon-Hyuk Chang, "Sound reinforcement based on context awareness for hearing impaired," *Journal of The Institute of Electronics Engineers of Korea*, vol. 48-SP, no. 5, pp. 109-114, Sep. 2011.
- [8] Jae-Hun Choi and Joon-Hyuk Chang, "Robust speech reinforcement based on gain-modification incorporating speech absence probability," *Journal of The Institute of Electronics Engineers of Korea*, vol. 47-SP, no. 1, pp. 175-182, Jan. 2010.
- [9] E. Zwicker and H. Fastl, *Psychoacoustics-Facts and Models*, Berlin: Springer, 1990.
- [10] J. W. Shin and N. S. Kim, "Perceptual reinforcement of speech signal based on partial specific loudness," *IEEE Signal Processing Letters*, vol. 14, no. 11, pp. 887-890, Nov. 2007.
- [11] ITU-T P.800, *Methods for Subjective Determination of Transmission Quality*, Aug. 1996.

저 자 소 개



박 준 형 (학생회원)
2008년 한국항공대학교
전자공학과 학사 졸업.
2013년 3월 ~ 광주과학기술원
정보통신공학부 석사과정.
<주관심분야 : 음성 및 음향신호
처리>



신 종 원 (정회원)
2002년 서울대학교 전기공학부
학사 졸업.
2008년 서울대학교 전기컴퓨터
공학부 박사 졸업.
2008년 ~ 2012년 Senior engineer,
Qualcomm Inc.
2012년 ~ 광주과학기술원 정보통신공학부 조교수
<주관심분야 : 음성 및 음향신호처리>