

지각 특성 및 라우드니스를 이용한 입체음향의 성능 개선

구 교 식^{a)}, 차 형 태^{a)‡}

Enhancement of the 3D Sound's Performance using Perceptual Characteristics and Loudness

Kyo-Sik Koo^{a)} and Hyung-Tai Cha^{a)‡}

요 약

인간의 청각기관은 소리의 방향과 거리인지에 있어서 양 귀에 들어오는 소리세기의 차이, 위상의 차이, 그리고 주파수 스펙트럼의 차이 등의 정보들을 복합적으로 이용한다. 이런 정보를 종합적으로 포함하고 있는 것을 머리전달함수라 하며, 이를 이용하여 실질적으로 존재하지 않는 음원을 근사적으로 생성할 수 있는 입체음향 시스템을 구현할 수 있다. 그러나 비개인화 된 머리전달함수는 음상정위 성능을 떨어뜨리는 주된 원인이 된다. 이에 본 논문에서는 이 머리전달함수를 이용한 3차원 음상정위를 위해 사람의 청각 특성을 이용한 알고리즘을 제안한다. 청자가 정확한 방향을 인지하기 위해 1차로 측정된 머리전달함수의 청각 자극 에너지를, 2차로 전역 마스킹 임계치와 라우드니스를 이용하여 방향감을 강조하였다. 제안된 알고리즘의 성능 평가를 위해 청감 테스트를 수행하였으며 실험 결과 기존의 알고리즘에 비해 입체음향의 전체적인 인지도가 50% 이상 증가하였음을 확인할 수 있었다.

Abstract

The binaural auditory system of human has ability to differentiate the direction and the distance of the sound sources by using the information which are inter-aural intensity difference(IID), inter-aural time difference(ITD) and/or the spectral shape difference(SSD). These information is generated from the acoustical transfer of a sound source to pinna, the outer ears. We can create a virtual sound system using the information which is called Head related transfer function(HRTF). However the performance of 3D sound is not always satisfactory because of non-individual characteristics of the HRTF. In this paper, we propose the algorithm that uses human's auditory characteristics for accurate perception. To achieve this, excitation energy of HRTF, global masking threshold and loudness are applied to the proposed algorithm. Informal listening test shows that the proposed method improves the sound localization characteristics much better than conventional methods.

Keyword : Head related transfer function, 3d sound, Psychoacoustics, Sound localization, Loudness

1. 서 론

나날이 발전하는 기술로 인하여 디지털 시대가 도래한 현재, 멀티미디어 기술은 신호처리 기술의 발달과 저렴한 DSP의 등장으로 인하여 우리 사회의 커다란 이슈로 자리

a) 숭실대학교 전자공학과
The School of Electronic Engineering, Soongsil university

‡ 교신저자 : 차형태 (senia2@mms.ssu.ac.kr)

※이 연구는 2011년도 숭실대학교 교내연구비 지원에 의한 연구임
· 접수일(2011년6월3일), 수정일(1차:2011년8월8일, 2차:8월23일), 게재확정일(2011년8월23일)

매김 하였다. 이 중에서도 가상현실 분야는 현재 가장 활발히 연구가 진행되고 있는 분야 중 하나라 할 수 있다. 가상현실이란 어떤 특정한 환경이나 상황을 컴퓨터를 이용하여 시뮬레이션 함으로서 그것을 사용하는 사람이 재현된 공간 안에 실제로 존재하는 것처럼 만들어주는 일을 뜻한다^[1]. 구체적인 예로서는 비행기 조종 훈련이나 수술 실습부터 우리가 실제로 겪을 수 없는 원자로 내에서의 작업이나 깊은 심해, 우주 저편에서의 활동 등 다양하며 이는 인간의 오감 중에서 시각, 청각, 촉각에 해당하는 그래픽, 사운드 및 운동감 기술의 실시간적인 상호작용으로 구현되어질 수 있다. 즉 인간의 오감을 만족시키는 멀티모달 기술이라 할 수 있으며 각 기술 중에서 사운드 기술은 가상현실 구현에서 가장 현실감을 느끼게 해주는 부분으로 3차원 입체음향이라는 말로서 대변할 수 있다^[2]. 입체음향이란 모노음의 음의 고저, 음색, 음원의 방향이나 거리감을 조절하여 듣는 사람으로 하여금 현장감이나 입체감을 가지게 하는 음향이다. 이것에 의하여 음원의 위치, 주위의 환경, 자신의 위치, 물체의 움직임, 음상의 형태와 표정 등을 얻을 수 있고, 각테일 파티 효과도 얻어진다^[3]. 이러한 입체음향을 구현하기 위하여 DVD 등에서는 5.1 채널 등의 다채널 음향재생 시스템을 적용하고 있다. 그러나 다채널 음향 재생 시스템은 스피커를 설치하기 위한 공간이나 비용 등의 문제가 발생하게 된다. 따라서 2개의 채널로 입체음향을 구현하는 바이노럴 시스템이 주목을 받고 있으며 대표적인 방법으로는 머리전달함수(HRTF; Head Related Transfer Function)를 이용한 방법을 들 수 있다. 머리전달함수는 자유 공간 상에서 음원과 귀 간의 경로를 모델링하는 함수로서 음원의 방향 및 위치에 따라 변하고 개인마다의 고유한 특성을 가지고 있다. 즉 음상의 공간화와 관련한 중요한 음향 파라미터이므로 헤드폰과 같은 바이노럴 재생 환경에서 음원의 방향성을 포함하는 공간감을 표현하는데 널리 적용되고 있다^[4]. 그러나 머리전달함수를 개개의 사람마다 측정하고 모델링하는 것은 많은 시간과 노력을 요구하는 일이므로 많은 시스템에서는 머리전달함수를 표준화된 더미 헤드 또는 청취자에게 양이 마이크를 착용시켜 직접 측정한 후 이를 DB 형태로 만들어 사용하고 있다. 그 결과로 입체음향의 성능이

떨어지는 현상이 발생하게 되는데, 특히 청취자의 두 귀 사이의 거리가 같은 혼돈원추상에서는 상/하, 전/후 등의 방향에 대한 정확한 방위를 판별할 수 없게 된다^[4]. 이에 수십 년 전부터 표준화된 머리전달함수를 이용하여 정확한 입체음향을 구현하고자 하는 연구가 진행되어 왔다. 그 대표적인 예로 Tan, Gan 및 Kim 등은 입체음향의 스펙트럴 큐를 부각시켜 방향감을 강조하고자 하였다^[5-8]. 그리고 Kim 및 Hwang 등은 거리 변화나 지저함수 등을 이용한 머리전달함수의 모델링에 대한 연구를 수행하였으며^{[9][10]} Lee, An 등은 머리전달함수 측정시스템을 개선하고자 하였다^{[11][12]}. 그러나 기존의 연구들은 머리전달함수의 다양한 특성을 모두 모델링하지 못할 뿐 아니라 머리전달함수가 적용될 모노 음원에 대해서도 전혀 고려하지 않고 있다. 만약 강조되는 주파수 대역에 모노 음원의 주파수 에너지가 강하게 존재할 경우 입체음향의 인지 단서들이 어긋나게 되고, 매 방향마다 음상의 인지거리나 음색이 달라지므로 정확한 음상의 정위가 불가능하다. 더불어 각각 개인화된 머리전달함수를 모두 측정하기에는 비용, 시간 등의 무리가 따르며 마스크와 같은 인간의 청각 특성에 따라 해당 방향과 관련한 단서가 인지되지 못할 수도 있어 이에 대한 고려가 필요하다. 이에 본 논문에서는 입체음향의 지각적인 특성을 이용하여 정확한 음상정위의 혼돈 문제를 개선하는 알고리즘을 개발하였다. 각 방향에 따른 입체음향의 고유한 변화가 인간의 청각에 미치는 영향을 계산한 후, 심리음향 모델과 라우드니스를 이용하여 해당 방향의 청각적 인지 단서를 부각시킴으로서 방향감을 강조하게 된다.

본 논문의 순서는 아래와 같다. 먼저 2장 및 3장에서 인간의 청각 특성을 기반으로 한 심리음향 모델 및 입체음향의 생성에 관해 설명한다. 그리고 4장과 5장에서 제안한 알고리즘에 대한 설명 및 시뮬레이션에 관한 테스트 결과를 제시한다. 마지막으로 6장에서 결론을 맺는다.

II. 지각 특성을 이용한 신호 해석

일반적인 음성 및 오디오 신호는 사람의 귀를 통하여 전

달되어지고 분석된다. 사운드에 대한 사람의 지각은 귀의 마스킹 현상에 가장 큰 영향을 받게 되고 이러한 지각적 특성을 이용하여 수학적 모델링을 하게 된다. 심리음향 분야는 이러한 청각적 지각 특성, 특히 내이에서의 시간 또는 주파수 성분들에 대한 해석 능력 등을 이해하는 분야로, 오디오 분야에서는 이러한 청각적 지각 특성을 이용하여 고음질, 고압축의 실현이 가능하다. 즉 청각적으로 무관한 정보는 인간이 지각할 수 없다는 특성으로서 신호의 음질을 개선하게 되는데, 이렇게 지각적으로 무관한 정보는 신호 분석 시 심리음향적 특징들, 즉, 절대 가청 한계(Absolute hearing threshold), 임계대역 주파수 분석(Critical band frequency analysis), 마스킹 현상(Masking phenomena)등을 이용하여 분석하게 된다^[13-15].

1. 임계대역 분석

인간의 청각 기관 중 내이의 기저막에서의 임의의 신호에 대한 주파수 분석은 기저막의 역학적인 특성에 영향을 받는다. 이러한 이유로 해서 입력 사운드의 각 주파수 성분들은 각각의 주파수 성분에 따라 기저막의 특정 위치에서 자극에 대한 반응을 보이게 된다. 일반적으로 임의의 두 개의 순음 성분의 주파수 차이를 천천히 변화시킬 때 그러한 변화의 차를 청자가 지각하게 되는 순간의 주파수 차이 폭을 임계대역(Critical band)이라고 한다. 임계대역 내에서는 두 주파수 성분이 존재하더라도 두 주파수를 갖는 신호가 아닌 하나의 주파수를 갖는 신호의 에너지로 인지한다. 이때 임계대역의 대역폭은 중심 주파수를 중심으로 저주파 영역에서는 대략 0.1kHz의 대역폭을, 고주파 영역에서는 대략 4kHz의 대역폭을 갖게 된다. 이러한 임계대역은 바크울(Bark ratio)로 표현하며, 심리 음향의 모든 연산은 이러한 바크 스케일(Bark scale)을 기준으로 이루어진다. 이러한 바크 울과 임계대역의 대역폭은 임의의 주파수 f에 대해서 다음 식 (1)과 (2)로 표현되어 진다^{[14][15]}.

※ Critical band ratio (Bark ratio)

$$z/Bark = 13\arctan(0.76f/kHz) + 3.5\arctan(f/7.5kHz)^2 \tag{1}$$

※ Critical band ratio

$$\Delta_{\omega_c}/Hz = 25 + 75[1 + 1.4(f/kHz)^2]^{0.69} \tag{2}$$

2. 절대 가청 한계

또 다른 인간의 청각 특성인 절대 가청 한계(Absolute hearing threshold)는 사람의 귀가 소리의 자극 세기에 대해 반응하여 인식하게 되기까지 모든 사람이 기본적으로 갖고 있는 저항치 이다. 즉 신호의 세기가 이러한 기본적인 저항치의 정도를 넘어서야만 그러한 자극을 지각할 수 있게 되는 것이다. 이러한 절대 가청 한계치는 임의의 주파수 f에 대해서 다음 식 (3)을 통해 표현되어 진다^[15].

$$T_q(f) = 3.64(f/1000)^{-0.8} - 6.5e^{-0.6(f/1000-3.3)^2} + 10^{-3}(f/1000)^4, \quad (dB SPL) \tag{3}$$

3. 마스킹 현상

임의의 두 가지 또는 그 이상의 신호가 함께 들릴 경우 일반적으로 임의의 신호가 다른 신호의 존재로 인해 들리지 않거나 감지하기 어렵게 되는 마스킹(masking)이라는 현상이 발생할 수 있다. 이때 마스킹 영향으로 감지할 수 없게 된 신호를 마스크(maskee)라 하고, 이러한 음의 인식에 영향을 미친 신호를 마스크(masker)라고 한다. 이와 같이 마스킹 현상은 임의의 신호 에너지가 기저막에서 강한 자극을 생성시켜 에너지가 약한 신호의 감지를 제약하는 현상이며 신호의 에너지와 매우 밀접한 관계를 갖고 있다. MPEG-1 Audio Layer I에서는 인간이 인지하지 못하는 정보를 생략하여 부호량을 절감하는 지각부호화를 수행하는 과정에서 Psychoacoustic Model-1을 이용하여 마스킹 곡선을 계산하는데 그 절차는 다음과 같다.

3.1 순음 성분과 비순음 성분의 검출

우선 생성된 입체음향의 각 채널에서 프레임 단위로 신호를 추출한 후, 푸리에 변환을 통해 전력 스펙트럼 밀도(PSD; Power Spectral Density)를 계산한다^[13].

$$\begin{aligned}
 X(k) &= \frac{1}{N} \sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi kn}{N}} \\
 P(k) &= 90.302 + 10\log_{10}|X(k)|^2 \\
 \text{where, } 0 \leq n \leq N-1, 0 \leq k \leq \frac{N}{2}
 \end{aligned}
 \tag{4}$$

이때 N은 1프레임의 샘플 수를 의미하며 w(n)은 해닝 윈도우(Hann Window) 함수를 사용한다.

다음으로 입체음향 신호의 마스킹 특성을 고려하기 위하여 각 프레임의 순음 성분 및 비순음 성분을 검출한다. 순음 성분은 신호의 주기적인 특성이 강하지만 비순음 성분은 잡음과 같은 특성을 가지고 있기 때문에 각각의 마스킹 특성이 틀리다. 따라서 전역 마스킹 임계치를 계산하기 위해 주파수 스펙트럼으로부터 순음과 비순음 성분들을 각각 구분지어 고려할 필요가 있다. 우선 임의의 주파수 인덱스 k에 대해서 P(k)>P(k±1)인 조건을 만족하는 P(k)를 local maximum으로 결정하고, 이렇게 결정된 local maximum 값들 중 Δ_k로 구분되어지는 구간 중 최소 7dB 이상 차이 나는 주파수 성분들을 순음 성분으로 간주한다. 여기서 Δ_k는 임계대역을 기반으로 한 주파수 범위이다.

$$\begin{aligned}
 S_T &= \{P(k)|P(k) > P(k \pm 1), P(k) \geq P(k \pm \Delta_k) + 7dB\} \\
 \text{where, } 0 \leq k \leq \frac{N}{2}
 \end{aligned}
 \tag{5}$$

이렇게 검출된 순음 성분들은 다음과 같이 이웃하는 순음 성분들을 이용하여 순음 마스커의 음압 레벨을 계산한다^[13]. 이 때 순음 성분 이외의 인덱스에 해당하는 순음 마스커의 음압 레벨은 0으로 정의한다.

$$P_{TM}(k) = 10\log_{10} \sum_{j=-1}^1 10^{\frac{P(k+j)}{10}} \quad (dB \text{ SPL}) \tag{6}$$

이어서 비순음 성분은 각각의 임계대역에 대해서 순음 성분이 검출된 주파수 인덱스 구간 ±Δ_k를 제외한 주파수 성분들을 합산하여 비순음 마스커(masker)의 음압 레벨로 계산한다. 이때 비순음 마스커의 인덱스 \bar{k} 는 각각의 임계대역에 기하평균(geometric mean)에 근접한 주파수 인덱스를 사용한다^[13]. 이 때 \bar{k} 이외의 인덱스에 해당하는 비순음 마스커의 음압 레벨은 0으로 간주한다.

$$\begin{aligned}
 P_{NM}(\bar{k}) &= 10\log_{10} \sum_j 10^{\frac{P(j)}{10}}, \quad (dB \text{ SPL}) \\
 \text{where, } \forall P(j) &\notin P_{TM}(k, k \pm 1, k \pm \Delta_k)
 \end{aligned}
 \tag{7}$$

3.2 마스커의 십진화(decimation) 및 재배열(reorganization) 수행

다음으로 전역 마스킹 임계치 계산 시 마스커의 수를 줄이기 위해서 decimation을 수행한다. decimation시 P_{TM,NM}(k) ≥ T_q(k)인 절대 가청 한계치 이상의 순음 및 비순음 마스커만 고려하며 각 임계대역에 대해 0.5 bark-width sliding window를 적용하여 큰 음압 레벨을 갖는 마스커만 고려한 후 다음과 같은 sub-sampling scheme을 적용하여 마스커의 수를 감소시킨다.

$$\begin{aligned}
 P_{TM,NM}(i) &= P_{TM,NM}(k) \\
 P_{TM,NM}(k) &= 0 \\
 \text{where, } i &= \begin{cases} k & 1 \leq k \leq 48 \\ k + (k \bmod 2) & 49 \leq k \leq 96 \\ k + 3 - ((k-1) \bmod 4) & 97 \leq k \leq 232 \end{cases} \tag{8}
 \end{aligned}$$

이러한 sub-sampling은 Bark index 18~22의 임계대역에 대해 2:1로 decimation을 취하는 것과 같고, Bark index 22~25의 임계대역에 대해서는 4:1로 decimation을 취하는 것과 같다. 그림 1은 임의의 1프레임 사운드 신호에 대한

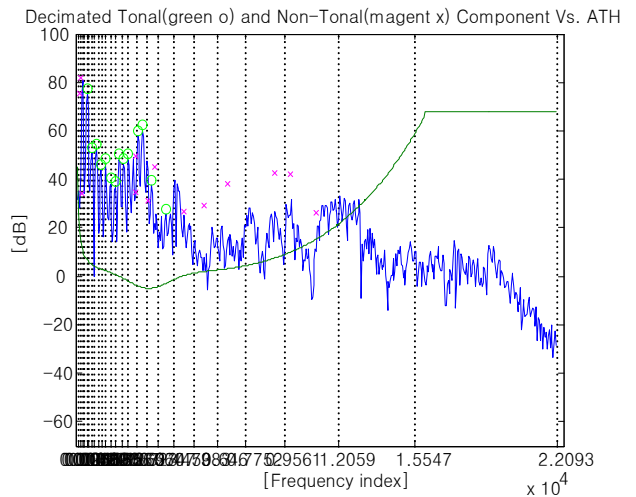


그림 1. 순음 및 비순음 성분의 검출
Fig. 1. Detection of tonal & non-tonal components

순음 마스크(o 표시) 및 비순음 마스크(x 표시) 검출 결과를 나타낸다.

3.3 개별 마스크 임계치 계산

decimation 과정과 reorganization 과정을 거쳐 얻어진 순음 및 비순음 마스크의 개별 마스크 임계치(Individual masking threshold)를 계산한다^{[13][14]}.

※ 순음 마스크의 개별 임계치

$$T_{TM}(i,j) = P_{TM}(j) - 0.275z(j) + SF(i,j) - 6.025, \quad (dB \text{ SPL}) \quad (9)$$

※ 비순음 마스크의 개별 임계치

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025, \quad (dB \text{ SPL}) \quad (10)$$

여기서 $P_{TM}(j)$ 과 $P_{NM}(j)$ 은 각각의 식 (6)과 (7)에서 결정된 마스크들에 대한 임의의 주파수 인덱스 j 에서의 순음 마스크와 비순음 마스크의 음압 레벨을 나타낸다. $z(j)$ 는 주파수 인덱스 j 에 대한 식 (1)에 의해 계산되어진 바크 율(Bark ratio)이다. $SF(i,j)$ 는 주파수 인덱스 j 의 마스크에 의한 주파수 인덱스 i 의 성분들에 대한 지각적 에너지 확산의 영향을 나타내는 확산함수(Spreading function)이다. 이러한 $SF(i,j)$ 는 다음과 같이 표현되어 진다^{[13][14]}.

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025, \quad (dB \text{ SPL}) \quad (10)$$

이 때 $\Delta_z = z(i) - z(j)$ 는 마스크의 주파수 인덱스 j 와 마스크 주파수 인덱스 i 에 대한 바크율로 표현된 주파수 간격이다.

$$SF(i,j) = \begin{cases} 17\Delta_z - 0.4P_{TM,NM}(j) + 11, & -3 \leq \Delta_z < -1 \\ (0.4P_{TM,NM}(j) + 6)\Delta_z, & -1 \leq \Delta_z < 0 \\ -17\Delta_z, & 0 \leq \Delta_z < 1 \\ (0.15P_{TM,NM}(j) - 17)\Delta_z - 0.15P_{TM,NM}(j), & 1 \leq \Delta_z < 8 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

$$T_g(i) = 10 \log_{10} \left(10^{\frac{T_q(i)}{10}} + \sum_{c=1}^C 10^{\frac{T_{TM}(i,c)}{10}} + \sum_{m=1}^M 10^{\frac{T_{NM}(i,m)}{10}} \right) \quad (12)$$

그림 2는 검출된 순음과 비순음 마스크들의 개별 마스크 임계치(점선)을 나타낸다.

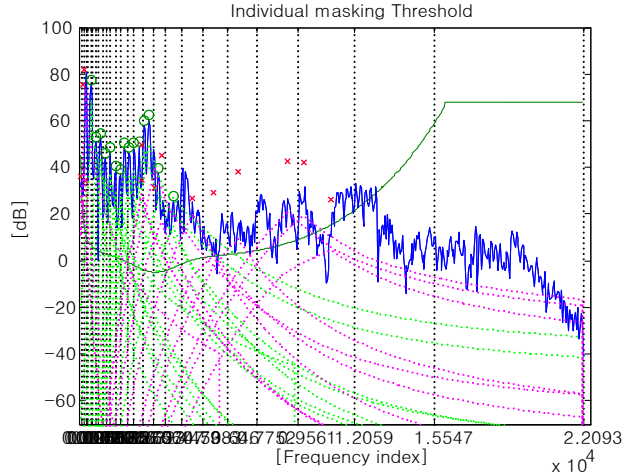


그림 2. 개별 마스크 임계치
Fig. 2. Individual masking threshold

3.4 전역 마스크 임계치 계산

이와 같이 순음 성분과 비순음 성분의 마스크 영향을 계산한 후, 계산되어진 각각의 마스크 임계치와 절대 가청 한계를 고려하여 다음과 같이 전역 마스크 임계치(GMTH; Global masking threshold), $T_g(i)$ 를 계산하게 된다. C 는 순음 성분의 개수를, M 은 비순음 성분의 개수를 나타낸다^{[13][14]}. 여기서 $T_g(i)$ 는 식 (3)의 discrete version이며 음성은 시간에 따라 주파수 특성이 변하므로 C 와 M 은 매 프레임마다 가변적인 값을 갖는다.

이러한 각 주파수 성분들에 의한 전역 마스크 임계치는 주파수 영역에서의 모든 순음과 비순음 성분에 의한 기저막에서의 자극에 의한 절대 가청 한계의 신호 의존적인 변

형이라고 할 수 있다. 그림 3은 임의의 신호에 대한 임계대역간의 이러한 순음 성분과 비순음 성분의 마스킹 영향과 절대 가청 한계치의 영향에 의한 전역 마스킹 임계치(굵은 선)을 나타내고 있다.

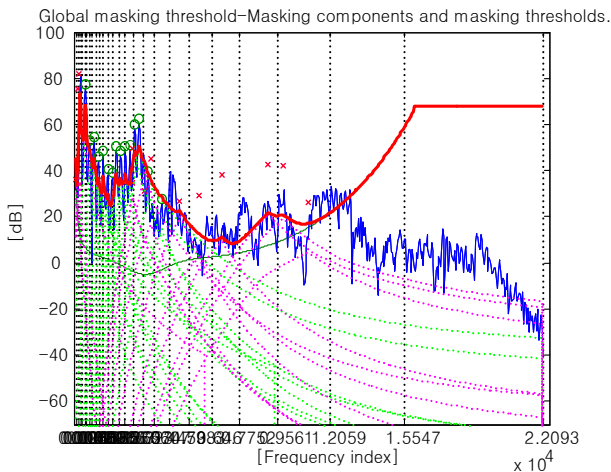


그림 3. 전역 마스킹 임계치
Fig. 3. Global masking threshold

III. 입체음향

소리가 3차원 공간상의 어떤 위치에서 방사될 때 청취자는 방사된 소리를 듣고 음원의 위치 및 자신이 속한 공간의 특성을 인지하게 된다. 이 때 청취자의 청각기관에 도달하는 음을 직접 녹음한 후 헤드폰을 이용하여 재생시켜 준다면 그 공간에 속해 있지 않더라도 마치 그 동일 공간 안에 존재하는 것처럼 느끼게 된다. 이것을 바이노럴 재생이라고 하며 더미헤드 또는 실제 사람의 귀에서 녹음하고 헤드폰으로 재생하는 방법을 말한다^[4].

사람이 방향 지각을 인지할 때 일반적으로 양이 단서(binaural cue)와 편이(여과) 단서(Monastral(filtering) cue)를, 그리고 거리 지각을 인지할 때 음량감(Loudness perception), 확산감(Spectral perception) 및 잔향감(Reverberation perception)을 이용한다^[3]. 이중 이론(Duplex theory)으로도 불리는 양이 단서는 청취자의 방위각(azimuth)에서의 정위를 인지하는 단서로써 두 귀 사이의 시간차

(Interaural Time Difference; ITD)와 두 귀 사이의 음압차(Interaural Intensity Difference; IID)로 구분되며 ITD는 중저역대(100~1.5kHz; 100Hz 이하는 무지향성)에, IID는 고역대(1.5kHz~)에 영향을 끼치게 된다^[3]. 즉 이를 자세히 풀어 보면 음파의 중요한 성질 중 하나는 바로 자신의 파장보다 긴 길이의 물체에서는 멈춰지고 짧은 길이의 물체에서는 비켜 나아간다는 것이다. 1500Hz 언저리의 음파는 대략 머리의 직경(약 20cm)과 같은 파장을 가지고 있다. 이로 인해 파장이 머리의 직경보다 긴, 약 1500Hz 이하의 저역에서는 음파는 머리 직경에 따른 음여과 현상에 영향을 받지 않으므로 양 귀 사이의 위상차(Phase Difference)가 발생하여 양이 단서 인지의 주 역할을 하게 된다. 그러나 파장이 머리 직경보다 짧아지는 1500Hz 이상에서는 위상차가 발생하지 않으므로 ITD의 단서가 모호해지게 된다. 따라서 IID가 방위각의 인지를 대신하게 된다. 이는 고음이 머리를 돌아 나아가지 못하고 멈춰지게 되어 머리의 음여과 현상(Head Shadow)에 의한 시간차가 점점 커지게 되는 것이다^[16]. 이밖에 편이 단서는 귓바퀴, 머리, 어깨, 몸통 등 청취자의 신체 부위가 특정음을 여과하는 현상을 말한다. 이런 현상에 관한 정보들은 자유 공간 상에서 음원과 귀 간의 경로를 모델링하는 함수인 머리전달함수(Head Related Transfer Function; HRTF)에 내포되어 있다. 머리전달함수는 어떠한 특정 점에서의 음향신호가 귀까지 도달하는 전달 경로의 특성으로서 음상의 공간화와 관련된 중요한 음향 파라미터로 사용된다^[16]. 이런 특성으로 인하여 머리전달함수는 가상음향에서 빼놓을 수 없는 역할을 하며 이를 취득하게 되면 어떠한 음원이라도 머리전달함수와 컨벌루션 연산을 통하여 청취자에게 가상음원의 효과를 줄 수 있다^{[3][4]}.

$$\begin{aligned} Y_L &= S * hrtf_L(\theta, \phi) \\ Y_R &= S * hrtf_R(\theta, \phi) \end{aligned} \tag{13}$$

여기서 θ 와 ϕ 는 특정한 방위각과 고도각을, *는 컨벌루션 연산을 의미한다. 그리고 S는 모노음원, Y_L 과 Y_R 은 각 채널에 따라 생성된 입체음향 신호를 나타내고 있다.

IV. 제안된 알고리즘

일반적으로 여러 방향 지각 단서들 중 가장 많은 영향을 미치면서 직관적으로 제어가 가능한 요소로는 각 방향에 따른 주파수 스펙트럼 특성의 변화라 할 수 있다. 따라서 바이노럴 시스템에서 방향 지각의 혼돈을 개선하기 위하여 각 방향에 대한 스펙트럴 큐를 강조하는 연구가 주를 이루고 있다. 그러나 기존의 연구들은 머리전달함수의 다양한 특성을 제대로 고려하지 않거나 모든 인간의 신체 특성을 모델링하지 않기 때문에 정확한 음상의 정위가 불가능하게 되는 주요 원인이 된다.

이에 본 논문에서는 머리전달함수의 주파수 단서뿐만 아니라 생성된 입체음향의 지각적인 특성까지 고려하고 이를 바탕으로 하여 방향감을 강조하는 알고리즘을 제안하고자 한다. 제안된 알고리즘의 Block diagram은 그림 4와 같다.

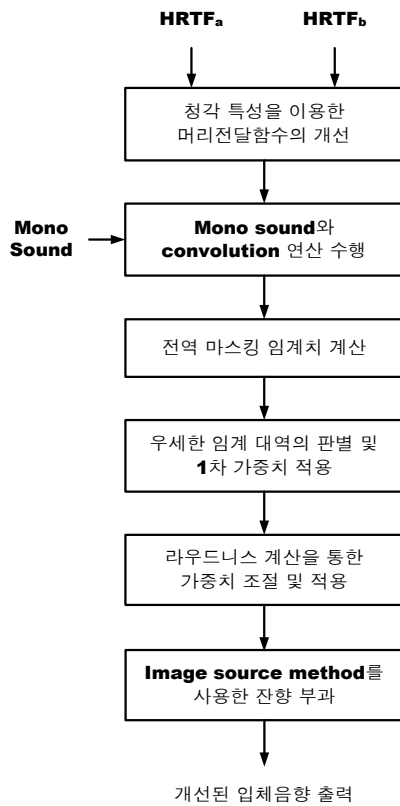


그림 4. 제안된 알고리즘의 순서도
Fig. 4. Block diagram of the proposed algorithm

1. 청각특성을 이용한 머리전달함수의 개선

우선 생성된 입체음향의 각 채널에서 프레임 단위로 신호를 추출한 후, 푸리에 변환을 통해 전력 스펙트럼을 계산한다. 이를 $X_p^i(k)$ 라고 정의 할 때, 신호의 전력 스펙트럼으로부터 외이(外耳)에서 중이(中耳)까지 전송을 고려한 임계대역에서의 전체 신호 에너지를 나타내는 임계대역 에너지는 식 (14)와 같이 표현할 수 있다^[4].

$$X_u^i(z, q) = a_0(z, q) \sum_{k=k_{lz}}^{k_{hz}} X_p^i(k, q), \quad (\text{for}, 0 \leq z \leq Z-1) \quad (14)$$

이 때, z 는 바크 인덱스(Bark index)를 나타내며, $a_0(z, q)$ 는 외이로부터 중이까지의 다양한 전달 특성 요소(Transformation factor)들에 의한 주파수 의존 감쇠 특성을 나타내는 감쇠 성분이다. k 와 i 는 주파수 축과 프레임 윈도우의 인덱스를, k_{lz} , k_{hz} 는 임의의 임계대역 z 에서의 주파수 경계를 나타낸다. 그리고 Z 는 전체 임계대역의 수를, q 는 좌(L)/우(R) 채널을 나타낸다. 이러한 임계대역 에너지는 인접 대역을 비롯한 다른 대역에 마스킹 영향을 끼치게 되는데 이러한 마스킹 확산현상을 고려하기 위해 확산 함수, $SF(v, z)$ 를 통해 컨벌루션 연산을 수행함으로써 신호의 에너지에 대한 각각의 임계대역에서의 청각 자극 에너지를 나타낼 수 있다^[4].

$$X_e^i(z, q) = \sum_{v=0}^{Z-1} [SF(v, z) X_u^i(v, q)], \quad (\text{for}, 0 \leq z \leq Z-1) \quad (15)$$

여기서 확산 함수 $SF(v, z)$ 는 식 (13)과 같이 표현되어질 수 있으며, $\Delta_z = z-v$ 는 바크 인덱스로 표현된 임계대역 간 차이이다^[4].

$$SF(v, z) = \frac{15.81 + 7.5(\Delta_z + 0.474)}{17.5 \sqrt{1 + (\Delta_z + 0.474)^2}} \quad (16)$$

이런 특성을 이용하여 음상을 정위하고자 하는 a방향의 머리전달함수, $HRTF_a$ 및 이와 혼돈 원추상에 있는 b방향의 머리전달함수, $HRTF_b$ 를 이용하여 각 임계대역, z 에서의 청

각 자극 에너지, $energy_a(z,q)$, $energy_b(z,q)$ 를 계산한다^[4]. 이때 a방향의 혼돈원추는 상/하, 전/후의 2경우로 발생하는데 본 논문에서는 전/후 지각과 관련한 혼돈원추의 경우로 한정한다.

다음으로 계산된 a와 b방향의 청각 자극 에너지를 이용하여 가중치를 구한다^[4]. 머리전달함수는 저주파 부분에서 각 방향의 주파수 스펙트럼 차이가 크지 않으므로 청각 자극 에너지의 차이도 크지 않지만 고주파 부분에서는 주파수 스펙트럼의 차이가 커지므로 인간의 귀에 끼치는 영향도 달라진다. 따라서 계산된 가중치는 원 $HRTF_a$ 주파수 스펙트럼에 적용되어 각 방향에 따른 우세한 임계대역을 증폭시키고 열세한 임계대역을 감소시킨다. 이렇게 각 임계대역을 조절한 머리전달함수는 모노음과의 컨벌루션 연산을 통하여 청취자에게 1차적으로 방향감이 강조된 입체음향, $pro1_a$ 를 생성한다.

$$pro1_a = S * enHRTF_a(\theta, \phi) \quad (17)$$

여기서 θ 과 ϕ 는 특정한 방위각과 고도각을, *는 컨벌루션 연산을 의미한다. 그리고 S는 모노음원을 나타낸다.

2. 인지 가능한 주파수 특성 강조에 따른 방향성 강조

이어서 생성된 입체음향에 지각 특성 해석 기법을 통해 사람의 기저막에서의 주파수 분석을 나타내는 지각 영역에서의 마스킹 특성을 해석하고 전역 마스킹 임계치를 계산한다. 인간의 청각 특성에서 마스킹은 사람의 음 지각과 무관한 신호성분을 검출할 수 있게 하기 때문에 지각 오디오 코더 설계에 있어 가장 중요한 심리음향 현상 중 하나이다. 마스킹은 마스커와 마스크의 스펙트럼 구조와 시간 특성, 신호의 세기 등에 의존하며 각 프레임에서 고유한 형태를 가진다. 이에 정위된 음상의 좌/우측 채널 신호가 각각 인간의 청각에 미치는 영향을 고려한 후, 그에 따라 2차적으로 방향감을 부각시키는 작업을 수행한다.

먼저 $pro1_a$ 의 각 채널에서 frame 단위로 신호를 추출한 후, MPEG-1 Audio Psychoacoustic Model-1을 이용하여 전역 마스킹 임계치를 계산한다.

그림 5와 6은 고도각 0°, 방위각 45°/135°에 정위된 입체음향의 우측 채널에서 추출된 임의의 한 프레임의 전역 마스킹 임계치를 나타내고 있다. 이 방위각들은 IID와 ITD가 같은 혼돈 원추 상에 위치하지만 이 두 방위각에 정위된 음상들은 머리카락에 따라 서로 다른 영향을 받게 되어 지각적인 특성이 달라진다. 인간의 심리음향 특성과 관련한 많은 연구에서 살펴보면 청자는 각 주파수 대역에서 전역 마스킹 임계치보다 큰 신호만을 인지할 수 있다^{[14][15]}. 따라서 인간이 지각할 수 있는 에너지가 많은 임계대역이 인간의 청각에 보다 많은 영향을 끼친다고 생각할 수 있다. 이에 각 임계대역의 인지 가능한 신호들의 에너지를 합산하여 전/후 혼돈 방향에 따른 우세한 임계대역을 추출할 수 있다. 인지 가능한 에너지는 전력 스펙트럼 밀도와 전역 마스킹 임계치의 차로써 계산된다. 만약 임의의 인덱스, k에서 전력 스펙트럼 밀도가 전역 마스킹 임계치보다 클 경우에는 인간이 인지 가능한 에너지라 판단할 수 있지만, 전력 스펙트럼 밀도가 전역 마스킹 임계치보다 작거나 같을 경우에는 청각 특성상 인지를 할 수 없는 부분이기 때문에 0의 값을 가지게 된다.

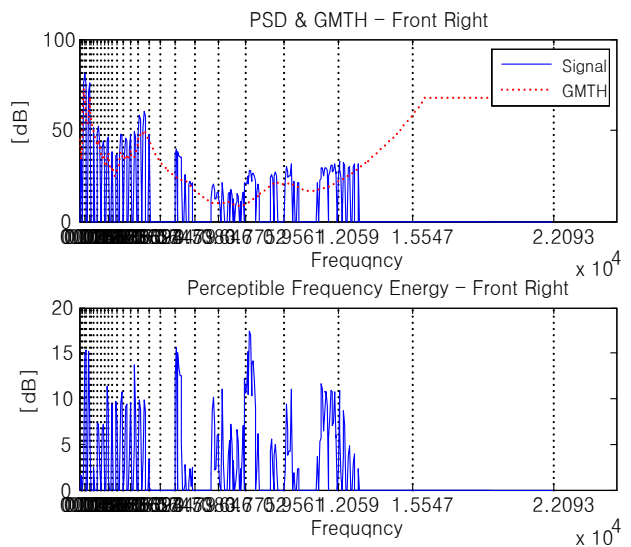


그림 5. 계산된 에너지(고도각 0°, 방위각 45°), (a) 전력스펙트럼밀도 & 전역 마스킹임계치, (b) 각 임계대역별 인지 가능한 주파수 에너지
Fig. 5. Calculated energy(Ele. 0°, Azi. 45°), (a) PSD & GMTH, (b) Perceptible frequency energy in each critical band

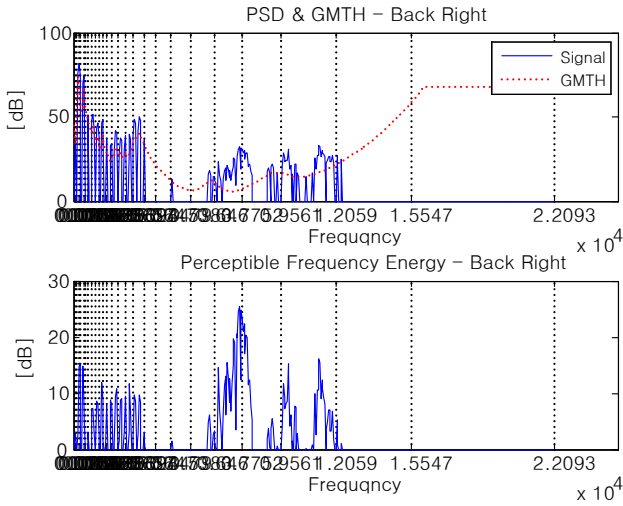


그림 6. 계산된 에너지(고도각 0°, 방위각 135°), (a) 전력스펙트럼밀도 & 전역 마스크임계치, (b) 각 임계대역별 인지 가능한 주파수 에너지
 Fig. 6. Calculated energy(Ele. 0°, Azi. 135°), (a) PSD & GMTH, (b) Perceptible frequency energy in each critical band.

$$Y_u^i(z, q) = \sum_{k=k_{l_z}}^{k_{h_z}} d^i(k, q) \quad (18)$$

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

$$d^i(k, q) = \begin{cases} P^i(k, q) - T_g^i(k, q), & (P^i(k, q) > T_g^i(k, q)) \\ 0, & (P^i(k, q) \leq T_g^i(k, q)) \end{cases} \quad (19)$$

최종적으로 계산된 a방향의 채널당 임계대역별 합산 에너지는 $Y_a^i(z, q)$ 로 정의되며 이어 동일한 방법으로 혼돈 방향인 b방향의 채널당 임계대역별 합산 에너지, $Y_b^i(z, q)$ 를 계산한다.

그림 7은 고도각 0°에서 방위각 45°/135° 방향의 인지 가능한 주파수 에너지들의 합을 나타낸다. Duplex theory에 따르면 주파수 영역에서 1.5kHz 이하의 대역에서는 IID보다 ITD가 주된 단서가 되기 때문에 1.5kHz 이하의 대역에 해당하는 Bark index 15이하의 임계대역에서는 인지 가능한 에너지도 비슷하다. 그러나 1.5kHz 이상의 대역에서는 IID가 방향 인지의 단서가 되므로 각 방향에 따라서 인지할 수 있는 에너지의 크기가 달라지고, 그에 따라 각 방향의 우세한 대역이 존재해 있음을 볼 수 있다. 이는 생성된 입체 음향이 마스크 등의 인간의 청각 특성으로 인하여 그 특성

이 달라지고 이웃 임계대역에 영향을 받기 때문이다. 더불어 Bark index 25의 임계대역은 인간의 절대 가청 한계 특성상 인지를 할 수 없는 대역이기 때문에 총합은 0이 된다.

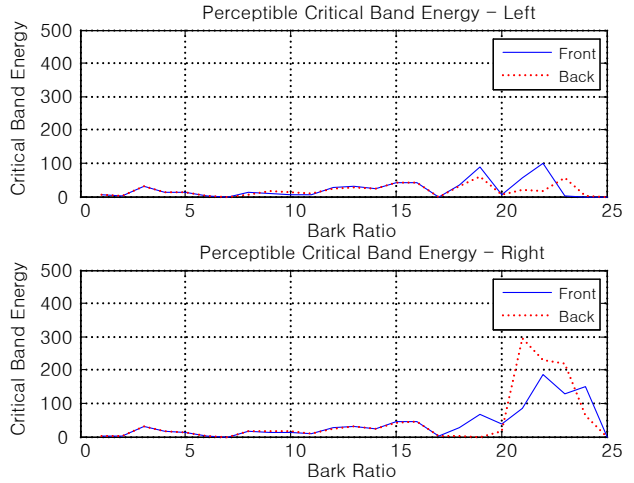


그림 7. 임계대역별 합산 에너지(고도각 0°, 방위각 45° & 135°), (a) 좌측 채널 (b) 우측 채널
 Fig. 7. Summation of energy in each critical band(Ele. 0°, Azi. 45° & 135°), (a) Left channel, (b) Right channel.

이어 인지 결과에 따라 각 임계대역에 가중치, $weight_1$ 을 적용함으로써 인간의 청각에 미치는 에너지를 조절한다. 일반적으로 인간의 두 귀에 도달하는 입체음향은 방향과 시간에 따라서 주파수 특성이나 세기가 변화하기 때문에 각 프레임마다 청각에 미치는 영향이 달라진다. 따라서 이러한 청각적인 영향을 강조한다면 청치는 해당 방향의 느낌을 강하게 지각할 수 있게 된다. 따라서 가중치를 통해 각 임계대역을 조절함으로써 혼돈 방향과의 청각적 자극 에너지 차이 및 음색을 부각시키게 된다. 가중치는 음이 손상되지 않는 범위 내에서 실험적으로 추출된 1.1에서 3사이의 실수 값으로 정의된다.

$$pro2_a^i(k, q) = X^i(k_z, q) \times weight_1, \quad (Y_a^i(z, q) > Y_b^i(z, q)) \\ X^i(k_z, q) / weight_1, \quad (Y_a^i(z, q) < Y_b^i(z, q)) \quad (20)$$

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

여기서 k_z 는 각 임계대역에 해당하는 주파수 범위의 인덱스를 나타낸다.

본 논문에서 제안된 방법은 3차원 공간상에 정위된 음상에서 혼돈 방향보다 강하게 지각할 수 있는 대역을 추출하고 가중치를 통해 각 대역의 에너지를 조절함으로써 방향감을 강조한다. 그러나 만약 가중치가 매우 크거나 증폭되는 대역에 주파수 에너지가 과다하게 존재할 경우 방향지각과 관련한 단서가 달라져 정확한 음상의 정위가 불가능하며 음질까지 손상되어 청자에게 불쾌감을 주는 경우가 발생할 수도 있다. 이에 라우드니스 계산을 통하여 가중치를 가변적으로 조절하는 단계를 거친다. 주관적으로 느끼는 소리의 크기로 정의되는 라우드니스는 그 자체로서 음질을 결정하는 중요한 인자이며, 샤프니스, 러프니스, 변동강도와 같은 다른 음질 인자 발생의 원인을 설명하는데 필요한 기본적인 청감이다. 본 연구에서는 Zwicker에 의하여 제안된 라우드니스 계산 방법을 사용하는데, 이는 Stevens에 의해 제안된 방법보다 복잡하긴 하지만 확산 음장(diffuse sound field) 및 자유 음장(free sound field)에 모두 적용 가능하고, 순음 성분이 있거나 불연속적인 형태의 주파수 특성을 갖는 음의 해석에 유용하기 때문이다^[17]. 라우드니스(N: 단위 sone)는 각 임계대역에 대한 비라우드니스(Specific loudness: 단위 sone/Bark)를 모두 더한 값이 되는데, 라우드니스는 소리의 전체 크기를 나타내는 반면, 비라우드니스는 각 임계대역에 대한 크기를 나타내기 때문에 소리의 특성을 주파수 대역과 대응시켜 분석할 수 있는 장점이 있다.

머리전달함수의 개선을 통해 1차적으로 방향감이 개선된 입체음향, pro1_a와 전역 마스킹 임계치를 이용하여 2차적으로 방향감이 개선된 입체음향, pro2_a의 각 프레임 별 비라우드니스, $lou_1^i(z,q)$, $lou_2^i(z,q)$ 를 계산한다. 이어 임의의 임계대역에서 $lou_2^i(z,q)$ 가 $lou_1^i(z,q)$ 보다 설정 임계치, threshold 이상 클 경우 해당 임계대역의 가중치를 $weight_2$ 로 변경하여 2차 개선 절차를 다시 수행한다. 이 과정을 통해 방향지각 단서 중의 하나인 IID의 변형을 최대한 방지하여 원하는 방향에 개선된 입체음향을 정위할 수 있게 된다.

식 (21)은 비라우드니스를 이용한 임계대역의 가중치 조절을 정의하고 있으며, $weight_1 > weight_2$ 로 정의된다.

그림 8과 9는 원 머리전달함수만을 사용해 생성된 입체음향과 본 알고리즘을 통해 출력된 입체음향에서 임의로 추출된 우측 채널 한 프레임의 비라우드니스를 나타낸다.

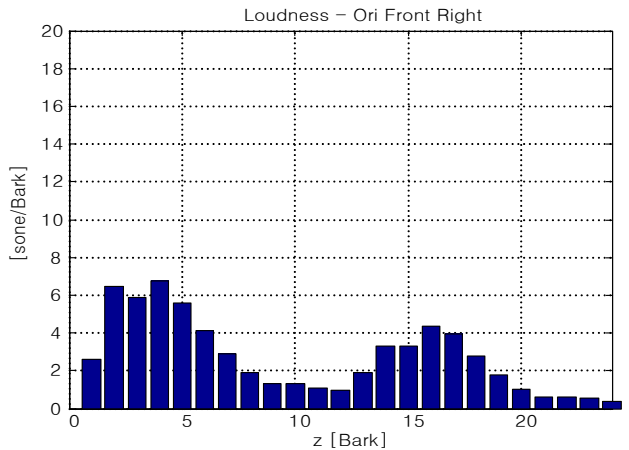


그림 8. 비라우드니스 해석(원머리전달함수 이용, 고도각 0°, 방위각 45°)
Fig. 8. Specific loudness analysis(applied original HRTF, Ele. 0°, Azi. 45°)

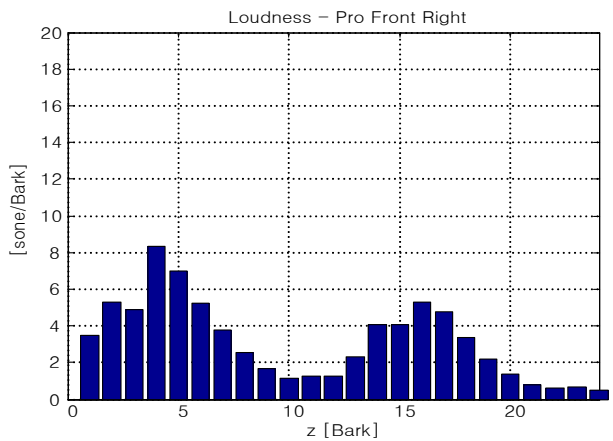


그림 9. 비라우드니스 해석(제안된 방법 이용, 고도각 0°, 방위각 45°)
Fig. 9. Specific loudness analysis(applied proposed method, Ele. 0°, Azi. 45°)

$$\begin{aligned}
 final_a^i(k,q) = & X^i(k_z,q) \times weight_1, \quad (Y_a^i(z,q) > Y_b^i(z,q) \text{ 및 } (lou_2^i(z,q) - lou_1^i(z,q)) < threshold) \\
 & X^i(k_z,q) \times weight_2, \quad (Y_a^i(z,q) > Y_b^i(z,q) \text{ 및 } (lou_2^i(z,q) - lou_1^i(z,q)) > threshold) \\
 & X^i(k_z,q) / weight_1, \quad (Y_a^i(z,q) < Y_b^i(z,q)) \\
 & (where, 0 \leq z \leq Z-1)
 \end{aligned} \tag{21}$$

계산된 비라우드니스 패턴을 살펴보면 제안된 방법의 비라우드니스에서 Bark index 4~8과 16~20의 임계대역은 전방향의 에너지가 우세한 대역이므로 원 머리전달함수 방법에 비해 증가되었음을 보여준다. 이를 통해 음의 재생 대역이 넓고 풍부해져 콧바퀴의 영향을 거의 받지 않는 전방향의 음색을 강조한다. 이와는 반대로 후방향에 음상을 정위시킬 경우에는 비라우드니스가 감소하여 음이 콧바퀴의 영향을 받는 효과를 더해지게 될 것이다.

이어 개선된 입체음향에 외재화 효과를 부여하기 위해 Image source method를 이용하여 인공 잔향효과를 생성한다^[18]. 즉 머리전달함수에 의해 방향 정보가 부가된 음원에 Image source method를 이용하여 생성한 룬 임펄스 응답을 처리함으로써 직접음과 함께 인공적으로 생성한 잔향이 더해지게 된다. 이를 통해 음상이 청자의 머리 밖에 맺히게 되므로 입체음향의 현실감을 상승시키게 된다. 본 논문에서 사용된 잔향은 20m × 20m × 3m 크기의 공간에서 음원이 10m, 18m, 1.5m, 왼쪽 마이크 및 오른쪽 마이크가 9.8m, 10m, 1.5m 지점 및 10.2m, 10m, 1.5m에 위치하며, 반사율 차수와 반사 계수는 각각 24와 0.3인 환경을 가정하여 생성하였다.

본 논문에서 제안한 방법을 통하여 최종적으로 생성된 입체음향은 각 방향에 대하여 유동적으로 청각에 영향을 미치는 부분만을 제어한다. 그 결과로 혼돈 방향과의 청각 에너지 차이 및 음색이 부각되는 효과를 얻게 되어 청자는 해당 방향의 방향감을 강하게 느낄 수 있게 된다.

V. 실험 및 결과 고찰

1. 실험 환경

본 논문은 청각 자극 에너지와 전역 마스킹 임계치를 이용하여 방향감을 강조하기 위한 알고리즘을 제안한다. 제안된 알고리즘의 성능을 확인하기 위하여 헤드폰 환경에서의 시뮬레이션 및 테스트를 수행하였다. 테스트는 일반 생활 잡음이 존재하는 연구실 공간으로 피실험자들은 중앙에 설치된 의자에 편안하게 앉도록 하였다. 더불어 조명을 끄거

나 눈을 가림으로서 실험에 영향을 줄 수 있는 외부 자극을 차단하였다. 사용된 신호들은 일반 영어 관련 CD에서 추출된 44.1kHz, 16bits/sample 음성 신호를 사용하였다. 사용된 머리전달함수로는 CIPIC에서 측정된 것^[19]을 사용하였으며 해닝 윈도우(Hann window) 함수와 오버랩애드(Overlap Add) 방식을 주파수축 변환에 사용하였다. 테스트 장비로는 청취자의 귀에 밀착되는 헤드폰(audio-technica ATH-ES7), Notebook computer(i3 core), 스피커 등을 사용하였다. 그리고 재생시스템과 피실험자 간에 거리를 두고 랜덤한 순서로 입체 음향을 들려주어 청취자가 음원의 종류나 방향을 사전에 알 수 없도록 하였다.

표 1. 사용된 머리전달함수

Table 1. Head related transfer function for simulations

샘플 수	200
샘플링 주파수	44.1 kHz
샘플당 비트수	16 bits/sample
수평각	-80° ~ 80°
고도각	-45° ~ 231°
측정 대상	2 KEMAR dummy head, 43 human subjects

비교 대상은 원 머리전달함수 및 기존 방법으로 Tan과 Gan이 제안한 머리전달함수의 필터뱅크를 이용한 방식^[5] 그리고 제안한 방식을 사용하였다. 그리고 정면을 방위각 0°, 오른쪽 90°, 뒷면 180°, 왼쪽을 270°로 가정할 때, 0°를 기준으로 하여 좌/우측이 서로 대칭이다. 따라서 채널만 서로 바꾸어 재생한다면 대칭 방향에 음상을 정위시킬 수 있다. 더불어 두 방향에 정위된 음상의 각도 차이가 작을 경우에 청자는 동일한 각도로 인지한다. 이에 본 논문에서는 우측 방향에 정위된 음상 중 청자가 그 차이를 명확히 인지할 수 있도록 각 경우마다 30°의 편차를 두었다. 그리고 정밀한 결과를 확인하고자 45°/135°의 경우를 추가하여 전체적으로 방위각 0°/180°, 30°/150°, 45°/135°, 60°/120°의 네 경우에 대해서 테스트를 실시하였다.

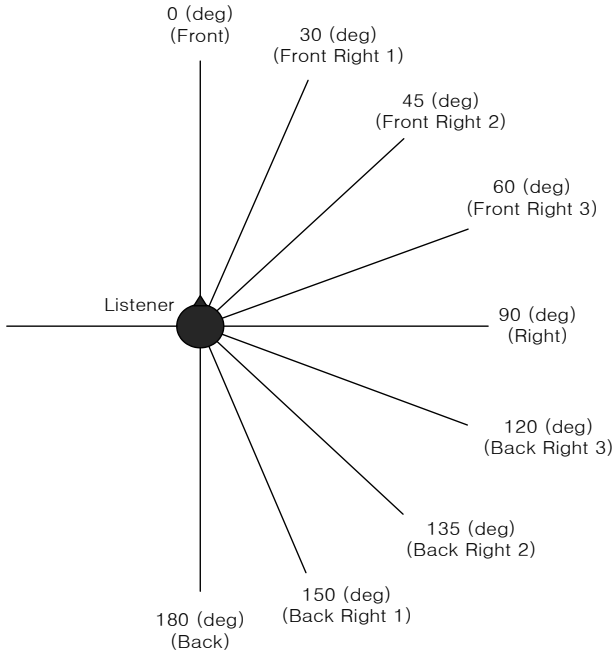


그림 10. 입체음향을 생성하기 위한 방위각의 분포
 Fig. 10. Geometrical arrangement of virtual sources and listener

입력 신호로는 인간의 음성 신호를 3차원 공간상에 정위시키는 상황에서의 성능을 확인하기 위해 표 2와 같이 3가지 음성 신호를 사용하였다. 이 신호들은 가청 주파수 대역 전체에 걸쳐서 에너지가 분포하고 있으며 시간에 따라 신호가 변화하기 때문에 다양한 경우에 대하여 테스트를 사용하기에 알맞다.

표 2. 사용된 음원
 Table 2. Vocal sources for simulations

Source No.	No. 1	No. 2	No. 3
이름	남성 음성	여성 음성 1	여성 음성 2
샘플링 주파수	44.1 kHz		
샘플 당 비트수	16 bits/sample		
샘플 수 / 샘플 길이(시간)	49140 samples / 1.11 초		

테스트에 참가한 피실험자는 입체음향에 사전 지식이 없는 2~30대의 일반 남녀 10명이며 실험 전에 헤드폰과 스피

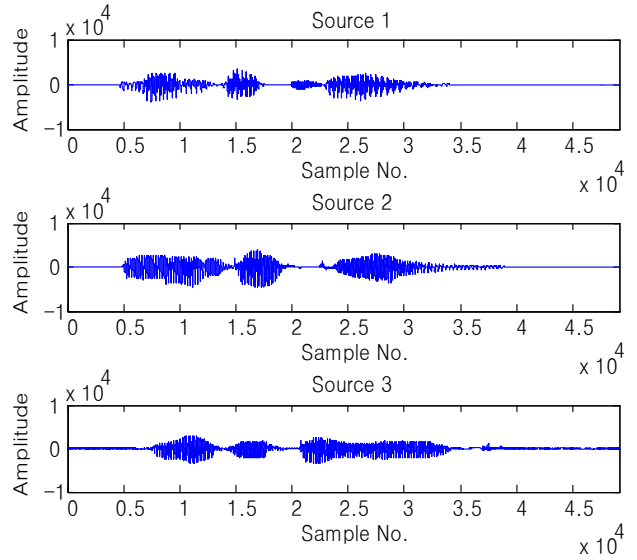


그림 11. 시뮬레이션에 사용된 음성 신호
 Fig. 11. Voice signals for simulations

커를 이용하여 간단한 샘플을 들려주고 이에 대한 간단한 설명을 진행하였다. 이어 헤드폰 환경에서 각 방법을 통해 생성된 입체음향을 재생하고 인지한 음상에 대해 구두로서 답하도록 하였다.

2. 인지도 테스트

본 논문은 바이노럴 재생 환경에서 음상을 정위하고자 할 때, 인간의 심리음향 특성을 이용하여 방향감을 강조하는 알고리즘을 제안한다. 이에 제안된 알고리즘의 성능을 확인하기 위하여 헤드폰 환경에서 선호도 테스트를 수행하였다. 먼저 원 머리전달함수와 기존의 방법에도 공간감을 주기 위해 제안한 방법과 동일한 환경으로 Image source method를 사용하여 인공 잔향을 부가하였다^[18]. 이어 앞에서 말한 3가지 방법에 따라 생성된 각 방향의 입체음향을 들려준 후, Kim의 평가 방법^[6]을 참조하여 전/후 각각의 경우에서 방향감이 가장 많이 개선되었다고 인지하는 방법을 선택하도록 하였다.

그림 12, 13 그리고 14는 세 방법을 이용하여 생성된 입체음향에 대한 선호도 테스트 결과를 나타낸다.

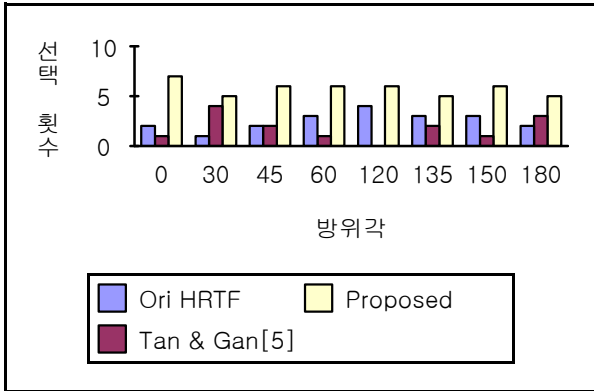


그림 12. 선호도 테스트 결과(남성 음성)
Fig. 12. Preference test results(Male voice)

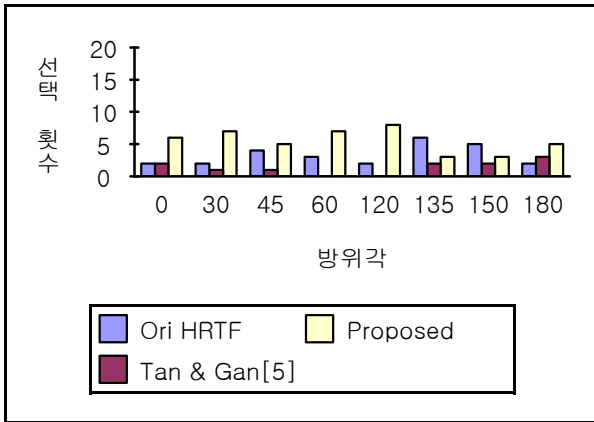


그림 13. 선호도 테스트 결과(여성 음성 1)
Fig. 13. Preference test results(Female voice 1)

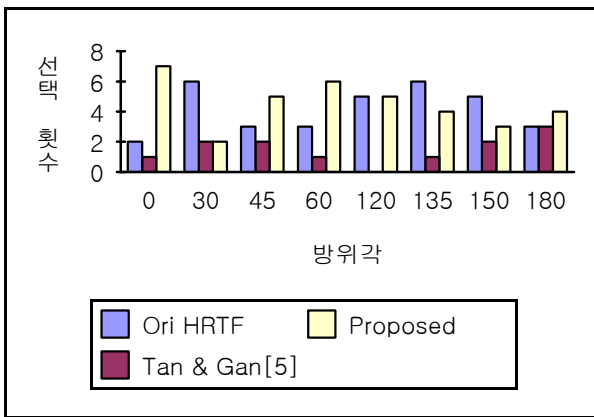


그림 14. 선호도 테스트 결과(여성 음성 2)
Fig. 14. Preference test results(Female voice 2)

전방향과 관련한 테스트 결과를 살펴보면 남성 음성의 경우에는 대부분의 청자들이 제안된 알고리즘을 이용하여 생성한 입체음향을 선호하였다. 여성 음성의 경우에서도 음원마다 약간의 차이는 있으나 방향감이 개선됨을 알 수 있었다. 이는 제안된 알고리즘을 통하여 전방향과 관련한 입체대역의 지각 단서를 부각시켜 전체적인 음색을 전방향의 음색 및 느낌과 유사하도록 조절하기 때문이다. 그 결과로 청자는 전방향 입체음향을 명확히 지각할 수 있다. 그러나 기존 알고리즘의 결과를 살펴보면 음상의 방향이 정면에서 우측으로 이동할수록 좋지 않은 성능을 보여준다. Tan & Gan의 알고리즘^[5]은 고정된 주파수 대역을 증폭 및 감쇄를 하게 되는데 만약 증폭되는 대역에 음원 신호의 주파수에너지가 클 경우 생성되는 입체음향 역시 매우 큰 값을 갖는다. 따라서 우측(방위각 0~180°) 면에 음상을 정위시키는 경우 우측 채널의 신호가 매우 커지기 때문에 인지된 음상의 방향이 90° 근방으로 몰리게 된다. 이런 현상은 음상을 방위각 45~135°사이에 정위시킬 때, 가장 영향을 크게 미쳐 성능을 저하시키는 원인이 된다.

후방향과 관련한 결과를 보면 후방 측면각도에서 음원 신호에 따라 성능 개선의 결과가 약간 상이하게 나타났다. 이는 각 음원의 다양한 스펙트럼 특성이 인간의 청각에 다른 작용을 함으로서, 해당 방향의 인지에 영향을 주기 때문이라 생각된다. 특히 방위각 135~150°의 방향에 대해서 원 머리전달함수를 사용한 방법이 제안한 방법보다 다소 용이한 성능을 나타내었다. 그러나 전체적인 결과로 비추어 볼 때 제안된 알고리즘의 성능이 기존의 알고리즘에 비하여 개선되었음을 확인할 수 있다.

3. 음질 테스트

두 번째 테스트로는 제안된 알고리즘이 음질에 미치는 영향을 확인하기 위하여 음질 테스트를 실시하였다. 평가 방법은 원 머리전달함수를 적용한 소리를 기준으로 삼아서 개선된 소리의 음질이 기준에 얼마나 충실한지를 비교하는 Degradation Category Rating Method를 사용하였다. 이 방법은 5가지 Degradation Category Scale를 적용하며 그 내용은 다음과 같다^[20].

표 3. 음질 테스트 점수
Table. 3. Degradation category scale

5	Degradation is inaudible.
4	Degradation is audible but not annoying
3	Degradation is slightly annoying
2	Degradation is annoying
1	Degradation is very annoying

그림 15는 각 방향에 대한 음질 테스트 결과를 보여주고 있다. 음질테스트 결과는 각 방향에 대한 평균값을 제시하였다. 제안된 알고리즘은 원 머리전달함수를 적용한 입체음향에 비해 음질의 열화가 매우 적다. 이는 인간이 지각할 수 있는 임계대역에서만 주파수 에너지를 조절하여 음질의 변화를 가능한 적게 유지할 수 있으며 가변적인 가중치를 이용하여 IID를 보존할 수 있기 때문이다. 그러나 Tan과 Gan의 방법^[5]은 다양한 사운드 특성을 고려하지 않고 단순히 5개의 고정된 대역에서 증폭 및 감쇄를 하기 때문에 정확하게 단서를 부각시킬 수 없다. 특히 Tan과 Gan의 방법^[5]을 사용하여 음상을 정위시킬 경우 특정 주파수 대역의 에너지가 매우 강해지기 때문에 음색의 변화뿐만 아니라 음질이 손상되는 부작용이 발생하게 된다. 그 결과로 음상이 원하는 방향에 정위되더라도 인지되는 거리가 달라지거나 전혀 다른 느낌의 입체음향이 생성된다. 따라서 제안된 알

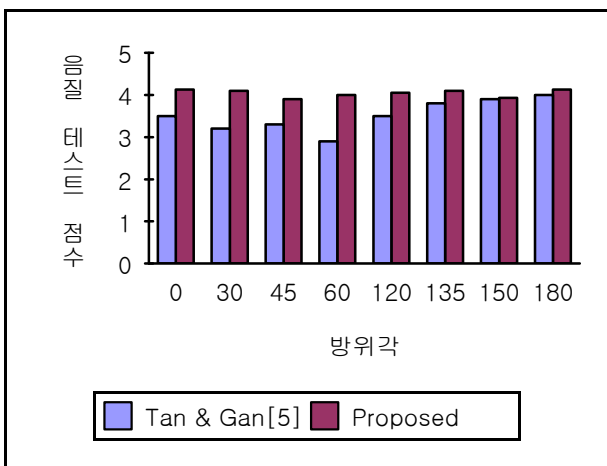


그림 15. 음질 테스트 결과
Fig. 15. MOS test results

고리즘을 사용하게 되면 개선된 입체음향을 생성할 수 있을 것이다.

VI. 결론

본 논문에서는 헤드폰 기반의 바이노럴 환경에서 표준화된 머리전달함수를 이용하여 3차원 공간상에 음상을 정위시킬 때, 청자가 음상의 위치를 혼돈하는 문제를 개선하기 위한 방법을 제안한다. 먼저 머리전달함수가 인간의 청각에 미치는 에너지를 계산하여 1차적으로 방향감을 강조한다. 이어 생성된 입체음향의 전역 마스킹 임계치 및 라우드니스를 이용하여 음상정위의 단서가 달라지는 것을 방지하고 해당 대역의 가중치를 조절하여 2차적으로 음상정위감을 개선한다. 그 결과로 제안된 방식을 사용해 생성된 입체음향은 청자는 기존 알고리즘에 비해 음상이 정위되는 방향의 느낌을 강하게 지각할 수 있게 된다. 이에 기존 알고리즘과 비교하여 입체음향의 전체적인 인지도는 50% 이상 증가하였다. 그러나 현재 시뮬레이션에서는 주로 음성 신호를 사용하였으므로 앞으로는 심리음향적인 특성을 이용하여 다양한 종류의 사운드 신호를 효율적으로 정위시키는 알고리즘에 대한 연구가 계속 진행될 예정이다.

참고 문헌

- [1] 김기홍, 김용완, 명현, 정혁, 김기호, “3D Sound 기술”, 정보과학회지, 19권, 5호, 30-37쪽, 2001년 5월
- [2] 구교식, 차형태, “머리전달함수의 심리음향적 특성을 이용한 전/후 음상 정위 특성 개선,” 방송공학회논문지, 11권, 4호, 448-457쪽, 2006년 12월
- [3] 강성훈, 강경욱, 입체음향, 기전연구사
- [4] 구교식, 차형태, “개선된 머리전달함수를 이용한 3차원 입체음향 성능 개선 연구,” 한국음향학회지, 28권, 6호, 557-565쪽, 2009년 8월
- [5] Chong-Jin Tan, Woon-Seng Gan, “User-defined spectral manipulation of HRTF for improved localisation in 3D sound systems, Electronics letters, Vol.34 No.25, pp. 2387-2389, Dec. 1998
- [6] 김경훈, 김시호, 배건성, 최송인, 박만호, “헤드폰 기반의 입체음향 생성에서 앞뒤 음상정위 특성 개선,” 한국통신학회논문지, 29권, 8c호, 1142-1148쪽, 2004년 8월
- [7] 이채봉, “고역 주파수 영역에서 HRTF의 간략화에 관한 연구,” 한국신호처리시스템학회논문지, 12권 1호, 1-6쪽, 2011년 1월
- [8] 김용국, 전찬준, 김홍국, 이용주, 장대영, 강경욱, “헤드폰 청취환경에서의

- 실감 오디오 재현을 위한 음상 외재화 기법,” 대한전자공학회논문지, 47권, 5호, 1-8쪽, 2010년 9월
- [9] 황성목, 박영진, 박윤식, “보편적인 기저함수를 이용한 중앙면상의 머리 전달함수 모델링,” 한국소음진동공학회논문집, 18권, 4호, 448-457쪽, 2008년 4월
- [10] 이영한, 김홍국, “가상현실 환경에서의 3차원 사운드 생성을 위한 거리 변화에 따른 구조적 머리전달함수 모델,” 말소리, 59권, 89-99쪽, 2006년 9월
- [11] 안태수, 이두호, “머리전달함수 측정법의 실험적 비교,” 한국소음진동공학회논문집, 19권, 12호, 1260-1268쪽, 2009년 12월
- [12] 이윤재, 박영진, 박윤식, “머리전달함수 측정시스템의 개발과 분석,” 제어로봇시스템학회논문지, 16권, 2호, 202-205쪽, 2010년 2월
- [13] ISO/IEC, JTC1/SC29/WG11 MPEG, Information technology-coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbits/s-Part 3: Audio, IS11172-3 (MPEG-1), 1992
- [14] E. Zwicker and H. Fastil,, Psychoacoustics, Facts and Models, Springer 2nd Edition, 1999
- [15] C. J. Moore, Hearing, Academic Press, 1995
- [16] C. I. Cheng and G. H. Wakefield, “Introduction to Head-Related Transfer Functions(HRTFs),” J. Audio Eng. Soc. pp. 231-249, April. 2001.
- [17] ISO 532: acoustics method for calculating loudness level. Geneva: International Organization for Standardization, 1975
- [18] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," Journal of the Acoustical Society of America, Vol. 65, No. 4, pp.943-950, Apr. 1979
- [19] <http://interface.cipic.ucdavis.edu/sound/hrtf.html>
- [20] ITU-T P.800, Methods for subjective determination of transmission quality

저 자 소 개



구 교 식

- 2005년 2월 : 송실대학교 정보통신공학부 졸업
- 2007년 2월 : 송실대학교 전자공학과 석사 졸업
- 2007년 3월 ~ 현재 : 송실대학교 전자공학과 박사과정
- 주관심분야 : 오디오 신호처리, 심리음향, 디지털 신호처리



차 형 태

- 1993년 : The University of Pittsburgh (공학박사)
- 1993년 ~1996년 : 삼성전자 신호처리 연구소 선임연구원
- 1996년 ~ 현재 : 송실대학교 정보통신공학부 교수
- 2001년 ~ 현재 : 신호 및 시스템학회 이사
- 2001년 ~ 현재 : 한국 방송공학회지 편집위원장, 편집위원
- 주관심분야 : Multimedia Systems and Applications Audio and Video Signal Processing, Communication System, ASIC Implementation of Digital System