한국음향학회지 제30권 제5호 pp. 255~264 (2011)

인간의 청각 특성을 이용한 입체음향의 방향감 개선

Improvement of 3D Sound Using Psychoacoustic Characteristics

구 교 식, 차 형 태
(Kyo-Sik Koo, Hyung-Tai Cha)

숭실대학교 전자공학과
(접수일자: 2011년 4월 13일; 수정일자: 2011년 5월 19일; 채택일자: 2011년 5월 25일)

3차원 공간에서 음원으로부터 사람의 귀로 음향적인 전달 과정을 표현하는 머리전달함수는 사람이 음원의 위치를 판단할 수 있는 중요한 정보를 포함하고 있으며, 이를 이용하여 실질적으로는 존재하지 않는 음원을 근사적으로 생성할 수 있다. 그러나 각 청자에 특화되지 않는 머리전달함수를 사용함으로 인해 전후 방향이나 상하 방향의 혼동이 발생하게 되어 음상정위 성능이 저하된다. 이 논문에서는 머리전달함수를 통해 생성된 입체음향을 인간의 청각 특성을 사용하여 개선하는 알고리즘을 제안한다. 머리전달함수가 적용된 사운드 신호의 전역 마스킹 값을 계산한 후, 청자에게 크게 영향을 미치는 주파수 대역만을 추출하여 이를 강조함으로서 각 방향에 해당하는 인지 특성을 부각시키는 방법을 제안하였으며, 제안된 방법의 성능을 청감테스트를 통해 증명하였다.
핵심용어: 음상정위, 심리음향, 입체음향, 머리전달함수, 가상현실
투고분야: 음향 신호처리 분야 (1,1)

The Head Related Transfer Function (HRTF) means a process related to acoustic transmission from 3d space to the listener's ear. In other words, it contains the information that human can perceive locations of sound sources. So, we make virtual 3d sound using HRTF, despite it doesn't actually exist. But, it can deteriorate some three-dimensional effect by the confusion between front and back directions due to the non-individual HRTF depending on each listener. In this paper, we proposed the new algorithm to reduce the confusion of sound image localization using human's acoustic characteristics. The frequency spectrum and global masking threshold of 3d sounds using HRTF are used to calculate the psychoacoustical differences among each directions. And perceptible cues in each critical band are boosted to create effective 3d sound. As a result, we can make the improved 3d sound, and the performances are much better than conventional methods.
Keywords: Sound localization, Psychoacoustics, 3d Sound, Head related transfer function, Virtual reality
ASK subject classification: Acoustic Signal Processing (1.1)

I. 서 론

지난 수십 년 동안 멀티미디어 기술이 급격하게 발달함에 따라 카세트, PMP, MP3 player와 같은 개인용 멀티미디어 장치를 이용한 오디오, 비디오 등의 멀티미디어 데이터 이용이 포괄적으로 증가하고 있다. 또한 인터넷 스트리밍 기술을 사용하여 멀티미디어 데이터 전체를 다운로드 받지 않고 실시간으로 제공하는 서비스가 상용화되고 있다. 이러한 기술들에 힘입어 컴퓨터와 인터넷 접속이 가능한 곳에서는 언제, 어디서나 자신이 선택한 오디오, 비디오 데이터를 즐길 수 있게 되었다 [1]. 최근에는 스마트폰을 이용하여 보다 고음질, 고화질 서비스를 제공하기 위한 연구가 활발히 이루어지고 있으며 그 중 오디오 분야에서는 보다 생생한 공간감을 구현하기 위해 입체음향 기술이 적용되고 있다. 2채널 재생시스템 환경에서 입체음향을 생성하기 위해서는 3차원 공간상에 방사된 음원이 인간의 귀에 도달하기까지의 정보를 담고 있는 머리전달함수 (HRTF; Head Related Transfer Function)를 사용하는 방법이 대표적이다. 그러나 머리전달함수의 비개인화된 문제점으로 인하여 원하는 방향에 음상이 제대로 정위되지 않음으로서 사용자가 제대로 된 방향감을 느끼게 하기에는 아직도 해결해야 될 문제점이 남아 있다. 보다 넓은 대역폭, 빠른 처리 속도를 가지는 차세대

책임저자: 차 형 태 (hcha@ssu.ac.kr)
156-743 서울시 동작구 상도1동 숭실대학교 전자공학과
(전화: 02-820-0711; 팩스: 02-821-7653)

멀티미디어 환경에서도 음상정위의 부정확성 문제는 여전히 해결해야 할 중요한 문제로 남아있을 것이다 [2].

이에 본 논문에서는 머리전달함수를 이용하여 입체음향을 생성하는 과정에서 발생하는 음상 정위의 혼돈 문제를 개선하기 위한 알고리즘을 제안한다. 머리전달함수가 적용된 사운드 신호에서 순음 성분 (Tonal component)과 비순음 성분 (Non-tonal component)을 구한 후 그에 따른 마스킹 특성을 이용하여 각 방향에 따른 주파수 성분 중 청자가 인지 가능한 대역의 에너지를 강조하여 방향감을 개선하고자 하였다.

본 논문의 2절에서는 지각 특성을 이용한 사운드 신호 해석과 관련한 이론에 관해 기술하고 제 3절에서는 제안된 알고리즘에 대해 설명한다. 이어 제 4절에서는 테스트를 통한 실험 결과를 제시하며 마지막으로 5절에서 결론을 맺는다.

Ⅱ. 지각 특성을 이용한 신호 해석

일반적인 음성 및 오디오 신호는 사람의 귀를 통하여 전달되어지고 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 가청한계 (Absolute hearing threshold), 임계대역 주파수 분석 (Critical band frequency analysis), 마스킹 현상 (Masking phenomena)등을 이용하여 분석하게 된다 [3-5].

2.1. 임계 대역 분석

사람의 귀는 외이, 중이, 내이로 크게 세 부분으로 구분할 수 있다. 외이는 외부 소리 에너지를 수집하고, 소리의 정위 (sound location) 정보 해석에 유용하도록 미세한 지연을 발생시키며 고막에서 음압의 형태로 전달되어 지는 소리 에너지를 진동을 통해 역학적인 에너지로 변환하는 역할을 한다. 중이는 이소골 (Ossicles)로 구성되어져 있으며, 소리의 증폭과 과도한 소리 에너지로부터 귀를 보호하기 위한 자동 조절기의 역할을 수행한다. 내이의 경우 세반고리관 (Semicircular canals)은 몸이 평형감각을 제어하는 부분이며 난원창 (Oval window)에서는 이와 같이 이소골을 통해 전달되는 역학적인 에너지를 파동 에너지로 변환시켜주게 된다 [4]. 기저막 (basilar membrane)에서는 이렇게 전달되어진 음압 파동을 적절한 해석을 통해 전기적인 자극으로 변환하여 청신경 계에 전달하는 역할을 수행한다. 또한 기저막에서는 일반적으로 "place theory"라고 하는 메커니즘을 통해 신호에 대한 주파수 분석을 하게 되는데 이것은 기저막이 신호 각각의 주파수 성분에 따라 각각 다른 위치에서 자극에 대한 최대 응답을 발생시키는 것을 나타낸다 [5].

일반적으로 내이의 기저막에서의 임의의 신호에 대한 주파수 분석은 기저막의 역학적인 특성에 영향을 받게 되는데 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. 이러한 임계대역의 대역폭은 중심 주파수를 중심으로 저주파 영역에서는 대략 0.1 kHz의 대역폭을, 고주파 영역에서는 대략 4 kHz의 대역폭을 갖게 된다. 이러한 임계대역은 바크 율 (Bark ratio)로 표현하며, 심리 음향의 모든 연산은 이러한 바크 스케일 (Bark scale)을 기준으로 이루어지게 된다. 이러한 바크 율과 임계대역의 대역폭은 임의의 주파수 f에 대해서 다음 식 (1)과 (2)로 표현되어 진다 [4-5].

※ Critical band ratio (Bark ratio)

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

※ Critical band ratio

$$\triangle_{\omega_G}/Hz = 25 + 75[1 + 1.4(f/kHz)^2]^{0.69} \quad (2)$$

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 저항치이다. 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 [4-5].

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

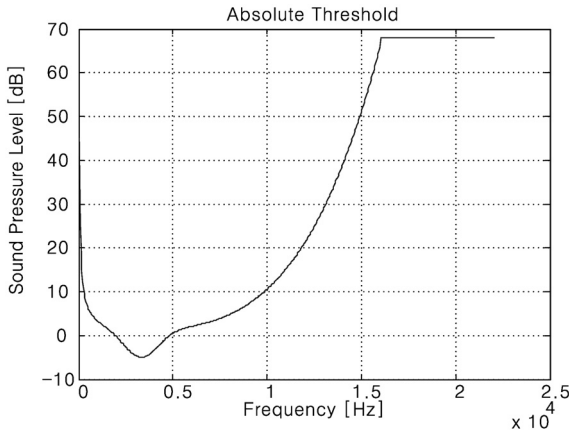

그림 1. 절대 가청 한계
Fig. 1. Absolute hearing threshold.

2.3. 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 마스키 (maskee)라 하고, 이러한 음의 인식에 영향을 미친 신호를 마스커 (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 [3,6,7].

2.3.1. 순음 성분 및 비 순음 성분의 검출

먼저 오디오 데이터 스트림에서 frame 단위로 신호를 추출한 후, 푸리에 변환을 통해 전력 스펙트럼 밀도 (PSD; Power Spectral Density)를 계산한다.

$$\begin{aligned} &X_a(n) = \frac{1}{N}\sum_{n=0}^{N-1} w(n)x(n)e^{-j\frac{2\pi\omega n}{N}} \\ &P(k) = PN + 10\log_{10}|X_a(n)|^2 \\ &\quad where, 0 \le n \le N-1, 0 \le k \le \frac{N}{2} \end{aligned} \tag{4}$$

이 때 N은 1프레임의 샘플 수를 의미하며 PN은 90.302 dB로 고정되어 있다. w(n)은 해닝 윈도우 (Hann Window) 함수를 사용한다.

다음으로 입체음향 신호의 마스킹 특성을 고려하기 위하여 각 프레임의 순음 성분 및 비 순음 성분을 검출한다. 우선 임의 주파수 인덱스 k에 대해 다음 식 (5)의 조건을 만족하는 성분을 순음 성분으로 결정한다. 즉 P(k) 〉 P(k ± 1)인 조건을 만족하는 P(k)를 local maximum 값으로 결정하고, 이렇게 결정된 local maximum 값들 중 Δ_k로 구분되어지는 구간 중 최소 7 dB 이상 차이나는 주파수 성분들을 순음 성분으로 간주한다.

$$S_T = (P(k) | P(k) > P(k \pm 1), P(k) \ge P(k \pm \Delta_k) + 7dB) \tag{5}$$

여기서 Δ_k는 임계대역을 기반으로 한 주파수 범위로서 다음과 같다 [1,3,4].

$$\Delta_k = \begin{cases} 2 & (2 < k < 63) \\ [2,3] & (6 \le k < 127) \\ [2,6] & (127 \le k < 255) \\ [2,12] & (255 \le k < 500) \end{cases} \tag{6}$$

이렇게 검출된 순음 성분들은 다음과 같이 이웃하는 순음 성분들을 이용하여 순음 마스커 (masker)의 음압 레벨을 계산한다 [1,3,4].

$$P_{TM}(k) = 10\log_{10}\sum_{j=-1}^{1} 10^{\frac{P(k+j)}{10}}, \quad (dB \;\; SPL) \tag{7}$$

다음으로 비순음 성분은 각각의 임계대역에 대해서 순음 성분이 검출된 주파수 인덱스 구간 $\pm\Delta_k$를 제외한 주파수 성분들을 합산하여 비순음 마스커 (masker)의 음압 레벨로 계산한다 [1.3.4].

$$\begin{aligned} &P_{NM}(\bar{k}) = 10\log_{10}\sum_{j} 10^{\frac{P(j)}{10}}, \quad (dB \;\; SPL) \\ &where, \forall P(j) \notin P_{TM}(k, k \pm 1, k \pm \Delta_k) \end{aligned} \tag{8}$$

이때 비순음 마스커의 인덱스 $\bar{k}$는 식 (9)와 같이 각각의 임계대역에 기하평균 (geometric mean)에 근접한 주파수 인덱스를 사용한다.

$$\bar{k} = \left(\prod_{j=i_{lz}}^{i_{hz}} j\right)^{\frac{1}{(i_{lz} - i_{hz} + 1)}} \tag{9}$$

여기서 i_{lz}, i_{hz}는 각각 임계대역의 저주파와 고주파 경계의 주파수 인덱스이다. 그림 2는 검출된 순음 (o 표시) 및 비순음 성분 (x 표시)을 나타내고 있다.

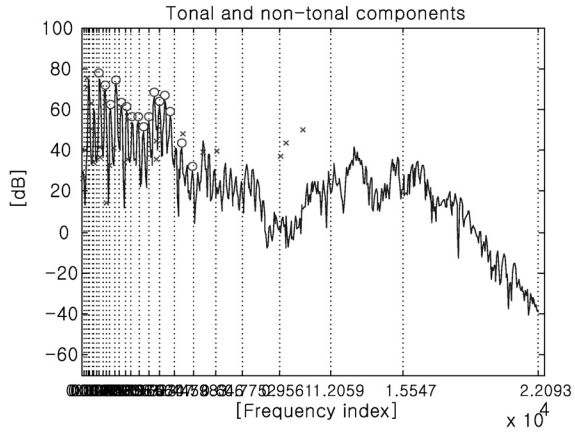

그림 2. 순음 및 비순음 성분의 검출
Fig. 2. Detection of tonal & non-tonal components.

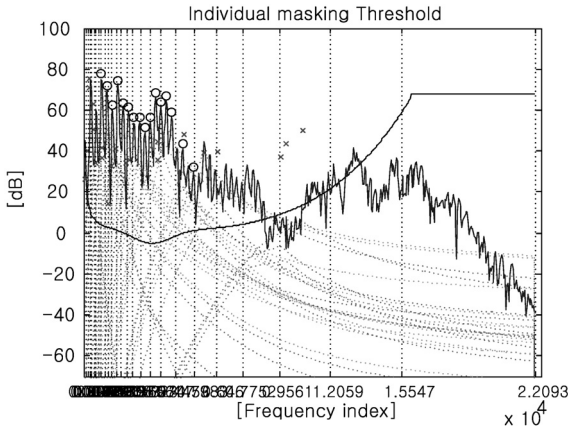

그림 3. 개별 마스킹 임계치
Fig. 3. Individual masking threshold.

2.3.2. 개별 마스킹 임계치 계산

다음으로 전역 마스킹 임계치 계산 시 마스커의 수를 줄이기 위해서 decimation을 수행한다. decimation 시 $P_{TM}(k) \geq T_q(k)$, 절대 가청 한계 이상의 순음 마스커만 고려하며 각각의 critical band에 대해 0.5 bark-width sliding window를 적용하여 큰 음압 레벨을 갖는 마스커만 고려한다. decimation 된 마스커들 중 다음과 같은 sub-sampling scheme을 적용하여 마스커의 수를 감소시킨다.

$$P_{TM,NM}(i) = P_{TM,NM}(k)$$
$$P_{TM,NM}(k) = 0$$
$$where,\ i = \begin{cases} k & 1 \le k \le 48 \\ k+(k \bmod 2) & 49 \le k \le 96 \\ k+3-((k-1) \bmod 4) & 97 \le k \le 232 \end{cases} \quad (10)$$

이어 decimation 과정과 재배열 과정을 거쳐 얻어진 순음 마스커의 개별 마스킹 임계치 (individual masking threshold)를 계산한다 [3-4].

※ Tonal Masking Threshold

$$T_{TM}(i,j) = P_{TM}(j) - 0.275z(j) + SF(i,j) - 6.025, \quad (dB\ SPL) \quad (11)$$

※ Non-tonal Masking Threshold

$$T_{NM}(i,j) = P_{NM}(j) - 0.175z(j) + SF(i,j) - 2.025, \quad (dB\ SPL) \quad (12)$$

여기서 $P_{TM}(j)$과 $P_{NM}(j)$은 각각의 식 (7)과 (8)에서 결정된 마스커들에 대한 임의의 주파수 인덱스 j에서의 순음 마스커와 비순음 마스커의 음압 레벨을 나타낸다. z(j)는 주파수 인덱스 j에 대한 식 (1)에 의해 계산되어진 바크 율 (Bark ratio)이다. SF(i,j)는 주파수 인덱스 j의 마스커에 의한 주파수 인덱스 i의 성분들에 대한 지각적 에너지 확산의 영향을 나타내는 확산함수 (Spreading function)이다. 이러한 SF(i,j)는 다음과 같이 표현되어진다 [1,3,4].

$$SF(i,j) = \begin{cases} 17\Delta_z - 0.4P_{TM,NM}(j) + 11, & -3 \le \Delta_z < -1 \\ (0.4P_{TM,NM}(j) + 6)\Delta_z, & -1 \le \Delta_z < 0 \\ -17\Delta_z, & 0 \le \Delta_z < 1 \\ (0.15P_{TM,NM}(j) - 17)\Delta_z - 0.15P_{TM,NM}(j), & 1 \le \Delta_z < 8 \end{cases} \quad (13)$$

이때 Δ_z=z(i)-z(j)는 마스커의 주파수 인덱스 j와 마스키 주파수 인덱스 i에 대한 바크율로 표현된 주파수 간격이다. 그림 3은 검출된 순음과 비순음 마스커들의 개별 마스킹 임계치 (점선)을 나타낸다.

2.3.3. 전역 마스킹 임계치 계산

이와 같이 순음 성분과 비 순음 성분의 마스킹 영향을 계산한 후, 계산되어진 각각의 마스킹 임계치와 절대 가청한계를 고려하여 다음과 같이 전역 마스킹 임계치 (GMTH; Global masking threshold)를 계산하게 된다. C는 순음 성분의 개수를, M은 비 순음 성분의 개수를 나타낸다 [1,3,4].

$$T_g(i) = 10\log_{10}\left(10^{\frac{T_q(i)}{10}} + \sum_{c=1}^{C} 10^{\frac{T_{TM}(i,c)}{10}} + \sum_{m=1}^{M} 10^{\frac{T_{NM}(i,m)}{10}}\right) \quad (14)$$

이러한 각 주파수 성분들에 의한 전역 마스킹 임계치는 주파수 영역에서의 모든 순음과 비순음 성분에 의한 기저막에서의 자극에 의한 절대 가청한계의 신호 의존적인 변형이라고 할 수 있다. 그림 4는 임의의 신호에 대한 임

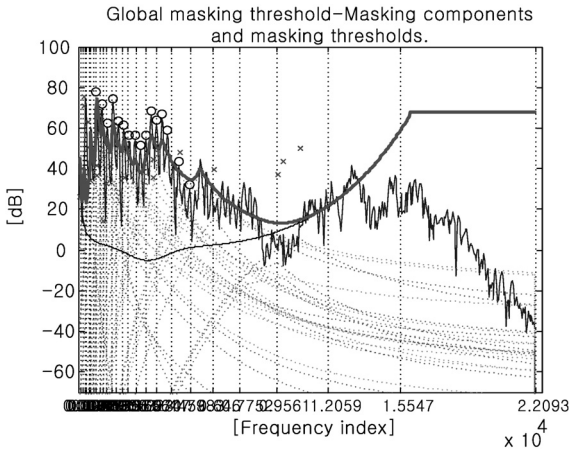

그림 4. 전역 마스킹 임계치
Fig. 4. Global masking threshold.

계대역간의 이러한 순음성분과 비순음 성분의 마스킹 영향과 절대 가청 한계치의 영향에 의한 전역 마스킹 임계치 (굵은 선)을 나타내고 있다.

III. 제안된 알고리즘

머리전달함수를 사용하여 입체음향을 생성하는 과정에서 청자에게 정확한 방향감을 느낄 수 있도록 하기 위한 연구는 이전부터 많은 관심을 받아 왔다. Tan과 Gan은 머리전달함수를 5개의 주파수 대역으로 나눈 후, 병렬로 구성된 필터 뱅크를 통과시킴으로서 방향감을 강조하고자 하였으며, Gupta는 전/후 방향 지각에 관련한 단서가 인간의 귀가 돌출된 정도와 관계가 있음을 알아냈다 [8-9]. 그리고 Cho와 Kim 등은 머리전달함수를 수정하여 정확한 방향감을 나타내고자 하였다 [10-11]. 이밖에 Zotkin과 Pec, Lee 등은 개인화 된 머리전달함수를 측정하여 혼돈 문제를 해결하고자 하였다 [12-14]. 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 마스킹과 같은 인간의 청각 특성에 따라 해당 방향과 관련한 단서가 인지되지 못할 수도 있어 이에 대한 고려가 필요하다. 따라서 음상을 정위하고자 하는 방향에 따른 다양한 스펙트럼 단서뿐만 아니라 사용될 모노 음원의 특성까지 고려해준다면 혼돈원추의 문제점을 개선할 수 있다.

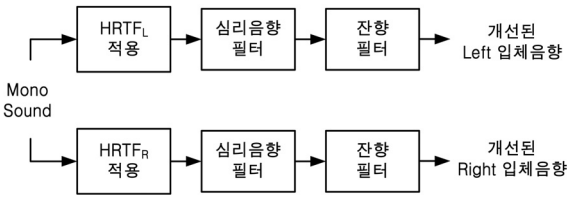

그림 5. 제안한 알고리즘의 블록다이어그램
Fig. 5. Block diagram of proposed algorithm.

이에 본 논문에서는 머리전달함수를 통해 생성된 입체음향이 인간의 청각 기관에 끼치는 영향을 고려하고 이를 바탕으로 하여 방향감을 강조하는 알고리즘을 제안하고자 한다.

제안된 알고리즘의 Block diagram은 그림 5와 같다.

심리음향을 이용하여 입체음향의 방향감을 강조하기 위한 Pre-processing으로 모노 음에 머리전달함수를 통해 원하는 방위각, a에 음상을 정위시킨다. 머리전달함수란 음원으로부터 소리가 인간의 두 귀에 도달하는 정보를 담고 있는 것으로 시간 영역에서의 임펄스 응답인 HRIR (Head Related Impulse Response)의 형태로 제공된다. 머리전달함수를 이용하여 입체음향을 구현하기 위해서는 모노 음원과 머리전달함수와의 컨벌루션 연산을 수행하게 되며 이를 통해 원하는 방향에 음상을 정위시키고 거리감까지 조절할 수 있다 [15]. 본 연구에 사용된 머리전달함수로는 CIPIC Interface Lab.에서 dummy head를 이용하여 측정한 것으로 44.1 kHz의 샘플링 율과 16 Bit/Sample의 해상도를 지니고 있다 [16]. 음상을 정위하고자 하는 방향은 정면을 방위각 0°, 오른쪽 90°, 뒷면 180°, 왼쪽 270°로 가정할 때, 고도각 0°, 방위각 0°/180°, 30°/150°, 45°/135°, 60°/120°의 4가지 경우에 대해서 시뮬레이션을 실시하였다. 각 경우에 해당하는 두 방위각은 방향을 구별하기 위한 가장 큰 단서인 음이 두 귀에 들어오는 세기차 및 시간차가 같으므로 서로 혼돈 방향에 있게 된다. 따라서 청자들은 정확한 방향을 구별할 수 없게 되므로 입체음향의 효과가 감소하는 원인이 된다. 그러나 머리나 귓바퀴와 같은 신체의 영향으로 인해 혼돈 방향에 정위된 두 입체음향의 주파수 특성이 달라지므로 인간의 청각에 미치는 영향도 달라진다. 따라서 정위된 음상이 각 방향에 따라 인간의 청각에 미치는 영향을 부각시킨다면 그에 따른 방향감도 개선되게 될 것이다.

먼저 생성된 입체음향의 각 채널에서 frame 단위로 신

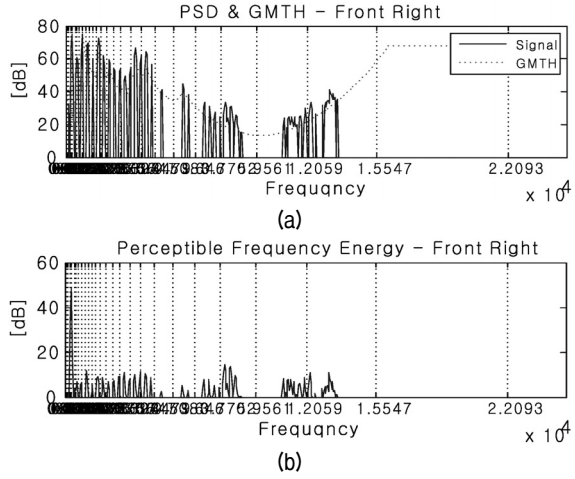

그림 6. 계산된 에너지 (고도각 0°, 방위각 30°)
(a) 전력스펙트럼밀도 & 전역마스킹임계치, (b) 각 임계대역별 인지 가능한 주파수 에너지

Fig. 6. Calculated energy (Ele. 0°, Azi. 30°).
(a) PSD & GMTH, (b) Perceptible frequency energy in each critical band.

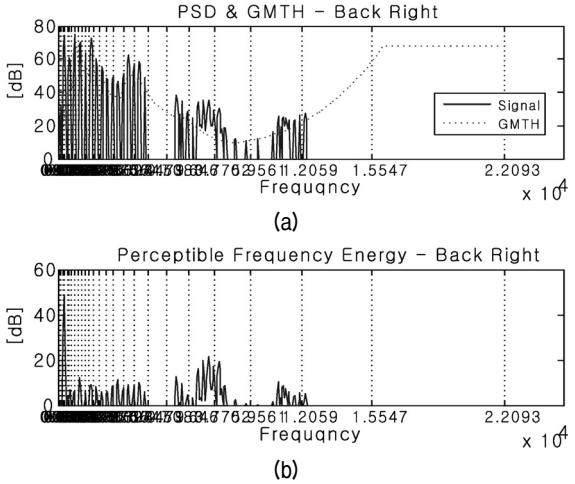

그림 7. 계산된 에너지 (고도각 0°, 방위각 150°)
(a) 전력스펙트럼밀도 & 전역마스킹임계치, (b) 각 임계대역별 인지 가능한 주파수 에너지

Fig. 7. Calculated energy (Ele. 0°, Azi. 150°).
(a) PSD & GMTH, (b) Perceptible frequency energy in each critical band.

호를 추출한 후, 2장에서 설명한 Psychoacoustic Model-1을 이용하여 실제로 원음을 들으면서 감지할 수 있는 한계인 전역 마스킹 임계치를 계산한다. 입체음향 신호는 시간에 따라 변화하며 정위된 방향에 따라 신체의 영향을 다르게 받게 된다. 그 결과로 전역 마스킹 임계치는 각 방향에 따라서 고유한 값을 가지게 된다.

그림 6과 7은 방위각 30°/150°에 정위된 입체음향의 우측 채널에서 추출된 한 프레임의 전역 마스킹 임계치를 보여주고 있다. 방위각 30°와 150°는 혼돈 관계에 있으므로 세기와 시간차는 같지만 지각적으로 신호를 해석한 경우에는 두 방향의 인지 특성이 다름을 확인할 수 있다. 인간의 청각 특성에서 청자는 각 주파수 대역에서 전역 마스킹 임계치보다 큰 신호만을 인지할 수 있다. 이를 근거로 살펴보면 현재 프레임에서는 6 kHz 근방의 주파수 대역에서 전방향에 비해 후방향에서 인지 가능한 에너지가 많이 분포하고 있다. 그러나 9 kHz 이상의 대역에서는 전방향이 우세하다. 따라서 각 임계대역에서 전역 마스킹 임계치보다 큰 주파수 에너지가 많은 대역이 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.

$$Y_a(z) = \sum_{w=i_{lz}}^{i_{hz}} d(w) \quad where,\ 0 \le z \le Z-1 \tag{15}$$

$$d(i) = \begin{cases} P(i) - T_g(i), & (P(i) > T_g(i)) \\ 0, & (P(i) \le T_g(i)) \end{cases} \tag{16}$$

여기서 i_{hz}와 i_{lz}은 전체 임계대역 Z에서의 임의의 임계대역 z에 대한 저주파 경계와 고주파 경계의 인덱스를 나타낸다.

최종적으로 계산된 a 방향의 채널당 임계대역별 합산 에너지는 $Y_a(z, q)$로 정의되며 q는 좌(l)/우(r) 각 채널을 의미한다. 이어서 동일한 방법으로 혼돈 방향인 b 방향의 채널당 임계대역별 합산 에너지, $Y_b(z, q)$를 계산한다.

그림 8은 고도각 0°에서 방위각 30°/150° 방향의 인지 가능한 주파수 에너지들의 합을 나타낸다. Duplex theory에 따르면 주파수 영역에서 1.5 kHz 이하의 대역에서는 IID보다 ITD가 주 단서가 되기 때문에 혼돈방향에서 머리전달함수의 에너지 차이가 크지 않다. 이에 따라 1.5 kHz 이하의 대역에 해당하는 11 bark 이하의 대역에서는 인지 가능한 에너지의 차이도 거의 없음을 볼 수 있다. 그러나 1.5 kHz 이상의 대역에서는 IID가 방향 인지의 단서가 되므로 각 방향에 따라서 인지할 수 있는 에너지의 크기가 달라진다. 그림 8의 (a)에서 20~22 bark는 후방향이 우세한 대역이며 (b)에서는 19~24 bark 대역에서 각 방향에 따라 우세한 대역이 혼재해 있음을 볼 수 있다. 이는 생성된 입체음향이 마스킹 등의 인간의 청각 특성으로 인하여 이웃 임계대역에 영향을 받기 때문이다. 더불어 25 bark 이상의 대역은 인간의 절대 가청 한

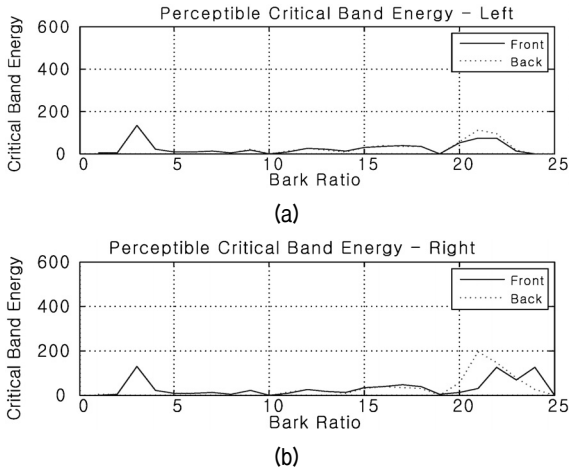

그림 8. 임계대역별 합산 에너지 (고도각 0°, 방위각 30° & 150°)
(a) 좌측 채널, (b) 우측 채널
Fig. 8. Summation of energy in each critical band (Ele. 0°, Azi. 30° & 150°).
(a) Left channel, (b) Right channel.

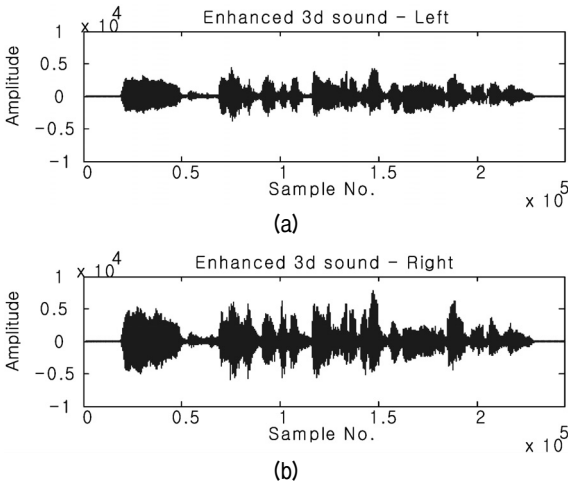

그림 9. 최종 개선된 입체음향 (고도각 0°, 방위각 30°)
(a) 좌측 채널, (b) 우측 채널역별 인지 가능한 주파수 에너지
Fig. 9. Finally enhanced 3D sound (Ele. 0°, Azi. 30°).
(a) Left channel, (b) Right channel.

계 특성상 인지를 할 수 없는 대역이기 때문에 총합은 0이 된다.

이어 인지 결과에 따라 a방향이 우세하다고 판단되는 각 임계대역에 가중치, weight를 적용함으로서 인간의 청각에 미치는 에너지를 강조한다. 가중치는 음이 손상되지 않는 범위 내에서 추출된 1.1에서 2사이의 실수값이며 본 연구에서는 Tan과 Gan이 제안한 방법 [8]과 비슷한 크기로 증폭 및 감쇄를 하기 위하여 1.5를 사용하였다.

$$en\,sound_a(f,q) = \begin{cases} X_a(f_z,q)\times weight, & (\,Y_a(z,q) > Y_b(z,q)) \\ X_a(f_z,q)/weight, & (\,Y_a(z,q) < Y_b(z,q)) \end{cases} \quad \text{for}, 0 \le z \le Z-1 \tag{17}$$

여기서 f_z는 각 임계대역에 해당하는 주파수 범위를 나타낸다.

이어 개선된 입체음향에 거리감을 부여하여 음상을 외부에 맺히도록 하기 위해 image source method를 이용하여 인공 잔향효과를 생성한다 [17]. 즉 머리전달함수에 의해 방향 정보가 부가된 음원에 image source method를 이용하여 생성한 룸 임펄스 응답을 처리함으로서 직접음과 함께 인공적으로 생성한 잔향이 더해지게 된다. 이를 통해 생성된 결과 신호로부터 청자는 정확한 음상 정위와 함께 거리감 및 공간감을 느낄 수 있게 된다 [18]. 본 논문에 사용된 잔향은 20 m × 20 m × 3 m 크기의 공간에서 음원이 10 m, 18 m, 1.5 m, 왼쪽 마이크 및 오른쪽 마이크가 9.8 m, 10 m, 1.5 m 지점 및 10.2 m, 10 m, 1.5 m에 위치하며, 반사음 차수와 반사 계수는 각각 24와 0.3인 환경을 가정하여 생성하였다.

그림 9는 최종적으로 생성된 입체음향을 나타낸다.

제안된 방법은 머리전달함수를 통해 정위된 음상의 주파수 대역에서 인간이 인지할 수 있는 신호를 추출하고 가중치를 통해 각 대역의 에너지를 조절함으로서 방향감을 강조시킨다. 그 결과로 청자는 소리의 위치를 보다 정확하게 지각할 수 있으므로 게임이나 시뮬레이터와 같은 가상현실 공간에서 보다 더 현실감을 느낄 수 있을 것이다.

IV. 실험 및 결과고찰

4.1. 실험 환경 및 방법

본 연구에서 테스트가 수행된 장소는 일반 생활 잡음이 존재하는 연구실 공간으로 피실험자들은 중앙에 설치된 의자에 편하게 앉도록 하였다. 실험에 사용된 사운드 신호들은 일반 오디오 CD에서 추출된 44.1 kHz, 16 bits/sample 음성 및 사운드 신호를 이용하였다. 사용된 머리전달함수로는 CIPIC Interface Lab.에서 측정한 것을 사용하였으며 해닝 윈도우 (Hann Window) 함수와 오버랩애드 (Overlap Add) 방식을 주파수축 변환에 사용하였다. 테스트 장비로는 청취자의 귀에 밀착되는 헤드폰 (beyerdynamic, DT880), Desktop computer (core2duo E8500), 음향 재생 프로그램으로는 Cool edit pro를 사용하였다. 그리고 비교 대상은 원 머리전달함수 및 기존 방법으로 Tan과 Gan이 제안한 머리전달함수의 필터뱅크

를 이용한 방식 [8]이며 입력 사운드 신호는 표 1과 같이 남자 및 여자 목소리 및 효과음의 3가지를 사용하였다.

그림 10은 시뮬레이션에 사용된 신호인 두 음성 신호와 비행기 소리 신호를 보여주고 있다. 두 음성 신호는 인간의 음성 신호를 3차원 공간상에 정위시키는 상황에서의 성능을 확인하기 위해 사용하였다. 가청 주파수 대역 전체에 걸쳐서 에너지가 분포하고 있으며 시간에 따라 신호가 변화하기 때문에 다양한 경우에 대하여 테스트를 사용하기에 알맞다. 또한 비행기소리 신호는 시간에 따른 변화도가 적은 경우에서의 성능 테스트를 위해 사용하였다.

테스트에 참가한 피실험자는 입체음향에 사전 지식이 없는 2~30대의 일반 남녀 10명이며 실험 전에 헤드폰과 스피커를 이용하여 간단한 샘플을 들려주고 이에 대한 간단한 설명을 진행하였다. 이어 헤드폰 환경에서 각 방법을 통해 생성된 입체음향을 재생하고 인지한 방향에 대해서 구두로서 답하도록 하였다.

표 1. 사용된 음원
Table 1. Sound sources for simulation.

	Source 1	Source 2	Source 3
이 름	남자 음성	여자 음성	비행기
샘플링 주파수	44.1 kHz		
샘플당 비트수	16 bits/sample		
샘플 수	244446		

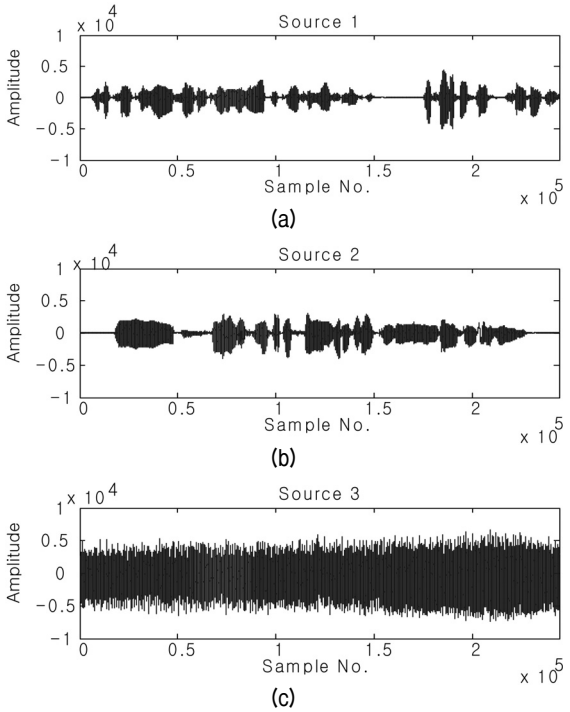

그림 10. 시뮬레이션에 사용된 음원
(a) 남자 음성, (b) 여자 음성, (c) 비행기
Fig. 10. Sound sources for simulation.
(a) Man's voice, (b) Woman's voice, (c) Airplane.

4.2. 음상정위 테스트

제안된 알고리즘의 성능을 확인하기 위하여 헤드폰 환경에서 음상 정위 테스트를 수행하였다. 앞에서 말한 3가지 방법에 따라 생성된 각 방향의 입체음향을 들려준 후, 각 방향에 대해 인지하는 정확도를 측정하였으며 ± 10° 내의 인지 오차의 경우는 정답으로 인정하였다.

그림 11과 12는 세 방법을 이용하여 생성된 입체음향에 대한 음상 정위 테스트 결과를 나타내고 있다. 원 머리전달함수와 기존의 방법에도 공간감을 주기 위해 제안한 방법과 동일한 환경으로 image source method를 사용하여 인공 잔향을 생성하였다. 이어 표 1에서 제시한 음원에 대한 테스트를 수행한 후, 각 방향에 따른 인지 결과를 제시하였다. 먼저 전방향 음상정위 결과를 보면 0° 방향에서 측정된 머리전달함수는 청자의 얼굴 등의 신체로

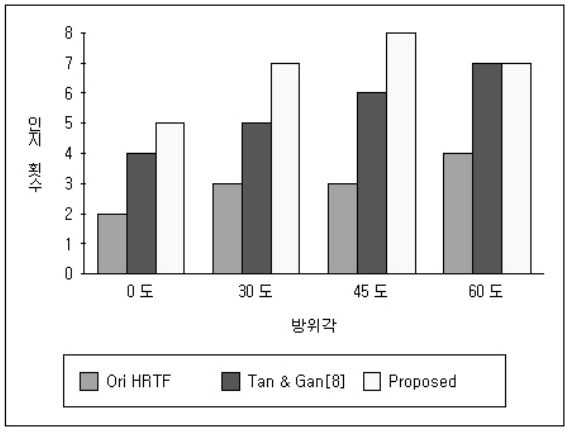

그림 11. 전방향 음상정위 테스트 결과
Fig. 11. Sound localization test result at front azimuth.

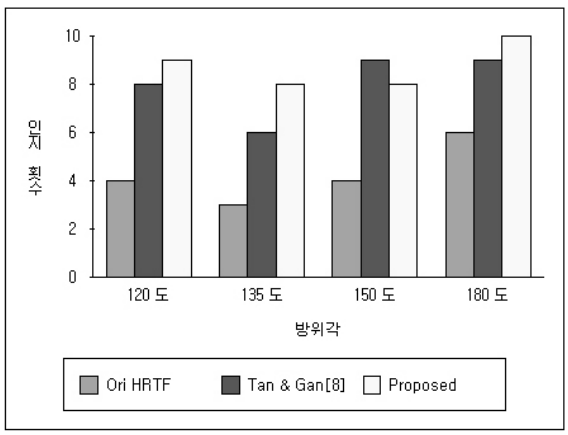

그림 12. 후방향 음상정위 테스트 결과
Fig. 12. Sound localization test result at back azimuth.

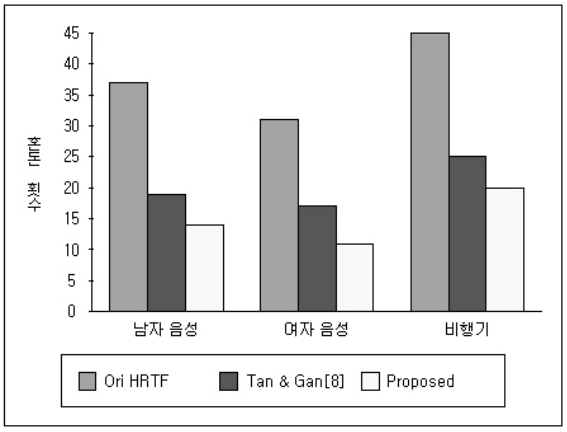

그림 13. 생성된 입체음향의 전/후 혼돈 횟수
Fig. 13. Front/Back confusion result of each method.

인한 영향이 크기 때문에 0° 방향에 정위된 음상은 180° 방향과 쉽게 혼동되었다. 그리고 90° 방향으로 갈수록 Tan과 Gan의 방법 [8]과 제안한 방법은 원 머리전달함수를 이용한 결과보다는 나은 결과를 보이고 있다. 한편 후방향 음상정위 테스트에서는 대부분 방향을 용이하게 구분할 수 있음을 나타냈다. 그러나 음원의 다양한 스펙트럼 특성이 인간의 청각기관에 미치는 영향이 다르므로 전체적인 결과에서 확인할 수 있듯이 제안된 방법이 기존의 방법에 비해 성능이 개선되었음을 알 수 있었다. 이는 청자가 인지할 수 있는 에너지가 큰 대역을 조절하여 해당 대역의 느낌을 강조하기 때문이라 생각된다.

그림 13은 생성된 입체음향의 혼돈 정도를 나타내고 있다. 음상정위 테스트 결과와는 별도로 정위된 음상을 방위각에 관계없이 전방향이나 후방향으로 구분하여 인지되는 결과를 비교하였다. 결과를 보면 원 머리전달함수에 비해 평균 60 %의 혼돈 횟수가 감소하였다. 머리전달함수를 사용한 입체음향은 머리전달함수의 스펙트럼 특성뿐만 아니라 모노 음원의 스펙트럼 특성에도 영향을 받기 때문에 이를 고려해주는 과정이 필요하다. 이에 제안된 알고리즘에서는 마스킹 특성 등을 고려하여 전/후 방향에 따라 청자가 인지하는 부분을 조절함으로서 혼돈 횟수를 줄이고 음상 정위 성능을 향상시킬 수 있다.

4.3. 음질 테스트

두 번째 테스트로는 제안된 알고리즘이 음질에 미치는 영향을 확인하기 위하여 음질 테스트를 실시하였다. 평가방법은 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 [19].

표 2. 음질 테스트 점수
Table 2. 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

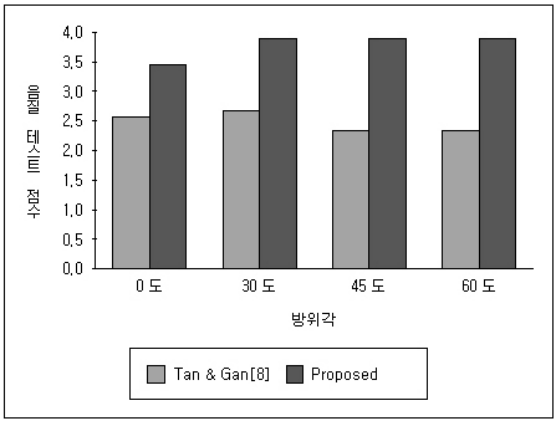

그림 14. 전방향 음질 테스트 결과
Fig. 14. MOS test result at front azimuth.

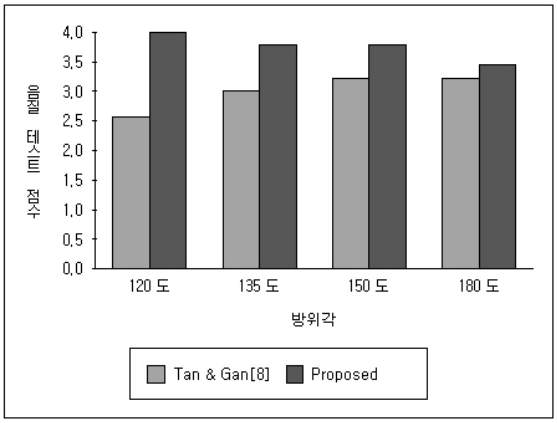

그림 15. 후방향 음질 테스트 결과
Fig. 15. MOS test result at back azimuth.

그림 14와 15는 각 방향에 대한 음질 테스트 결과를 보여주고 있다. 음상정위 테스트와 마찬가지로 각 방향에 대한 평균값을 제시하였다. Tan과 Gan의 방법 [8]과 비교해서 전체적으로 음질의 손상이 적음을 알 수 있다. Tan과 Gan의 방법 [8]은 모노 음원의 특성이나 주파수 에너지의 존재 유무에 상관없이 5개의 고정된 대역에서 증폭 및 감쇄를 하기 때문에 증폭되는 대역에서 모노음의 주파수 에너지가 큰 경우 IID가 달라지고 음질이 손상되

는 현상이 발생하게 된다. 특히 90° 주변으로 음상을 정위시킬 경우 이 현상은 더욱 심해진다. 그러나 본 알고리즘은 청자가 지각할 수 있는 부분만을 조절하기 때문에 음질의 변화를 최소화 할 수 있다. 이를 통해 개선된 입체음향의 생성이 가능해진다.

V. 결 론

본 논문에서는 헤드폰 재생 시스템에서 비 개인화 된 머리전달함수를 이용하여 입체음향을 생성할 때 발생하는 음상정위의 혼돈 문제를 개선하기 위한 알고리즘을 제안한다. 머리전달함수를 통해 생성된 입체음향이 인간의 청각에 미치는 영향을 고려하여 전역 마스킹 임계치를 계산하고 청자가 인지할 수 있는 에너지를 추출한다. 그리고 추출된 에너지를 이용하여 우세한 주파수 대역을 조절해줌으로서 해당 방향의 느낌을 강조하고자 하였다. 제안된 방식을 사용해 생성된 입체음향은 음상정위 테스트에서 성능이 향상되었으며 음상 인지의 혼동 횟수도 감소시킴을 확인할 수 있었다. 그러나 청각 모델링은 상당히 주관적이므로 모든 청취자들에게 동일한 효과를 거둘 수는 없다. 이에 향후 진행될 연구 과제는 방향 지각 단서들과 심리음향 특성과의 관계를 이용하여 보다 현실감 있는 입체음향을 구현하는 방법에 대한 연구가 진행될 예정이다.

감사의 글

이 연구는 2011년도 숭실대학교 교내연구비 지원에 의한 연구임.

참 고 문 헌

1. 차혁근, “지각적 특성을 이용한 저비트오율 압축 오디오 음질 개선,” *숭실대학교*, 2004.
2. 구교식, 차형태, “개선된 머리전달함수를 이용한 3차원 입체음향 성능 개선 연구,” *한국음향학회지*, 28권, 6호, 557-565쪽, 2009.
3. 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.
4. E. Zwicker and H. Fastil,, Psychoacoustics, *Facts and Models, Springer 2nd Edition,* 1999.
5. C. J. Moore, *Hearing,* Academic Press, 1995.
6. ISO/IEC, JTC1/SC29/WG11 MPEG, *Information Technology – Generic Coding of Moving Pictures and Associated Audio – Part 3 : Audio,* IS13818-3 (MPEG-2), 1994.
7. ISO/IEC, JTC1/SC29/WG11 MPEG, *MPEG-2 Advanced Audio Coding, AAC International Standard,* IS13818-7, 1997.
8. 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, Nov. 1998.
9. Navarun Gupta, Armando Barreto and Carlos Ordonez, “Spectral Modification of head-related transfer functions for improved virtual sound specialization,” *IEEE International Conference on Acoustics, Speech and Signal Processing,* pp. 1953-1956, May, 2002.
10. Sangjin Cho, Ovcharenko A. and Uipil Chong, “Front-Back Confusion Resolution in 3D Sound Localization with HRTF Databases,” *The 1st International Forum on Strategic Technology,* pp. 239-243, Oct. 2006.
11. 김경훈, 김시호, 배건성, 최송인, 박만호, “헤드폰 기반의 입체음향 생성에서 앞/뒤 음상정위 특성 개선,” *한국통신학회논문지,* 29권, 8c호, 1142-1148쪽, 2004.
12. Zotkin, D.Y.N., Hwang, J., Duraiswaini, R. and Davis, L.S., “HRTF personalization using anthropometric measurements,” *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics,* pp. 157-160, Oct. 2003.
13. Pec, M., Bujacz, M., Strumillo, P. and Materka, A., “Individual HRTF measurements for accurate obstacle sonification in an electronic travel aid for the blind,” *International Conference on Signals and Electronic Systems,* pp. 235-238, Sep. 2008.
14. 이윤재, 박영진, 박윤식, “머리전달함수 측정 시스템의 개발과 분석,” *제어로봇시스템학회논문지,* 16권, 2호, 202-205쪽, 2010.
15. 강성훈, 강경옥, *입체음향,* 기전연구사, 1997.
16. http://interface.cipic.ucdavis.edu/sound/hrtf.html
17. J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.*, vol. 65, no. 4, Apr. 1979.
18. 김용국, 전찬준, 김홍국, 이용주, 장대영, 강경옥, “헤드폰 청취환경에서의 실감 오디오 재현을 위한 음상 외재화 기법,” *대한전자공학회논문지*, 47권, SP편, 5호, 1-8쪽, 2010.
19. ITU-T P.800, *Methods for subjective determination of transmission quality.*

저자 약력

•구 교 식 (Kyo-Sik Koo)

2005년: 숭실대학교 정보통신전자공학부 (학사)
2007년: 숭실대학교 전자공학과 (석사)
2007년 ~ 현재: 숭실대학교 전자공학과 박사과정
※ 주관심 분야: 오디오 및 음성 신호처리, 통신 신호 처리

•차 형 태 (Hyung-Tai Cha)

1993년: The University of Pittsburgh (박사)
1993년 ~ 1996년: 삼성전자 신호처리 연구소 선임연구원
1996년 ~ 현재: 숭실대학교 정보통신전자공학부 교수
※ 주관심 분야: Multimedia Systems and Applications Audio and Video Signal Processing, Communication System, ASIC Implementation of Digital System