

청각에너지를 이용한 모노럴 시스템에서의 음상 정위 특성 연구

Research for Characteristics of Sound Localization at Monaural System Using Acoustic Energy

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

*송실대학교 전자공학과

(접수일자: 2011년 2월 18일; 수정일자: 2011년 3월 30일; 채택일자: 2011년 4월 18일)

디지털 신호처리 기술의 발달로 게임, 멀티미디어 콘텐츠, 가상현실 등에서 입체음향에 대한 관심이 증대되었다. 입체음향에 대한 많은 연구에서는 현실감을 증가시키기 위한 여러 가지 단서를 제시하고 있다. 그러나 이런 단서들은 입체음향이 인간의 두 귀가 정상인 바이노럴 환경에서 생성됨을 기본 바탕으로 삼고 있다. 따라서 이 단서들을 한쪽의 귀에만 적용하는 모노럴 환경에 적용하게 되면 그 성능이 감소된다. 이에 모노럴 환경에서도 효율적인 입체음향 구현을 위한 여러 가지 단서들이 연구되고 있는데 가장 대표적인 것이 양이 단서 (Duplex theory)이다. 양이 단서는 음이 방사되는 방향에 따라서 귓바퀴 등의 인간의 신체에 의해 영향을 받게 된다는 것으로 이 특성을 강조시키면 모노럴 환경에서도 각 방향에 따른 방향감을 개선시킬 수 있게 된다. 본 논문에서는 인간의 청각 특성을 이용하여 모노럴 환경에서 음상 정위감을 개선시키기 위한 알고리즘을 제안한다. 서로 대칭되는 각 머리전달함수의 여기에너지를 계산하고 각 bark 대역에 따른 비율을 추출한 후 이를 이용하여 각 방향에 해당하는 특성을 부각시키는 방법을 제안하였으며 청감테스트를 통하여 제안한 방식이 모노럴 환경의 입체음향 시스템에서 향상된 입체음향을 재생함을 확인할 수 있었다.

핵심용어: 음상정위, 심리음향, 입체음향, 머리전달함수, 가상현실

투고분야: 음향 신호처리 분야 (1.1)

According to developments of digital signal processing, 3D sound come into focus on multimedia systems. Many studies on 3d sound have proposed lots of clues to create realistic sounds. But these clues are only focused on binaural systems which two ears are normal. If we make the 3d sound using those clues at monaural systems, the performance goes down dramatically. In order to use the clues for monaural systems, we have studies algorithms such as duplex theory. In duplex theory, the sounds that we listen are affected by human's body, pinna and shoulder. So, we can enhance sound localization performances using its characteristics. In this paper, we propose a new method to use psychoacoustic theory that creates realistic 3D audio at monaural systems. To improve 3d sound, we calculate the excitation energy rates of each symmetric HRTF and extract the weights in each bark range. Finally, they are applied to emphasize the characteristics related to each direction. Informal listening tests show that the proposed method improves sound localization performances much better than the conventional methods.

Keywords: Sound Localization, Psychoacoustics, 3d Sound, Head Related Transfer Function, Virtual Reality

ASK subject classification: Acoustic Signal Processing (1.1)

I. 서론

최근 멀티미디어 기술의 발달, 특히 음향 기술의 급격

한 발달과 더불어 고품질 오디오에 대한 요구와 함께 현실감 있는 오디오를 재생하기 위한 실감 오디오 기술 개발이 요구되고 있다. 이러한 요구를 만족시키기 위해 사용자의 가상현실 증강 및 실감나는 오디오 효과를 제공해 줄 수 있는 입체음향에 대한 연구가 활발히 진행되고 있다 [1]. 일반적으로 입체음향이란 모노음의 음의 고저, 음

색, 음원의 방향이나 거리감을 조절하여 청자로 하여금 현장감이나 입체감을 가지게 하는 음향이다 [2]. 이것에 의하여 음원의 위치, 주위의 환경, 자신의 위치, 물체의 움직임, 음상의 형태와 표정 등을 얻을 수 있고, 각테일 파티 효과도 얻어진다. 2채널을 사용하는 바이노럴 시스템에서 입체음향을 구현하는 방법에는 여러 가지가 있으나 가장 대표적인 방법은 머리전달함수를 이용하는 방법이다. 머리전달함수는 자유 공간상에서 음원과 귀 간의 경로를 모델링하는 함수로서 음상의 공간화와 관련된 중요한 음향 파라미터이다 [3]. 이는 헤드폰과 같은 바이노럴 환경에서 여러 가지 단서들과 함께 음원의 방향성을 포함하는 공간감을 표현하는데 널리 사용되고 있다. 그러나 기존의 방법들은 바이노럴 환경에서의 구현을 목적으로 진행되었기 때문에 모노럴 환경에 그대로 적용하는 것은 무리가 따르며 구현된 입체음향의 성능도 제한적이게 된다. 여기서 모노럴 환경이란 청자가 선천적/후천적인 원인으로 인하여 한 쪽 귀의 청력을 상실하게 되어 나머지 한 귀로만 입체음향을 인지하는 경우를 말하며 보청기나 시뮬레이션 기기의 발전으로 인하여 점점 그 영역이 넓어지고 있다. 이에 수십 년 전부터 모노럴 환경에서도 음상을 정위하고자 하는 연구가 진행되어 왔다. 그 대표적인 예로 Rao는 머리전달함수를 최적화시켜 음상 정위의 성능을 향상시키고자 하였으며 [4], Wanrooij와 Yan은 소리가 인간의 신체에 영향을 받는다는 사실에 근거하여 head shadow effect 및 Pinna effect를 모노럴 음상 정위의 단서로 제시하였다 [5-6]. 그러나 모노럴 환경에서 수평면상에 음상을 정위하는 것은 한계가 있기 때문에 기존 연구들은 비교적 제약이 덜한 수직면상에서의 음상정위에 집중되어 있다. 더불어 개인용 멀티미디어 기기의 등장에 따라 블루투스 핸드프리와 같은 한 개의 채널을 이용하는 장치의 중요성이 증대되어 현재는 수직면뿐만 아니라 수평면 상에서도 정확한 음상 정위가 요구되고 있다.

이에 따라서 본 논문에서는 모노럴 시스템을 이용하여 수평면 상에서 입체음향을 구현하는 경우 성능을 향상시키기 위한 알고리즘을 제안하고자 한다. 본 논문의 2절에서는 입체음향 생성에 관련한 이론에 관해 기술하고 제 3절에서는 제안된 알고리즘에 대해 설명한다. 이어 제 4절에서는 테스트를 통한 실험 결과를 제시하며 마지막으로 5절에서 결론을 맺는다.

II. 입체음향

2.1. 입체음향 생성

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 [7].

사람이 방향 지각을 인지할 때 일반적으로 양이 단서 (binaural cue)와 편이 (여과) 단서 (Monaural (filtering) cue)를, 그리고 거리 지각을 인지할 때 음량감 (Loudness perception), 확산감 (Spectral perception) 및 잔향감 (Reverberation perception)을 이용한다 [8]. 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 (100~1.5 kHz; 100 Hz 이하는 무지향성)에, IID는 고역대 (1.5 kHz~)에 영향을 끼치게 된다 [8-9]. 즉 이를 자세히 풀어 보면 음파의 중요한 성질 중 하나는 바로 자신의 파장보다 긴 길이의 물체에서는 멈춰지고 짧은 길이의 물체에서는 비껴 나아간다는 것이다. 1500 Hz 인저리의 음파는 대략 머리의 직경 (약 20 cm)과 같은 파장을 가지고 있다. 이로 인해 파장이 머리의 직경보다 긴, 약 1500 Hz 이하의 저역에서는 음파는 머리 직경에 따른 음 여과 현상에 영향을 받지 않으므로 양 귀 사이의 위상차 (Phase Difference)가 발생하여 양이 단서 인지의 주 역할을 하게 된다. 그러나 파장이 머리 직경보다 짧아지는 1500 Hz 이상에서는 위상차가 발생하지 않으므로 ITD의 단서가 모호해지게 된다. 따라서 IID가 방위각의 인지를 대신하게 된다. 이는 고음이 머리를 돌아 나아가지 못하고 멈춰지게 되어 머리의 음 여과 현상 (Head Shadow)에 의한 시간차가 점점 커지게 되는 것이다 [8]. 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 [3]. 이런 특성으로 인하여 머리전달함수는 가상음향에서 빼놓을 수 없는 역할을 하며 이를 취득하게 되면 어떠한 음원이 라도 머리전달함수와 컨벌루션 연산을 통하여 청취자에게 가상음원의 효과를 줄 수 있다 [10-11].

$$\begin{aligned} Y_L &= S * hrtf_L(\theta, \phi) \\ Y_R &= S * hrtf_R(\theta, \phi) \end{aligned} \quad (1)$$

여기서 θ 과 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.

2.2. 모노럴 환경에서의 음상 정위

대부분의 연구에서 입체음향 생성을 위하여 사용되는 단서들은 기본적으로 바이노럴 환경, 즉 청취자의 두 귀를 이용한다는 가정을 기본 바탕으로 삼고 있다. 이 때문에 여러 가지 이유로 한 귀만 사용해야 하는 경우이거나 음향 재생을 위해 1채널만 사용하는 기기에서는 이런 단서들을 사용할 수 없다. 따라서 바이노럴과는 별도로 모노럴 환경에서의 음상의 방향 인지와 관련한 연구가 꾸준히 진행되어 왔다. Angell은 한 쪽 귀가 전혀 들리지 않는 피험자를 이용해서 실험을 실시한 결과 순음을 자극했을 때, 정상인 두 귀 청취자와 비교해서 정위가 잘 되고 있음을 나타내었다. Gatehouse도 똑같이 피험자에 의한 실험에서 순음보다도 백색 잡음 쪽이 정위가 좋고, 한 귀 청취자보다도 정상인 두 귀 청취자 쪽이 좋음을 증명하였다. 또 Butler는 한 귀 청취에서는 음상이 정상인 귀 쪽으로 벗어난 것을 나타내었다 [12]. 그러나 Fisher는 학습에 의하여 두 귀 청취와 같은 정도의 정위 정밀도가 얻어지는 것을 나타냈다. 두 귀가 모두 정상인 청력을 가진 피험자의 오른쪽 귀를 막아서 역치를 상승시키고, 자극의 강도를 낮게 해서 의사적으로 한 귀 청취를 모의하였다. 자극은 지속 시간 120 μ s의 잡음을 사용하였으며 사전에 자극을 제시하여 음원 방향을 가르치는 것을 반복하는 강제 학습을 시켰다. 한편 Belendiuk는 자극의 주파수 성분의 영향에 대해서 Fisher와 유사하게 한 귀 청취자를 이용하여 실험하였다. 학습은 전혀 시키지 않고, 자극은 저역 잡음을 사용하였다. 그 결과로 5 kHz를 경계로 하여, 그 이하의 저역 잡음에서는 현저하게 정위가 나빠진다. 역으로 5 kHz 이상의 성분이 포함되어 있으면 정위가 개선됨을 알 수 있다. 한편 Gatehouse는 정중면에서는

한 귀 청취도 두 귀 청취와 같다는 것을 나타내었다. 이상의 연구 성과로 한 귀 청취의 수평면 정위에 대해서 다음과 같이 정리할 수 있다 [12].

- ① 순음보다 복합음이나 광대역 잡음 쪽이 정위는 좋다.
- ② 자극의 주파수 성분에 크게 영향을 받으며, 5 kHz 이상의 성분이 필요하다.
- ③ 수평면 정위에서는 두 귀 청취에 비해 정위가 떨어진다. 그러나 충분히 학습을 하면 동등의 정도를 얻을 수 있다.

III. 제안된 알고리즘

모노럴 환경에서의 음상 정위와 관련한 많은 연구결과에서 살펴보면 한 쪽 귀가 들리지 않거나 귀마개를 착용한 경우에도 일부 청자들은 음의 방향을 느낄 수 있다. 즉 한 귀 만으로도 음원 신호가 귀에 도달하기까지 생기는 변화를 인지함으로써 방향을 판단하게 되며 그 변화는 귓바퀴나 머리 등의 인간의 신체가 가장 크게 관여하게 된다. 이런 특성을 기반으로 하여 모노럴 시스템에서 정확하게 음상을 정위시키고자 하는 기존의 연구는 각 방향에 대한 음의 변화를 강조하는 연구가 주를 이루고 있다. Rao는 가중치를 이용하여 머리전달함수를 최적화시키는 과정에서 cost function을 계산한 후, 이를 최소화시킴으로서 실제로 측정된 머리전달함수와 최대한 비슷하도록 하였다 [4]. Wanrooij는 실험을 통해 모노럴에서 방향 지각에 관련한 단서가 인간의 신체 특성 중에서 머리 및 귓바퀴의 돌출과 관계가 있다고 밝혔다 [5]. 이밖에 Yan은 귓바퀴의 형태를 바꾸어가며 그 영향을 측정하고자 하였으며 [6] Slattery와 Middlebrooks은 모노럴 청자들이 정상인 귀에서 느껴지는 주파수 특성을 사용한다고 제안하였다 [13]. 그러나 기존의 연구들은 다양한 인간의 신체특성을 모두 모델링하기에는 무리가 있으며 단순히 주파수 단서만을 조절함으로써 마스크와 같은 인간의 청각 특성에 따라 해당 단서가 인지되지 못하는 상황이 발생할 수도 있다. 또한 대부분의 연구가 수직면 상에서의 인지를 목적으로 하고 있어 수평면 상에서는 충분한 효과를 거두지 못하고 있다.

이에 본 논문에서는 기존에 연구된 결과를 기반으로 하여 수평면 상에서 머리전달함수가 인간의 청각에 미치는 영향을 고려하고 이를 바탕으로 하여 머리전달함수를 개선하는 알고리즘을 제안하고자 한다. 한 귀 만을 사용하는 모노럴 환경이므로 본 논문에서는 좌/우측 채널의

머리전달함수 중 우측 방향만을 고려한다.

먼저 음상을 정위시키길 원하는 p방향의 머리전달함수 $hrtf(n,p)$ 및 이와 대칭 방향인 q방향의 머리전달함수, $hrtf(n,q)$ 에 푸리에 변환을 적용하여 전력 스펙트럼을 계산한다.

$$X(\omega, m) = \left| \frac{1}{N} \sum_{n=0}^{N-1} hrtf(n, m) e^{-j \frac{2\pi \omega n}{N}} \right|^2 \quad (2)$$

where, $0 \leq n \leq N-1, 0 \leq \omega \leq \frac{N}{2}$

이 때 N은 머리전달함수의 샘플 수를 의미하며 m은 p와 q방향, $X(\omega, p)$ 와 $X(\omega, q)$ 는 p와 q방향에 따른 전력 스펙트럼을 나타낸다. 이어서 p, q 방향 머리전달함수의 주파수 특성을 비교하기 위해 식 (3)을 이용하여 $X(\omega, p)$ 와 $X(\omega, q)$ 의 차를 계산한다.

$$diff(\omega, p) = X(\omega, p) - X(\omega, q) \quad (3)$$

그림 1은 고도 0°에서 p방향을 방위각 60°로 가정할 때 이에 대한 혼돈을 일으키는 q방향, 방위각 120°에 대한 $X(\omega, p)$ 와 $X(\omega, q)$ 차를 나타낸다. 그림 1에서 보면 1.5 kHz 이하의 대역에서는 전력 스펙트럼의 차이가 그리 크지 않는데 이는 1.5 kHz 이하의 대역에서는 방향인지와 관련한 양이 단서 중에 ITD의 영향이 크기 때문으로 방위각이 변해도 그 차이는 거의 변하지 않는다. 또한 15 kHz 이상의 대역에서는 대부분의 방향에서 머리전달함수의 주파수 에너지가 거의 존재하지 않기 때문에 역시 차이가 거의 없다. 그러나 1.5~15 kHz의 대역에서는 각 방향 고유의 피크나 노치로 인하여 큰 변화를 나타냄을 확인할 수 있다. 이 변화는 방향 결정 대역이라는 말로 표현이

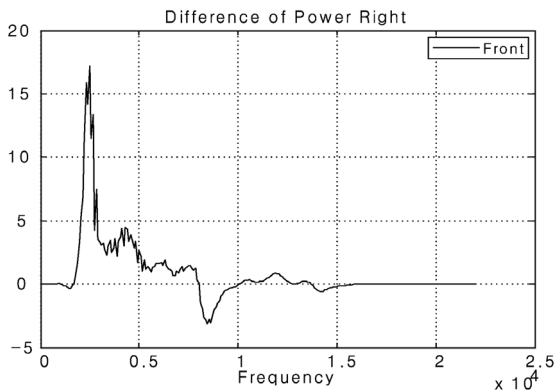


그림 1. 전력 스펙트럼 차 (고도각 0°, 방위각 60° & 120°)
Fig. 1. Difference of power spectrum (Ele. 0°, Azi 60° & 120°).

가능한데, 음상 방향은 음원 방향에 관계없이 각 주파수 대역에 의존하고 있다는 것이다. 즉 방향 결정 대역은 인간의 청각에 가장 많은 영향을 미치는 대역이므로 그에 따른 영향을 강조하게 되면 해당 방향감이 개선될 것이라 생각할 수 있다. 이를 위해 각 방향 머리전달함수의 주파수 특성이 청자의 귀에 미치는 영향을 고려함으로써 원하는 방향과 관련한 인지 특성을 부각시키고자 한다 [4]. 다음으로 사람의 청각 시스템에서의 주파수 변별력을 나타내는 임의의 임계대역 또는 바크 인덱스 (Bark index) z에서 $X(\omega, p)$ 에 대한 p방향의 임계대역 에너지 (critical band intensity)를 계산한다.

$$X_i(z, p) = \sum_{\omega=\omega_{lz}}^{\omega_{hz}} X(\omega, p), \quad (\text{for}, 0 \leq z \leq Z-1) \quad (4)$$

여기서 Z는 신호의 샘플링 주파수에 의해 결정되어지는 전체 임계대역의 수를 나타낸다. 각각의 주파수 성분 ω 에 대한 임계대역 율 (Critical Band Ratio) 또는 바크 율 (Bark Ratio)과 전체 임계대역의 수는 식 (5)를 통해서 계산되어진다. 또한 ω_{lz} 와 ω_{hz} 는 각각 임계대역에 대한 저주파 경계와 고주파 경계를 나타낸다. 이러한 임의의 임계대역에 대한 대역폭은 식 (6)을 통해 계산된다 [14].

$$z/bark = 13 \arctan(0.76f/kHz) + 3.5 \arctan(f/7.5kHz)^2 \quad (5)$$

$$\Delta f / Hz = 25 + 75 [1 + 1.4(f/kHz)^2]^{0.69} \quad (6)$$

이때 외이에서부터 중이까지의 다양한 전달 특성 요소들에 의한 주파수 의존 감쇠특성을 나타내는 감쇠성분, $a_0(z)$ 를 고려하여 임계대역 에너지를 계산하면 다음과 같다 [14].

$$X_a(z, p) = a_0(z) X_i(z, p), \quad (\text{for}, 0 \leq z \leq Z-1) \quad (7)$$

단 이때 $a_0(z)$ 는 외이부터 중이까지의 감쇠특성을 나타낸 것이므로, 신호를 직접 내이로 전달하는 경우 또는 난청환경과 같이 이러한 특성의 변화가 있을 경우를 제외하고는 그 영향을 무시할 수 있다.

다음으로 $X_a(z, p)$ 와 기저막에서의 신호의 에너지의 확산현상을 나타내는 지각적인 에너지 확산함수를 컨벌루션 연산을 수행으로써 기저막에서의 신호 에너지에 대한 응답을 나타내는 청각 자극 에너지를 식 (8)과 같이 계산

한다. 이 청각 자극 에너지는 마스킹과 같은 사람의 비선형적인 청각 영향으로 인해 임계대역 에너지의 특성이 변화하게 되는 것으로서 실제로 인간의 뇌에 전해져 청감을 일으키는 자극이다.

$$X_e(z, p) = \sum_{v=0}^{Z-1} [SF(v, z)X_a(v, p)], \text{ (for, } 0 \leq z \leq Z-1) \text{ (8)}$$

여기서 확산 함수 SF(v, z)는 식 (9)와 같이 표현되어질 수 있으며, Δz = z-v는 바크 인덱스로 표현된 임계대역 간 차이이다 [14].

$$SF(v, z) = 15.81 + 7.5(\Delta z + 0.474) - 17.5\sqrt{1 + (\Delta z + 0.474)^2} \text{ (9)}$$

마찬가지로 혼돈을 일으키는 q방향의 임계대역 에너지, X_a(z, q) 및 청각 자극 에너지, X_e(z, q)를 계산한다.

그림 2는 60°와 120° 방향의 머리전달함수의 임계대역 에너지 및 청각 자극 에너지를 나타낸다. 임계대역 에너지를 살펴보면 11 bark 이하는 주파수 영역에서 1.5 kHz 이하의 대역에 해당하며 앞서 설명한 대로 에너지의 차가 거의 없기 때문에 인간의 청각에도 비슷한 영향을 끼친다. 그러나 12~22 bark는 각 방향에 따라 우세한 대역이 달라지는데, 그림 2의 (a)에서는 12~21 bark는 앞 방향이, 그리고 22, 23 bark 대역은 뒷 방향이 우세한 대역임을 알 수 있다. 그러나 마스킹 등 청각 작용으로 인하여 이웃 bark 대역에 영향을 받게 되므로 이 특성은 달라진

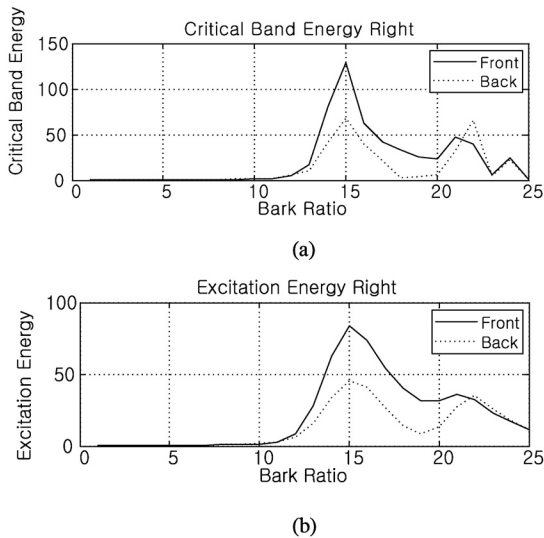


그림 2. 계산된 에너지 (고도각 0°, 방위각 60° & 120°)
 (a) 임계대역 에너지, (b) 청각자극 에너지
 Fig. 2. Calculated Energy (Ele. 0°, Azi 60° & 120°).
 (a) Critical band energy, (b) Excitation energy

다. 이에 최종적으로 22, 23 bark에서 우리가 인지 가능한 에너지는 앞 방향과 뒷방향이 거의 비슷하게 된다.

이러 계산된 p와 q방향의 청각 자극 에너지를 이용하여 원 머리전달함수에 적용될 가중치를 생성한다. 기존에 연구된 바이노럴 환경의 논문 [7]에서는 단순히 두 방향의 비율만을 가중치로 적용을 하였으나 모노럴 환경에서는 ITD와 IID 등의 단서를 이용할 수가 없고, 사운드 신호에서 주파수 에너지가 큰 부분을 과도하게 증폭시킬 경우 정위된 음이 손상될 우려가 있으므로 식 (10)을 이용하여 음이 손상되지 않는 범위 내에서 가중치를 생성하게 된다.

$$eHRTF(f_z, p) = HRTF(f_z, p) \times w(z, p) \\
w(z, p) = \log_{10}((X_e(z, p)/X_e(z, q))^{0.5}) + 1 \text{ (10)} \\
\text{(for, } 0 \leq z \leq Z-1)$$

여기서 f_z는 각 임계대역에 해당하는 주파수 범위를 나타낸다.

그림 3은 계산된 최종 가중치를 나타낸다. 대략 18 bark 주위에서 가장 큰 에너지를 나타내고 있는데 이 대역은 앞에서 구한 전력 스펙트럼의 차가 가장 큰 부분이기도 하다. 즉 앞/뒷 방향의 구분하는 느낌을 갖게 하는 중요한 대역이라 할 수 있다.

그림 4는 고도각 0°, 방위각 60°에 해당하는 개선된 머리전달함수의 주파수 스펙트럼을 나타낸다. 인간의 청각에 거의 영향을 미치지 않는 부분은 원래의 머리전달함수 에너지와 비슷하지만 1.5~9 kHz 이내의 대역은 청자가 지각적으로 인지할 수 있는 대역이므로 강조된 모습을 볼 수 있다. 만약 음상을 정위하고자 하는 방위각이 90°~180°라면 가중치를 통해 각 대역의 에너지를 조절하고 노치 등을 부각시켜 뒷 방향에 따른 느낌을 증가시키게 된다.

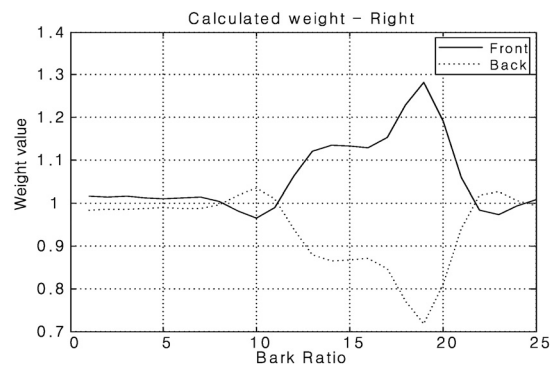


그림 3. 계산된 가중치 (고도각 0°, 방위각 60° & 120°)
 Fig. 3. Calculated weight (Ele. 0°, Azi. 60° & 120°).

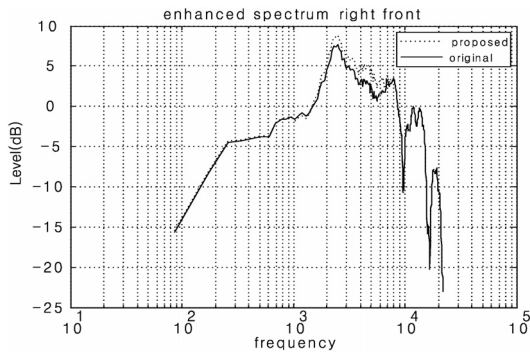


그림 4. 원 HRTF & 수정된 HRTF (고도각 0°, 방위각 60°)
Fig. 4. original HRTF & enhanced HRTF (Ele. 0°, Azi. 60°).

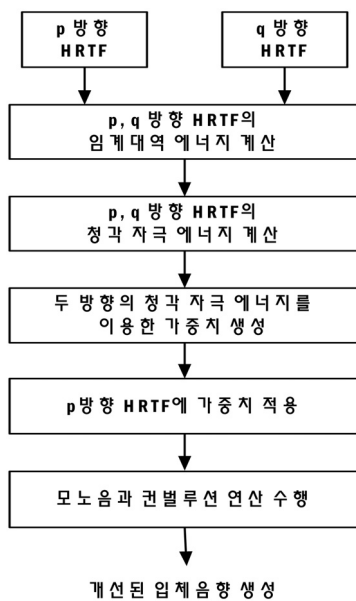


그림 5. 제안된 음상 정위 특성 개선 알고리즘의 전체 구조
Fig. 5. Overall structure of the proposed algorithm.

즉 제안한 방법은 머리전달함수의 전체 주파수 대역 중에서 인간이 해당 방향 인지에 필요한 대역을 찾고 강조한다. 그 결과로 방향감이 개선된 머리전달함수는 사운드 신호에 적용되어 입체음향을 생성하게 된다. 따라서 청자는 강조된 인지 단서를 이용하여 모노럴 환경에서도 음상의 위치를 정확하게 느낄 수가 있게 된다. 본 논문에서 제안된 음상 정위 특성 개선 알고리즘의 전체 구조는 그림 5와 같다.

IV. 실험 및 결과고찰

4.1. 실험 환경

본 장에서는 테스트가 수행된 장소는 일반 생활 잡음이 존재하는 연구실 공간으로 피실험자들은 중앙에 설치된 의자에 편하게 앉도록 하였다. 더불어 조명을 끄거나 눈

표 1. 사용된 머리전달함수

Table. 1. HRTF for simulation.

샘플 수	512 (full) / 128 (compact)
샘플링 주파수	44.1 kHz
샘플당 비트수	16 bits/sample
수평각	0° ~ 360°
고도각	-40° ~ 90°
측정 대상	KEMAR dummy head

표 2. 사용된 음원

Table. 2. Sound sources for simulation.

이름	Vocal	Helicopter
샘플링 주파수	44.1 kHz	
샘플당 비트수	16 bits/sample	
샘플 수	244446	

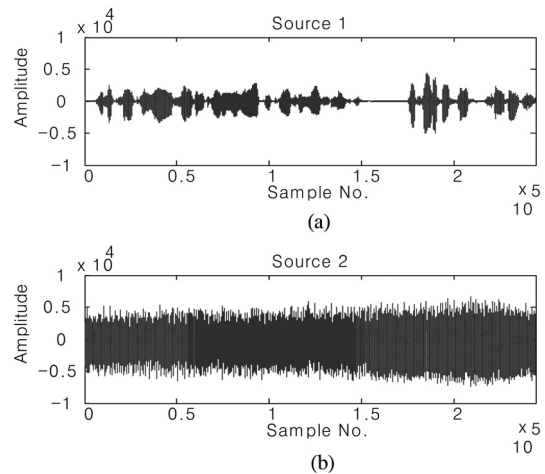


그림 6. 시뮬레이션에 사용된 음원

(a) 음성, (b) 헬리콥터

Fig. 6. Sound sources for simulation.

(a) Vocal, (b) Helicopter

을 가림으로서 실험에 영향을 줄 수 있는 외부 자극을 차단하였다. 사용된 사운드 신호들은 오디오 CD에서 추출된 44.1 kHz, 16 bits/sample 사운드 신호를 이용하였다. 사용된 머리전달함수로는 MIT Media Lab.에서 측정된 것을 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.

비교 대상은 원 머리전달함수 및 기존 방법으로 Kim이 제안한 머리전달함수의 주파수 차를 이용한 방식 [15] 그리고 제안한 방식으로 정면을 방위각 0°, 오른쪽 90°, 뒷면 180°로 가정할 때, 방위각 30°/150°, 45°/135°, 60°/

120°, 75°/115°의 네 방향에 대해서 테스트를 실시하였다. 입력 사운드 신호로는 표 2와 같이 Vocal, Helicopter sound의 2가지를 사용하였다.

그림 6은 시뮬레이션에 사용된 신호인 Vocal과 Helicopter 신호를 보여주고 있다. Vocal 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 Helicopter 신호는 저주파 대역에의 에너지가 큰 신호로서 시간에 따른 변화도가 적은 경우에서의 성능 테스트를 위해 사용하였다.

4.2. 실험 방법

테스트에 참가한 피실험자는 2~30대의 일반 남녀 10명이며 그 중 1명은 후천적인 원인으로 20대 중반에 한쪽 귀의 청력을 상실한 피실험자이다. 이들은 헤드폰으로 테스트를 수행하기 전에 음원 방향 인지와 관련하여 간단한 학습을 수행하였다. 학습은 피실험자의 한쪽 귀에 대한 역치를 높이고자 귀마개로 왼쪽 귀를 막은 후 스피커를 이용하여 1.5 m의 거리를 두고 전방부터 후방까지 오른쪽으로 음을 이동시켜가며 피실험자에게 들려주었다. 이 학습은 피실험자에게 모노럴 환경에 대한 이해를 높이고 한쪽 귀 청취에 대한 상세한 느낌을 부여함으로써 보다 정확한 테스트를 수행함에 그 목적이 있다. 기존의 연구 결과에서 확인할 수 있듯이 학습을 통해 음상 정위의 정확도가 상승할 수 있으므로 이는 매우 필요한 과정이라 할 수 있다. 학습이 종료된 후 본격적인 테스트를 위하여 헤드폰을 착용시킨 후 우측 채널에서만 음을 재생시켜 피 실험자가 이를 청취하는 방법으로 테스트가 진행되었다.

4.3. 음상정위 테스트

본 논문은 모노럴 환경에서 음상을 정위하고자 할 때, 머리전달함수의 지각적인 특성을 이용하여 방향감을 강조하는 알고리즘을 제안한다. 이에 제안된 알고리즘의 성능을 확인하기 위하여 헤드폰 환경에서 음상 정위 테스트를 수행하였다. 앞에서 말한 3가지 방법에 따라 생성된 각 방향의 입체음향 3가지를 들려준 후 각 방향에 대해 인지하는 정확도를 측정하였으며 $\pm 10^\circ$ 내의 인지 오차의 경우는 정답으로 인정하였다.

그림 7과 8은 3가지 방법을 이용하여 생성된 입체음향

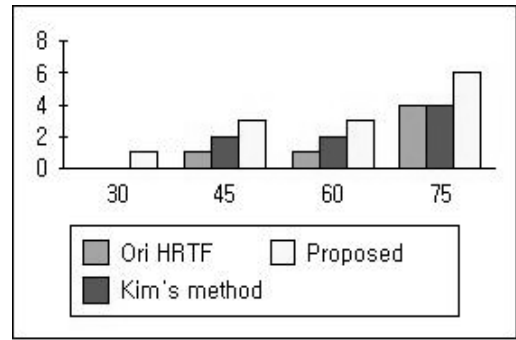


그림 7. 전방향 음상정위 테스트 결과

Fig. 7. Sound localization test result at front azimuth.

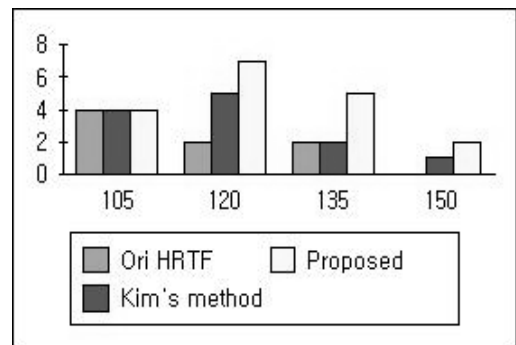


그림 8. 후방향 음상정위 테스트 결과

Fig. 8. Sound localization test result at back azimuth.

에 대한 음상정위 테스트 결과를 나타내고 있다.

만약 사운드 신호가 정면에서 좌/우측으로 이동하면 할수록 이동하는 방향의 귀에 들려오는 음의 느낌은 더욱 명확해지게 된다. 이는 사운드 신호가 귓바퀴 등의 신체 기관의 영향을 덜 받게 되어 음이 감쇄되거나 회절되지 않고 바로 청각 기관에 도달하기 때문이다. 이 특성은 모노럴에서 방향을 인지하기 위한 중요한 단서가 된다. 즉 IID나 ITD 같은 단서를 사용할 수 없기 때문에 음색이나 음의 명확성을 통해 방향을 인지할 수밖에 없기 때문이다. 테스트 결과에 따르면 순수한 머리전달함수 및 기존의 방법을 통해 생성된 음상은 대부분 $60^\circ \sim 120^\circ$ 사이에서만 정확하게 정위된다. 즉 $60^\circ \sim 120^\circ$ 이외의 범위에 음상을 정위시키기 위한 단서를 제대로 부각시키지 못하기 때문에 청자는 명확하게 들리는 음향의 방향만을 구별할 수 있기 때문이다. 그러나 제안된 방법의 테스트 결과를 살펴보면 넓은 범위에 음상이 제대로 정위됨을 볼 수 있다. 이는 머리전달함수를 이용하여 입체음향을 생성할 때 인간의 청각에 미치는 영향을 고려하여 영향이 큰 대역을 부각시킴으로서 결과적으로 인간이 지각하게 되는 방향감을 강조할 수 있기 때문이다.

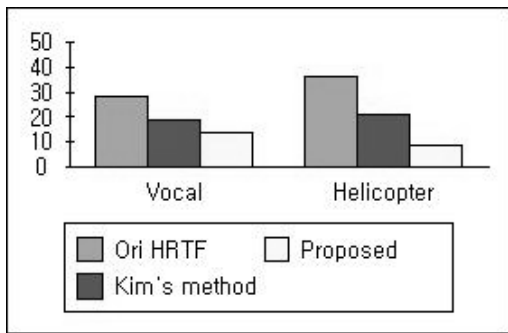


그림 9. 생성된 입체음향의 앞/뒤 혼돈 횟수
Fig. 9. Front/Back confusion result of each method.

그림 9는 음상정위 테스트 결과와는 별도로 생성된 입체음향의 혼돈 정도를 나타내고 있다. 입체음향의 혼돈이란 음원과 청취자의 귀 사이의 거리가 같은 혼돈원추상에서는 시간이나 레벨의 차 등을 이용하여 각 방향의 차이를 구분할 수가 없게 된다. 따라서 원하는 방향이 아닌 서로 대칭되는 방향에 음상이 정위됨을 뜻한다.

결과를 보면 원 머리전달함수에 비해 Vocal의 경우는 50%, Helicopter의 경우는 75%의 혼돈 횟수가 감소하였음을 알 수 있다. 머리전달함수를 사용하여 입체음향을 생성할 경우 앞/뒤 방향에 따라서 주파수 에너지가 우세한 대역이 존재하며 그에 따라 청각에 미치는 영향도 달라진다. 기존의 방법은 앞/뒤 방향의 단서를 사용한다 하더라도 청자가 지각하지 못하게 되어 혼돈을 가져오게 된다. 그러나 제안된 알고리즘은 청자가 앞/뒤 방향에 따라 청자가 지각하는 부분을 강조함으로써 혼돈 횟수를 줄일 수가 있어 더욱더 현실적인 음상 정위를 할 수 있다.

V. 결론

본 논문에서는 헤드폰을 이용한 모노럴 재생 시스템에서 머리전달함수를 이용하여 입체음향을 생성할 때 정확한 방향감을 강조하기 위한 알고리즘을 제안한다. 머리전달함수가 인간의 청각에 미치는 영향을 고려하여 가중치를 생성하고 이를 원 머리전달함수에 적용한다. 이때 보다 정밀한 음상정위를 위하여 대칭방향의 두 머리전달함수의 여기에너지 비율을 고려하게 된다. 제안된 방식을 적용하여 생성된 입체음향은 음상정위 테스트를 수행한 결과 기존의 방법에 비해 정확한 방향감을 생성함을 확인할 수 있었다. 이밖에 피실험자가 인지한 방향이 원 음상의 방향과의 오차가 일정 각도 이상 발생하더라도 앞/뒤 방향에 대한 혼돈 횟수가 크게 감소하였다. 이를

통해 모노럴 환경에서 입체음향 구현 시 본 알고리즘을 적용함으로써 보다 향상된 입체음향을 생성할 수 있을 것이라 생각된다.

감사의 글

이 연구는 2011년도 숭실대학교 교내연구비 지원에 의한 연구임.

참고 문헌

1. 이병호, 김용국, 전찬준, 김홍국, 이용주, 장대영, 강경옥, "헤드폰 청취환경에서의 실감 오디오 재현을 위한 음상 외재화 기법," *한국전자공학학회논문지*, 47권, SP편, 5호, 1-8쪽, 2010.
2. 윤대희, "오디오 신호처리 기술 동향," *전자공학회지*, 31권, 6호, 75-95쪽, 2004.
3. 이기승, 이석필, "원형 머리 모델을 이용한 머리 전달 함수의 보간," *한국음향학회지*, 27권, 7호, 333-341쪽, 2008.
4. K. R. Rao, J. Ben-Arie, "Optimal Head Related Transfer Functions for Hearing and Monaural Localization in Elevation: A Signal Processing Design Perspective," *IEEE Trans. on Biomedical Engineering*, vol. 43, no. 11, pp. 1093-1105, November, 1996.
5. M. V. Wanrooij, A. J. Van Opstal, "Contribution of Head Shadow and Pinna Cues to Chronic Monaural Sound Localization," *The Journal of Neuroscience*, vol. 24, no. 17, pp. 4163-4171, Apr. 2004.
6. T. Yan, T. Kochiyama and J. Wu, "Ability to Discriminate and Elevation in Sound: The Effects of an Altered Pinna in Subjects with Monaural Hearing," *Int'l Conference on Complex Medical Engineering*, pp. 1454-1457, 2007.
7. 구교식, 차형태, "개선된 머리전달함수를 이용한 3차원 입체음향 성능 개선 연구," *한국음향학회지*, 28권, 6호, 557-565쪽, 2009.
8. C. I. Cheng and G. H. Wakefield, "Introduction to Head-Related Transfer Functions(HRTFs)," *J. Audio Eng. Soc.*, pp. 231-249, April, 2001.
9. Durand R. Begault, *3-D Sound for Virtual Reality and Multimedia*, NASA, 2000.
10. 박장식, 김현태, "합성 스테레오 방식 3차원의 입체음향의 실시간 구현을 위한 머리전달 함수의 IIR 필터 설계," *한국콘텐츠학회논문지*, 5권, 6호, 74-86쪽, 2005.
11. D. N. Zotkin, R. Duraiswami and L. S. Davis, "Rendering Localized Spatial Audio in a Virtual Auditory Space," *IEEE Trans. on Multimedia*, vol. 6, no. 4, pp. 553-564, Aug. 2004.
12. 강성훈, 강경옥, *입체음향*, 기전연구사, 1997.
13. Slattery III WH, Middlebrooks JC, "Monaural Sound Localization: acute versus chronic unilateral impairment," *Hearing Res.*, vol. 75, pp. 38-46, May 1994.
14. E. Zwicker, H. Fastl, *Psychoacoustics : Facts and Models*, Springer 2nd Edition, 1999.
15. 김경훈, 김시호, 배건성, 최송인, 박만호, "헤드폰 기반의 입체음향 생성에서 앞/뒤 음상정위 특성 개선," *한국통신학회논문지*, 29권, 8호, 1142-1148쪽, 2004.

저자 약력

•구 교 식 (Kyo-Sik Koo)

2005년: 송실대학교 정보통신전자공학부 (학사)

2007년: 송실대학교 전자공학과 (석사)

2007년 ~ 현재: 송실대학교 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: 송실대학교 정보통신전자공학부 교수

* 주관심 분야: Multimedia Systems and Applications Audio and Video
Signal Processing, Communication System, ASIC Implementation
of Digital System