

헤드폰 기반의 입체음향 생성에서 앞/뒤 음상정위 특성 개선

정희원 김 경 훈*, 김 시 호**, 배 건 성*, 최 송 인***, 박 만 호***

Improvement of front-back sound localization characteristics in headphone-based 3D sound generation

Kyung-hoon Kim*, Si-ho Kim**, Keun-sung Bac**, Song-in Choi***, Man-ho Park*** *Regular Members*

요 약

일반적으로 2채널의 헤드폰에서 입체음향을 생성하기 위해서는 HRTF(Head Related Transfer Function) DB를 이용한 바이노럴(binaural) 합성기법을 사용하게 된다[1] 그러나 비개인화된 HRTF DB를 이용하여 생성된 2채널 입체음향은 사용자에 따라 앞뒤 또는 상하 방향에 대해서 혼돈을 가져와 입체감을 저하시킬 수 있다 본 논문에서는 앞/뒤 음상정위의 혼돈을 개선하기 위해 스펙트럼 단서를 부각시키는 기존의 알고리즘을 분석하고 이를 토대로 혼돈 방향의 HRTF의 지이 성분을 이용하여 스펙트럼 특성을 강조하는 방법을 제안하였다 또한 제안한 방식이 기존의 방법에 비해 앞/뒤 방향감이 개선됨을 청감평가 실험을 통해 확인하였다

Key Words 3D sound, HRTF, source localization, cone of confusion

ABSTRACT

A binaural filtering method using HRTF DB is generally used to make the headphone-based 3D sound. But it can make some confusion between front and back directions or between up and down directions due to the non-individual HRTF depending on each listener. To reduce the confusion of sound image localization, we propose a new method to boost the spectral cue by modifying HRTF spectra with spectrum difference between front and back directions. Informal listening tests show that the proposed method improves the front-back sound localization characteristics much better than the conventional methods.

1 서 론

일반적으로 입체음향이란 음원이 존재하는 공간에 직접 위치하지 않은 청취자가 재생된 음향을 들었을 때 음향으로부터 공간적 인식, 즉 방향감, 거리감, 공간감을 인지할 수 있는 음향을 의미한다^[1] 이때 사람이 공간적 인식을 할 때 사용하는 주된 단서로는 음이 두 귀에 도달하는 데 걸리는 시간차(ITD: Inter-aural Time Difference)와 두 귀에서

느끼는 음의 세기차(IID: Inter-aural Intensity Difference) 그리고 상악신과 귓바퀴에서의 반사, 회절에 의한 스펙트럼 단서(spectral cue) 등이 있다 이러한 특성을 포괄적으로 가지고 있는 것이 HRTF(Head Related Transfer Function)인데, 이는 음원의 방향 및 위치에 따라 변하고 청취자 개인마다의 고유한 특성을 가지고 있다 따라서 비개인화 된 HRTF DB를 사용하여 음상을 정위시킬 경우 개인의 고유한 스펙트럼 단서가 충분히 반영되지 않아

* 삼성달레스. 제조개발본부 Product Eng 팀, ** 경북대학교 전자공학과 신호처리연구실, *** 한국전자통신연구원 이송A/V연구실
논문번호 030524-1125, 접수일자 2003년 12월 5일

※본 연구는 한국전자통신연구원 이송통신연구소 이송부호화연구실의 연구비 지원으로 수행되었습니다

방향 지각에 혼동을 가져올 수가 있다. 그림 1과 같은 혼돈원추(cone of confusion) 상에 위치한 음원에 대해서는 청취자의 귀에 이르는 공간 사이의 거리가 같아지므로 ITD, IID 정보로는 방향의 차이를 구분할 수 없게 된다^[1,2]. 따라서 혼돈원추 상에 있는 음원을 구별할 수 있는 정보로는 스펙트럼 단서를 이용하는데, 이는 각 개인의 머리와 귓바퀴, 어깨 등 신체특성 차이에 크게 좌우 되기 때문에 MIT HRTF와 같이 비개인화 된 DB를 이용해 음상을 정위시킬 경우 특정 방향(예:0-180도, 30-150도)에 대한 정위에 한계를 가질 수 밖에 없다. 특히 헤드폰을 통해 다채널 입체음향을 재생하는 경우 앞/뒤 방향 채널의 음상이 제대로 구별되지 않아 입체감을 떨어뜨리게 된다. 각 청취자 자신의 HRTF를 측정하여 사용한다면 문제는 해결될 수 있겠지만, 모든 방향에 대한 HRTF를 개인마다 측정하고 이를 DB화 하는 것은 현실적으로 불가능하다. 따라서, 본 논문에서는 앞/뒤 음상정위의 혼돈을 개선하기 위해 스펙트럼 단서를 부각시키는 기존의 알고리즘을 분석하고 이를 토대로 혼돈 방향의 HRTF의 차이 성분을 이용하여 스펙트럼 특성을 강조하는 방법을 제안하고, 청감평가 실험을 통해 제안한 방식이 기존의 방법에 비해 앞/뒤 방향감이 개선됨을 보였다.

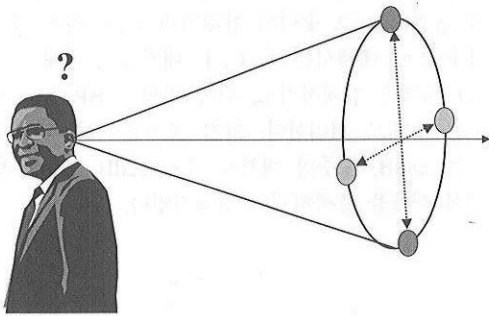


그림 1. 혼돈 원추

본 논문의 구성은 다음과 같다. 2장에서 HRTF DB와 입체음향 생성 방식에 대해 설명하고, 3장에서는 앞/뒤 방향의 스펙트럼 단서를 부각시키는 기존의 알고리즘을 분석하며 이를 토대로 앞/뒤 스펙트럼 차를 이용하여 스펙트럼 단서를 부각시키는 방법을 제안한다. 4장에서는 제안한 방식에 대한 실험

결과를 제시하고 마지막으로 5장에서 결론을 맺는다.

II. HRTF를 이용한 입체음향 생성

임의의 방향에서 방사된 소리는 방사되는 공간, 사람의 머리, 몸통, 귓바퀴, 중이를 거쳐 고막에 전달되고 이 과정에서 반사와 회절현상을 일으키며 두 귀에 도달한다. 이렇게 음원과 청취자 사이의 공간적 특성을 포괄적으로 나타내는 HRTF는 음이 두 귀에 입사하는 여러 각도에 대해 각각 시간영역에서의 임펄스 응답을 측정된 HRIR(Head Related Impulse Response) DB 형태로 만들어져 사용된다. 본 논문에서는 인터넷상에 공개되어 있는 MIT HRTF DB^[3,4]와 CIPIC HRTF DB^[5,6]를 사용하였는데, 표 1은 MIT HRTF DB와 CIPIC HRTF DB의 특징을 비교하여 나타내었다.

입체음향을 생성하기 위해서는 크게 방향감, 공간감, 거리감을 제어하는 과정을 거치게 된다. 먼저, 모노 사운드로 입력되는 신호에서 원하는 위치로 음상을 정위시키기 위해서는 정위시킬 방향의 HRTF와 바이노럴(binaural) 필터링을 수행하게 된다. 두 외이에 도달하는 물리적 단서로서 측정된 HRTF의 방향에 따른 필터링은 보통 컨볼루션 연산으로 수행되는데, 컨볼루션 연산은 식 (1)과 같이 표현된다.

표 1. MIT/CIPIC HRTF DB의 사양 비교

	MIT	CIPIC
샘플수	512(full) / 128(compact)	200
측정용 신호	MLS	Golay-code signal
샘플당 비트수	16	16
샘플링 주파수	44,100Hz	44,100Hz
수평각	0°~360°(5°간격)	-80°~80°(50°까지, 약 5°간격)
고도각	-40°~90°(10°간격)	-45°~231°(25가지, 5.625°간격)
측정대상	KEMAR dummy head Left: normal pinna Right: large red pinna	2 KEMAR dummy heads : large/small pinna, 43 human subjects(27men, 16women)

$$y(n) = \sum_{m=0}^{M-1} h(m)x(n-m) \quad (1)$$

여기서, $x(n)$ 은 음원샘플, $h(n)$ 은 HRIR 필터계수, $y(n)$ 은 필터링 된 출력 유흥신호, M 은 HRIR 필터의 탭 수이다 단일 음원에 대한 바이노럴 필터링은 그림 2의 같이 좌우 각 채널에 대한 HRTF 쌍을 이용하여 수행된다. 그러나 음원이 정면이나 후면 혹은 수직면 상이나 두 귀 축을 중심으로 원추형 방향에 존재할 때에는 바이노럴 필터링된 결과물을 헤드폰으로 들어보면 앞/뒤 방향과 상/하에 대한 혼돈이 생기게 된다. 즉, 정면이나 후면 또는 수직면 상에 존재하는 음원에 대해서는 방향지각의 주요 요소인 ITD나 IID의 차이가 없으므로 이들 외의 복합적인 요소들로 이뤄진 스펙트럼 단서나 방향 지각의 중요한 요소가 된다. 그러나 자신의 HRTF가 아닌 머니헤드에서 측정된 HRTF에서 생기는 오차는 정확한 스펙트럼 단서를 제공하지 못하므로 방향지각에 혼돈을 일으키게 된다.



그림 2 단일 음원에 대한 바이노럴 필터링 블록

한편, 공간의 형태 및 벽면 재질의 흡음율 등에 따라 선평되는 소리는 공간정보를 반영한 왜곡 특성을 보이게 되며, 이를 통해 청취자는 거리감이나 공간감을 지각하게 된다. 따라서 방향감만 고려된 음상정위 결과물으로는 현장감 있는 소리를 재생하기가 어려우며 특히 헤드폰 재생시에는 소리가 머리 안에 머물러 입체감을 저하시키기 때문에 잔향기를 사용하여 청취공간에 대한 공간정보를 부여해야 한다. 그림 3은 실내공간에서의 일반적인 음향 반사특성을 보이고 있는데 잔향기를 통해 이러한 공간정보를 부여할 수 있다.

거리감을 세이하기 위해서는 일반적으로 소리의 세기 기준 음에 대해 2배 거리가 될 때마나 6dB 씩 감쇄됨을 설명하는 역지승 법칙을 사용한다^[7]. 또한 잔향이 존재할 경우에는 직접음과 잔향음의 비율이 거리감 형성에 매우 중요한 요소로 작용하기 때문에 이를 이용한다. 잔향 레벨이 실내의 모든 공간에 대해 거의 일정한데 반해 직접음은 음원에서 멀어질수록 감소하게 되므로 직접음 내 잔향음

의 비는 거리가 멀수록 작아지는 특성을 나타낸다^[2].

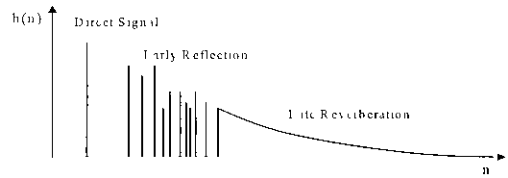


그림 3 실내에서의 음향 반사 특성

III 앞/뒤 방향의 음상정위감 개선

1 앞/뒤 방향감 개선을 위한 기존 알고리즘

비개인화된 HRTF의 앞/뒤 방향감을 개선하기 위해 스펙트럼 단서를 이용하는 기존의 알고리즘들을 정리하면 몇 가지로 요약될 수 있다. 첫째로 알고리즘으로 앞/뒤의 방향지각을 위한 스펙트럼 단서가 특정 주파수 대역에 있다는 지리음향의 연구결과를 바탕으로, 주어진 HRTF DB에서 대역별로 증폭/감쇄를 달리해 주는 방법이 있다^[8]. 그림 4와 같이 HRTF를 5개의 주파수 대역으로 분할하여 각 주파수 대역에 대하여 대역동적 필터링을 하게 된다. 필터링 과정은 원래의 HRTF가 4개의 대역통과필터와 1개의 고역통과필터가 각각 병렬로 구성된 필터뱅크를 통과하고 그 출력이 합해지게 된다. 앞쪽 방향의 HRTF에 대해서는 A, C, E 대역을 증폭해 주고 B, D 대역을 감쇄시키고, 뒤쪽 방향의 HRTF에 대해서는 반대로 처리한다. 크기 조절은 저주파 대역에서는 ±8dB, 고주파 대역에서는 ±12dB 정도의 범위에서 레벨을 증폭하거나 감쇄시킨다.

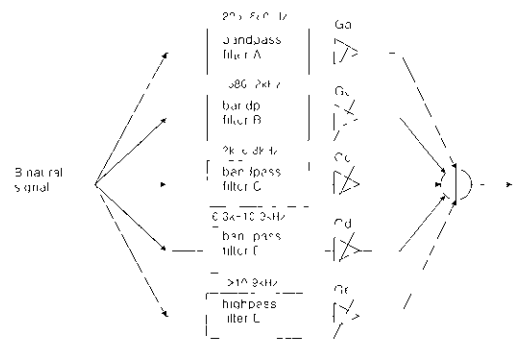


그림 4 앞/뒤 방향감 개선을 위한 대역동적 필터링 블록

또 다른 방식으로는 HRTF가 지지는 개인별 차이 중에서 그림 5처럼 귀 뒷부분의 돌출된 각도 (protrusion angle)의 차이가 앞뒤 빙향 지각에 영향을 미치는 것으로 보고 그 특성을 부각시키는 방법이다⁹⁾ 귀의 돌출 각도가 작은 것에 비해 큰 것이 앞뒤 방향을 잘 구분 하는 것으로 보고 큰 각도에 대한 주파수 특성을 부각시키게 된다 돌출된 정도에 따른 개인별 HRTF를 비교했을 때 저주파 영역에서는 앞 빈구의 방위각에 대한 HRTF가 반대 방향보다 저주파에서 부각 되고, 특정한 주파수 영역에서 감소하는 특성을 나타낸다 따라서 이러한 앞뒤 빈구에 따른 HRTF의 주파수 특성을 강조하여 앞뒤 음성정위의 명료도를 높이게 된다

마지막으로 혼돈을 야기하는 두 빙향의 HRTF 스펙트럼 차이를 부각시키기 위해 식 (2)와 같은 방법으로 HRTF를 수정하는 방법이 있다¹⁰⁾ 음상을 상위시각 p 방향의 HRTF $H_p(f)$ 를 정규화 시키고 그 스펙트럼 크기를 적당한 크기(m)로 지수승하이 가중치 함수를 만들고 이를 원래의 HRTF에 곱하여 $H'_p(f)$ 로 HRTF를 수정한다 그러면 스펙트럼 상의 피크(peak)나 노치(notch) 같은 성분이 강조되어 길게 피적으로 혼돈방향의 두 HRTF의 스펙트럼 차이가 부각되는 효과를 얻게 된다

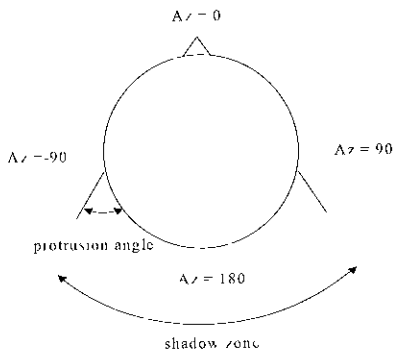


그림 5 귀의 돌출 각도와 shadow zone

$$\begin{aligned}
 H'_p(f) &= W_p(f) H_p(f) \\
 W_p(f) &= \left| \hat{H}_p(f) \right|^m \\
 \hat{H}_p(f) &= H_p(f) / \max_f |H_p(f)|
 \end{aligned} \quad (2)$$

2 제안된 앞뒤 방향의 음성정위 개선 알고리즘 기존 알고리즘의 분석을 바탕으로 하이 혼돈을 일으키는 방향의 HRTF 쌍의 유사 성분을 제거하고 해당 방향의 스펙트럼 특성을 확인하기 위해 두

DB의 스펙트럼 차를 분석하였다 먼저 식 (3)과 같이 상위시각 방향 p 의 HRTF인 $HRTF_p(f)$ 와 앞뒤 쌍을 이루는 대칭 방향의 HRTF인 $HRTF_{p_sym}(f)$ 의 크기 차 $DIFF_p(f)$ 를 분석하였다

$$DIFF_p(f) = |HRTF_p(f)| - |HRTF_{p_sym}(f)| \quad (3)$$

그림 6은 고도 0도에서 각 방위각에 대한 앞뒤 방향의 HRTF 스펙트럼 차를 나타낸 것이다 방위각이 0도에서 80도까지 증가함에 따라 앞뒤 HRTF 스펙트럼 차가 변하는 것을 볼 수 있는데, 2kHz 이하에서는 방위각이 변해도 스펙트럼 차는 큰 변화를 보이지 않으나 2kHz 이상에서는 다소 큰 변화를 보이는 것을 알 수 있다 그러나 방위각이 변함에 따라 두 스펙트럼 차이가 변화를 보이는 주파수 대역은 대체로 일정하게 나타난다 이는 특정 주파수 대역이 앞뒤의 방향지각에 영향을 미치는 스펙트럼 뒤서가 됨을 보여준다 따라서 이러한 스펙트럼 특성을 원래의 스펙트럼에 반영하기 위해 식 (4)와 같은 방식으로 HRTF를 수정하였다

$$\begin{aligned}
 HRTF'_p(f) &= HRTF_p(f) \times W_p(f) \\
 W_p(f) &= \begin{cases} \log_2(2^{DIFF_p(f)} + 1), & DIFF_p(f) \geq 0 \\ 2^{DIFF_p(f)}, & DIFF_p(f) < 0 \end{cases} \quad (4)
 \end{aligned}$$

이때 원래의 HRTF에 곱해지는 가중치 함수로 $DIFF_p(f)$ 를 지수로 하는 지수함수를 사용하는데 $DIFF_p(f)$ 가 0보다 큰 경우에 원음의 스펙트럼의 피크나 노치 성분들이 과도하게 증폭될 수가 있다 이러한 급격한 변화는 모노 음원과의 선분루신 피정에서 특정 주파수 성분이 과도하게 부각되어 음색이 듣기에 힘들 정도로 변하게 된다 따라서 스펙트럼의 차가 0보다 큰 경우에는 지수형태가 아닌 갈수록 완만한 증가 추세를 보이는 로그함수를 통해 그 변화율을 낮추고 특성 구간의 갑작스러운 변화가 아닌 주위 값들과 완만한 변화를 수도록 하였다 그림 7(a)에 실선은 고도 0도, 방위각 0도의 180도의 HRTF 스펙트럼 차 $DIFF_0(f)$ 를 나타내며, 점선은 식 (4)의 방법으로 만든 가중치 함수를 나타내고 있다 피크와 노치들이 가중치 함수로 바뀌면서 원만하게 나타남을 확인할 수 있다 그림 7(b)에서는 원래의 HRTF 스펙트럼이 가중치에 의해 스펙트럼

이 변화되는 것을 보여주고 있다. 1kHz 이상의 피크와 노치 성분들이 강조되면서 원래의 스펙트럼이 변화된 것을 볼 수 있다. [10]의 경우 단순히 피크나 노치가 날카롭게 변하는 반면 제안한 방식은 완만한 스펙트럼 변화를 나타내기 때문에 음질변화의 문제를 다소 줄일 수 있었다.

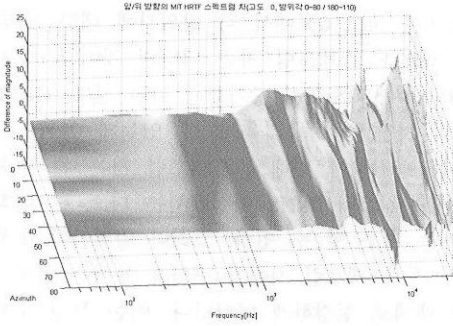
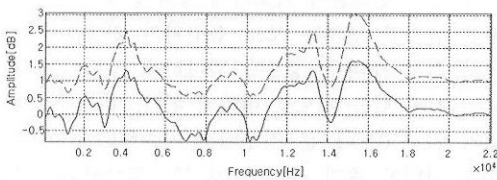


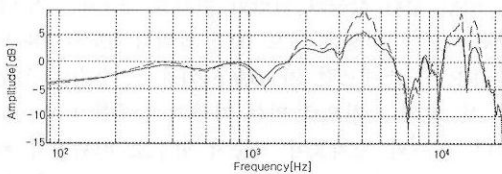
그림 6. 앞뒤 방향의 MIT HRTF 스펙트럼 차(고도: 0°, 방위각: 0°~80°/110°~180°)

IV. 청감평가 실험

제안된 알고리즘의 성능을 평가하기 위해 비전문가 6명을 대상으로 청감평가를 실시하였다. 비교대상은 단순한 HRTF 필터링, 앞에서 소개한 기존 알고리즘들 가운데 가장 양호한 성능을 보이는 대역별로 가중치를 주는 방식, 그리고 제안한 알고리즘으로 하였다. 평가방법은 웹에 공개되어 있는 MIT HRTF DB와 CIPIC HRTF DB 각각에 대한 잔향을 준 경우와 그렇지 않은 경우로 나누어 실험하였



(a) 스펙트럼 차(실선)와 가중치 함수(점선)



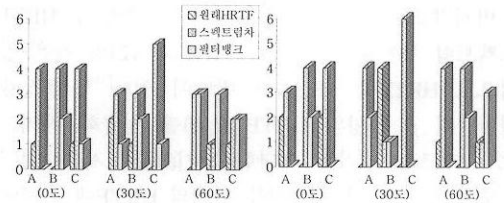
(b) 원래의 HRTF(실선)와 수정된 HRTF(점선)

그림 7. 수정된 HRTF 스펙트럼 특성(0도)

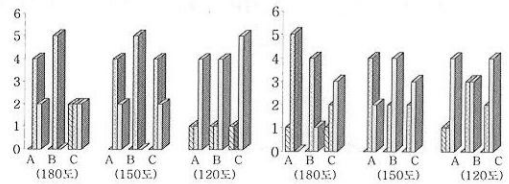
으며, 0°/180°, 30°/150°, 60°/120°의 3가지 방향에 대해 실험하였다. 입력 음원으로는 표 2와 같이 백색잡음, 음성, 효과음(총소리) 3가지를 사용하였다. 앞에서 말한 3가지 방법에 따라 생성된 앞뒤 방향의 입체음향 3가지를 들려 준 뒤 앞/뒤 각각의 방향감이 개선되는 정도를 순위로 나타내도록 하였다.

표 2. 실험에 사용된 입력음

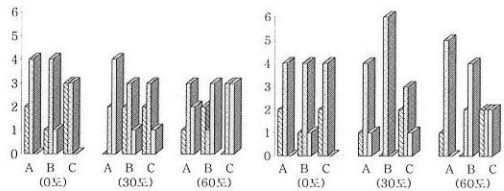
Sample A	백색잡음
Sample B	효과음(총소리)
Sample C	음성



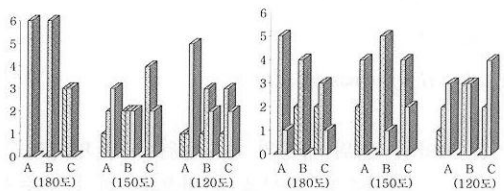
(a) MIT HRTF DB 전방감 (좌:무잔향, 우:잔향)



(b) MIT HRTF DB 후방감 (좌:무잔향, 우:잔향)



(c) CIPIC HRTF DB 전방감 (좌:무잔향, 우:잔향)



(d) CIPIC HRTF DB 후방감 (좌:무잔향, 우:잔향)

그림 8. 청감평가 결과

그림 8은 각 DB에 대해 수행한 실험의 결과를 보여주고 있다. 음질의 변화에 대해서는 반응의 차이가 있었으나 제안한 방법이 MIT 및 CIPIC 두 DB에서 앞쪽 방향감 개선의 효과가 제일 큰 것으로 나타났다. 뒤쪽 방향의 방향감 개선에 대해서는 측면각도에 대해서 필터뱅크에 의한 방식이 제안한 방식보다 다소 구분이 용이한 것으로 나타났다. 그러나 잔향이 첨가되면서 그 차이가 조금씩 줄어드는 것을 볼 수 있다. 한편 입력 음원에 따라 성능의 차이를 보이고 있는데 이는 스펙트럼을 통한 방향감 개선의 방법이 음원의 스펙트럼 특성에 영향을 받기 때문으로 생각된다.

V. 결론

본 논문에서는 비개인화 된 HRTF DB를 이용한 헤드폰 기반의 입체음향 재생 시스템에서 앞/뒤 방향의 음상정위 혼돈을 개선하기 위한 방법을 제안하였다. 먼저 기존의 알고리즘에서 혼돈 방향의 HRTF의 스펙트럼 차에 대한 분석을 수행하고, 이를 토대로 스펙트럼 차를 원래의 스펙트럼에 적절히 반영하여 HRTF를 수정하는 방식을 사용하였다. 성능평가 실험에서 제안한 방법이 기존의 알고리즘 중 비교적 양호한 특성을 나타내는 대역별 가중치 방법보다 개선된 성능을 보였다. 따라서 제안한 방식을 이용할 경우, 헤드폰 기반의 입체음향 재생 시스템에서 문제되는 앞/뒤 방향감의 혼돈 문제를 개선할 수 있을 것으로 판단된다.

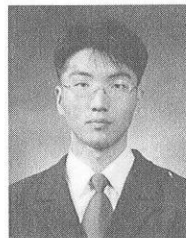
참고 문헌

[1] 강성훈, 강경욱, 입체음향, 기전연구사, 1997.
 [2] Begault, Durand R, *3-D Sound for Virtual Reality and Multimedia*, Academic press, Boston, MA, 1994.
 [3] Gardner. W, "HRTF Measurement of a KEMAR Dummy-Head Microphone," Technical Report# 280, MIT Media Lab, 1994.
 [4] <http://sound.media.mit.edu/KEMAR.html>
 [5] V.R. Algazi, R.O. Duda and D.M. Thompson, "The CIPIC HRTF Database," IEEE Workshop on Application of Signal Processing to Audio and Acoustics 2001, 2001.

[6] http://interface.cipic.ucdavis.edu/CIL_tutorial/3D_HRTF/3D_HRTF.htm
 [7] 김영오, 고대식, "멀티캐스트 화상회의를 위한 3-D 음향시스템 설계," 한국통신학회 논문지, 25(1B), pp.71-76, 2000.
 [8] Chong-Jing Tan and Woon-Seng Gan, "User-defined spectral manipulation of HRTF for improved localization in 3D sound systems," ELECTRONICS LETTERS, 34(25), pp.2387-2389, 1998.
 [9] Navarum Gupta, Armando Barreto, "Spectral Modification of Head-Related Transfer Functions for Improved Virtual Sound Spatialization," ICASSP2002, Vol.2, pp.1953-1956, 2002.
 [10] Ming Zhang, Kah-Chye Tan and M.H.Er "A Refined Algorithm of 3D Sound Synthesis," ICSP98, pp.1408-1411, 1998.

김 경 훈(Kyung-hoon Kim)

정희원



2002년 2월: 경북대학교 전자공학과 졸업
 2004년 2월: 경북대학교 전자공학과 석사
 2004년 3월~현재: 삼성탈레스 제2개발본부 Product Eng. 팀

<관심분야> 음성신호처리, 디지털신호처리, 오디오 코딩 등

김 시 호(Si-ho Kim)

정회원

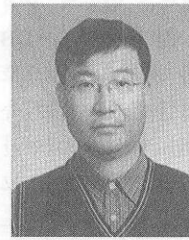


1998년 2월: 경북대학교 전자공학과 졸업
2001년 2월: 경북대학교 전자공학과 석사
2001년 3월~현재: 경북대학교 전자공학과 박사과정

<관심분야> 음성신호처리, 오디오/음성 코딩, 적응신호처리, 오디오 워터마킹, DSP 등

최 승 인(Song-in Choi)

정회원



1982년 2월: 광운대학교 응용전자공학과 졸업
1987년 2월: 광운대학원 전자계산기공학과 석사
1982년 7월~현재: 한국전자통신연구원 이동A/V연구팀 책임연구원

<관심분야> 이동통신, 음성신호처리, 영상신호처리 등

배 건 성(Kcun-sung Bae)

정회원



1977년 2월: 서울대학교 전자공학과 졸업
1979년 2월: 한국과학기술원 전기 및 전자공학과 석사
1989년 5월: University of Florida 공학박사
1979년 3월~현재: 경북대학교 전자·전기공학부 교수

<관심분야> 음성신호처리, 디지털신호처리, 디지털통신, 웨이브렛이론, 오디오신호처리 등

박 만 호(Man-ho Park)

정회원



1998년 2월: 경북대학교 전자공학과 졸업
2000년 2월: 경북대학교 전자공학과 석사
2000년 3월~현재: 한국전자통신연구원 이동A/V연구팀 연구원

<관심분야> 이동통신, 음성신호처리, 음향신호처리 등