

헤드폰을 이용한 가상음향 재현의 주관적, 객관적 평가 Objective and Subjective Test of a Virtual Sound Reproduction Using a Headphone

최 원 재*·김 상 명**
Won-jae Choi and Sang-Myeong Kim

Key Words : Virtual Sound (가상음향), Headphone(헤드폰), Sound Image(음상), Externalization(머리 밖 음상정위)

ABSTRACT

The headphone is regarded as the most effective means for reproducing 3-dimensional virtual sound due to its channel separation property. However, there still exist several serious problems in headphone reproduction, such as, 'front-back confusion' and 'in-head localization'. These well-known problems are in general assessed by the subjective test that is based on human judgment. In this paper, an objective test is conducted in parallel with the subject test in order to validate the objective reproduction performance. Such a combined approach may be a more scientific and systematic approach to the reproduction performance.

1. 서 론

가상음향이란 음향 신호를 이용하여 사람으로 하여금 방향감, 공간감, 그리고 거리감을 느끼게 해주는 것을 의미한다. 기존의 시각에 의존한 가상현실표현의 한계를, 음향을 가미함으로써 좀더 풍부한 느낌을 얻을 수 있고, 가정의 홈 오디오에서부터 산업, 군사용 시뮬레이션까지 그 적용의 범위는 매우 광범위하다.

이러한 가상음향을 실현 할 수 있는 가장 확실한 방법은, 특정 음향 상황에서 인간의 고막의 진동을 측정하여 같은 신호로서 고막을 진동시키면 청취자는 같은 음향 상황에 있는 것으로 느끼게 될 것이다. 이러한 개념에서 출발한 기술이 Binaural technology 이다. 이 기술은 공간에서 사람의 귀에 이르는 전달 함수인 머리전달함수(HRTF : Head-Related Transfer Function)로 구현이 가능하다. 컴퓨터 환경이 발달 되면서 디지털 필터로 HRTF 를 구현하여 실제의 입체 음향을 생성하는 연구가 활발히 진행되고 있다.

Binaural technology 는 주로 헤드폰을 이용하여 구현을 하는데 이는 양쪽 귀에 각각 하나의 음원만이 배치되어있고 상호간의 간섭이 없기 때문에 좀더 수월한 입체음의 구현이 가능하기 때문이다. 하지만 헤드폰 재현 시 앞의 음원이 뒤쪽에서 들리는 현상 (또는 그 반대 현상)인 앞뒤혼동(front-back confusion), 공간 음의 재현 시 음원이 머리 속에서 들리는 듯한 현상인 머리 내 음

상 정위(In-head localization) 등 해결해야 할 문제가 남아있다.

입체 음향은 다분히 사람의 지각에 의존하기 때문에 Wightman and Kistler [1], Martin et al [3], Pernaux et al [6], Hartmann et al. [9] 등이제까지 많은 연구자들은 사람의 판단에 의존하는 주관적 평가를 해오고 있다.

본 논문은 이러한 주관적 평가와 더불어 재현 시스템의 정확성을 나타낼 수 있는 객관적 평가를 병행하여 평가의 신뢰도를 높였다. 또한, 원음재생필터를 도입하여 머리 밖 음상정위 성능을 높일 수 있음을 보였다.

2. 헤드폰 재현 실험

본 연구에서는 헤드폰을 이용한 입체음향의 재현 평가를 위하여 27 세에서 37 세 사이의 정상 청력을 가진 5 명의 남자를 시험하였다.

실험 과정은 1) 개개인의 헤드폰-귀 충격 응답을 측정하고, 2) 그 전달함수의 역함수를 설계한다. 이를 이용하여 헤드폰 재현의 3) 객관적 그리고 4) 주관적 테스트를 실시 한다. 다음은 그 과정과 세부 내용이다.

2.1 헤드폰-귀도 충격응답(Headphone-ear canal impulse response) 측정

입체 음향의 재현을 위하여 원하는 입력신호를 고막에 전달해 주는 것이 우리가 목적하는 바이다. 하지만 음향 재현 시스템에서는 트랜스듀서의 전기적 특성과 입력부(스피커, 헤드폰 등)에서 출력부(마이크로폰 또는 고막)까지의 음향적 특성이

* 광주과학기술원 기전공학과
E-mail : wjchoi@kjist.ac.kr
Tel : (062) 970-2409, Fax : (062) 970-2384

** 광주과학기술원 기전공학과 조교수

필연적으로 포함되어 있어 이를 보상해줄 역필터를 설계해야 한다. 역필터를 설계하기 위하여 헤드폰에서 고막까지의 충격 응답을 측정한다.

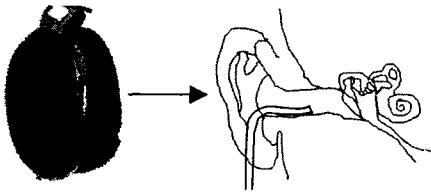


Fig.1 Headphone to ear impulse response measurement. The headphone emit the MLS to the ear canal, and the probe microphone in the ear canal acquire the sequence.

헤드폰은 Sennheiser 사의 HD600 을 사용하고, Etymotic 의 ER-7C 프로브 마이크로폰을 이용하여 귀도 입구로부터 10mm 거리에서 충격 응답을 측정하였다(Fig. 1). 모든 데이터의 획득 및 재현은 Lynx two soundcard 와 MATLAB DAQ Toolbox 를 이용하였다. 헤드폰의 입력 신호와 마이크로폰의 출력 신호로서 헤드폰-귀 충격 응답을 구한다. 실험에 쓰인 가진 신호는 $2^{14}-1$ 개의 MLS (Maximum Length Sequence)이고 충격 응답의 계산법은 MIT[5], Algazi et al.[8]에 의거하였다

2.2 원음 재생 필터 설계

2.1 절에서 측정한 충격 응답을 이용하여 역필터를 설계하는 방법으로, Wiener filter [7]를 이용한다. 원음 재생 필터 $h(t)$ 와 획득한 필터 $g(t)$ (헤드폰-귀 충격 응답)를 컨볼루션 연산을 한 결과값 $y(t)$ 와 지연된 충격함수 $\delta(t-\tau)$ 의 에러가 최소화 되도록 하는 최적화된 $h(t)$ 가 Wiener filter 이다. 원음 재현 필터 $h(t)$ 를 설계하는데 안정성과 재현성을 위하여 강건성 계수 ($\beta = 10^{-4}$) [7]를 고려하였다.

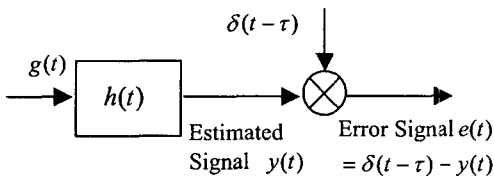


Fig 2. Inverse filter design using Wiener filter theory.

Fig. 3 에서 헤드폰-귀 충격 응답을 $g(t)$ 라고

하면, 특정 음원 $x(t)$ 를 디지털 필터 $g^{-1}(t)$ 를 통과시켜 재현하면 결과값은 결국 지연된 입력 함수와 같은 형태 $x(t-\tau)$ 를 취할 것이다. 그러므로 원음 재생을 위해선 역필터를 통과한 신호를 발생시켜야 한다.

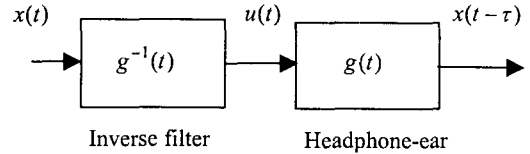


Fig. 3 Inverse filtering in headphone reproduction

이러한 역할을 하는 역필터는 원음 $x(t)$ 를 재생시켜주므로 앞으로는 디지털 역필터 대신 원음 재생필터라 호칭한다

2.3 객관적 테스트

객관적 테스트는 기존의 사람의 판단에 의한 주관적 테스트에 덧붙여, 원음 재생 필터를 이용한 헤드폰 재현 결과의 정확성을 나타낼 수 있는 테스트이다. Fig. 4 는 MLS 입력을 이용해 얻어진 전기-음향계(헤드폰-귀도-마이크)의 충격응답과 주파수 응답이다. Fig. 5 는 강건계수 $\beta = 10^{-4}$ 을 고려한 원음재생 필터응답을 보여주고 있다. 원음 재생 필터는 시간영역의 특성을 고려하여 약 256 샘플의 지연을 갖도록 설계하였다. 플랜트응답과 필터응답이 공히 1024 개의 길이를 갖는 FIR 필터로 잘 모델링됨을 알 수 있다. 플랜트 응답인 Fig. 6 에서 알 수 있듯이 고가 헤드폰 임에도 불구하고 주파수응답이 고르지 못하여 원음재생을 위해선 역필터 설계가 필수적임을 알 수 있다.

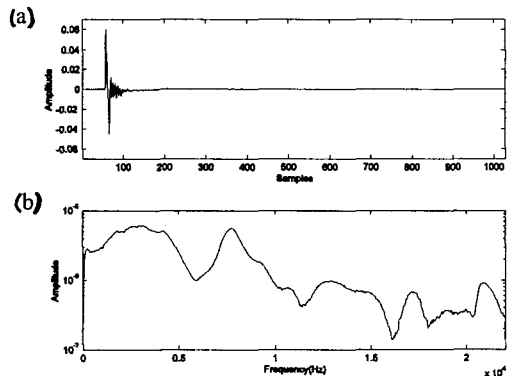


Fig.4 Headphone-ear canal response in a) time domain and b) frequency domain.

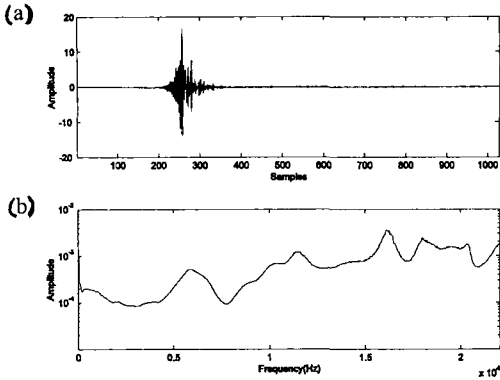


Fig. 5 the corresponding inverse filter in a) time domain and b) frequency domain.

Fig.6 은 원음재생필터를 통과한 랜덤신호의 귀도에서 측정된 음압의 파워 스펙트럼이다. 전영역에서 평탄한 특성을 보임을 알 수 있다.

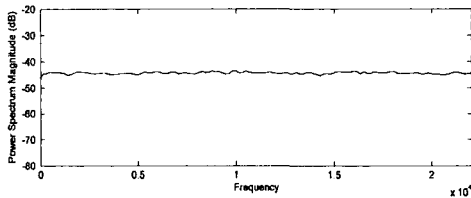


Fig. 6 Power spectrum of the output with the inverse filter

다음 기술할 헤드폰에서의 가상음향 재현의 주관적 및 객관적 평가는 이러한 원음재생필터를 고려한 경우와 고려치 않은 경우 두가지에 대해서 실행한다. 실험 장비 세팅은 헤드폰-귀 충격 응답 측정 시와 마찬가지로이며, 다른 점은 헤드폰 입력 신호로서 MLS 가 아닌 랜덤 신호를 쓴다는 점이다. 객관적 평가를 위해 모델링/역필터 설계/재생 시 각각의 오차를 정의하였다. 시간영역에서의 오차 및 전체 평균 오차는 다음 식에 의해 정의된다[7].

$$e(t) = d(t) - x(t - \tau) \quad (1)$$

$$J(\%) = \sum_{i=1}^2 \frac{E[e^2(t)]}{E[x^2(t - \tau)]} \times 100 \quad (2)$$

출력 신호 $d(t)$ 와 지연된 입력신호 $x(t - \tau)$ 의 차이를 식(1) 의 $e(t)$, $E[\bullet]$ 를 \bullet 의 기대값이라고 하면, 재현 에러 $J(\%)$ 는 (2)와 같이 정의한다.

식(1)의 오차는 측정된 출력신호 $d(t)$ 와 입력신호 $x(t)$ 를 비교하여 알 수 있는데, 이때 시간지연 τ 는 상호상관함수 (Crosscorrelation)를 이용해 구할 수 있다. 입력과 출력 신호의 상호상관 함수의 한 예를 Fig. 7 에 보였다. 입력 신호가 랜덤신호이기 때문에 상호상관함수가 단지 한 지점에서 현저하게 큰 값을 갖게 되는데, 그 값이 존재하는 샘플이 (즉 Fig. 4 에서 샘플 9 에 해당하는) 바로 출력의 지연 샘플 개수이다.

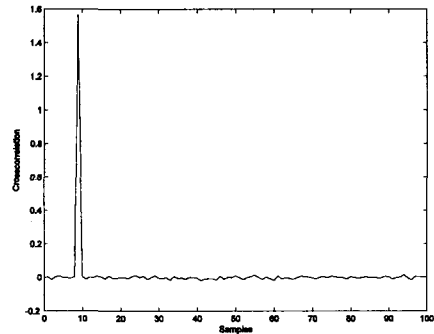


Fig. 7 Crosscorrelation between the input and output signals : The function has its maximum at the sample delay 9 which means the input and output signals are best correlated when the input is delayed by 9 samples.

Table 1 에서 각각의 시험자에 대한 모델링/역필터 설계/재생 오차를 비교하였다. 오차계산은 2 채널에 대한 일반식인 식(1)을 1 채널경우로 변형해(단순히 Sigma 를 없앴) 계산하였다. 표에서 보이다시피, 원음재생 필터 사용시 피 시험자 5 명 평균 약 2.4%의 재생에러를 보인 반면, 원음 재생 필터가 없이 약 80.64%의 에러를 나타내었다. 재생오차는 모델링 및 역필터 설계오차 전부를 포함하므로 시스템 전체의 정확도를 나타내는 값이다.

Table 1 Modeling, inverse, and reproduction error (%)

Error(%)	modeling	inverse	reproduction
S	0.20	0.33	1.95
M	0.18	0.22	3.34
G	0.18	0.44	1.09
W	0.20	0.15	1.75
D	0.19	0.19	3.85
average	0.19	0.26	2.40

다음 실험에서 가상음원을 일정각도로 돌리면서 재생오차를 시험하였다. Fig. 8 은 측정좌표계를 보이고 있으며, 측정은 인간의 머리의 중심을 지나는 수평면상의 음원 재현만을 고려한다. 머리

중심을 중심으로 피 실험자의 정면을 0° 로 하고 오른쪽 귀 방향이 양의 방향이 된다. 예를 들면 오른쪽 귀 방향이 약 90° , 왼쪽 귀는 약 270° 가 된다. 방향감을 주기 위한 HRTF는 MIT HRTF database [5]를 사용하였고, 0° 부터 15° 간격의 24개의 수평면상의 충격 응답을 사용하였다. 테스트는 각각의 각도에 대한 HRTF와 원음재생필터를 컨벌루션시킨 랜덤 신호를 헤드폰을 통해 약 70dB SPL로 가진하였다. 각각의 랜덤 신호는 0.25ms의 지속시간을 가지고, 신호의 시작 부분과 끝부분 각각 0.02ms에는 서서히 커지도록 또는 작아지도록 램프를 주었다. 그리고 200Hz ~ 14kHz의 bandpass 필터를 통과 시켰다.

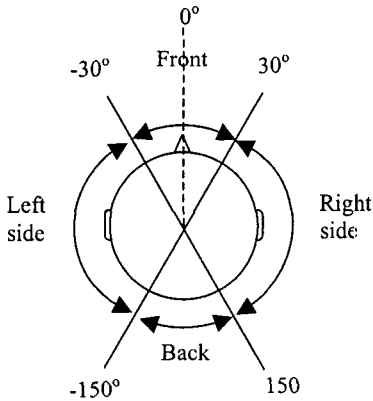


Fig. 8 Area division in the horizontal plane as used in Wightman and Kistler [1]

Fig. 9의 실선은 피 실험자 W의 오른쪽 귀에 서만의 재생에러와 HRTF의 음압의 에너지를 보여준다. 에러는 최소 1.29%에서 최대 46.23%의 분포를 가지며 평균 12.20%의 에러를 가진다. 특이할 점은 270° 에서 비교적 낮은 에러를 나타내고 전후 20° , 즉 250° , 290° 에서의 에러가 가장 높게 나타난다는 점이다. 이러한 결과는 각 각도의 HRTF의 에너지의 크기에 거의 정확히 반 비례한다(Fig. 9 점선). 즉 HRTF가 250° , 290° 에서 가장 적은 음압을 가지고 있고 상대적으로 270° 에서는 높은 음압을 갖는다. 결과적으로 각각의 각도에 의한 재생 에러는 신호 대 잡음 비가 작아짐에 따른 에러라 볼 수 있다.

Fig.10은 피 실험자 W의 각각의 각도에서의 양쪽 귀에 대한 재생 에러를 보여 준다. 원음재생필터를 이용한 재생에러를 측정된 결과 6.24%에서 최대 47.73%까지 평균 24.16%의 에러를 나타내었고, 이에 비하여 필터 없이 재생을 한 경우는 평균 113.22%로서 상대적인 에러가 매우

크게 나타남을 알 수가 있다. 원음재생필터를 고려치 않은 경우는 재생오차가 너무 크므로 자세한 오차해석을 생략하였다.

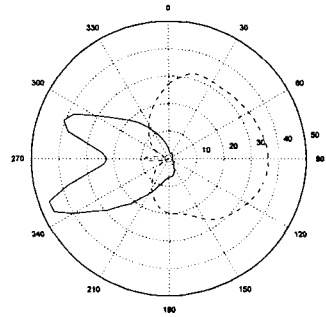


Fig.9 Reproduction error at the right ear only(solid lines) and the HRTF signal energy (dB) at every 5 degrees (dashed lines)

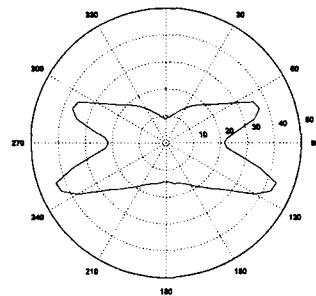


Fig.10 Reproduction error considering both ears at every 5 degrees

2.4 주관적 테스트

다음은 피 실험자의 판단에 의존하는 주관적 평가 결과를 기술한다. 시험은 앞의 객관적 평가와 마찬가지로 두 가지 경우에 대해 실행하였다. 첫째는 원음재생 필터를 적용한 경우와, 둘째는 적용시키지 않은 경우이다. 시험신호는 랜덤신호를 이용하였으며, 이를 개인의 원음 재생필터와 컨벌루션시키고 각 각도의 HRTF와 재 컨벌루션하여 헤드폰을 통하여 출력시켰다

재생 각도는 0도에서 시작하여 한바퀴를 돌아 360도에서 끝나도록 했으며, 매 재생 각도는 15도간격으로 정했다(Fig. 8 참조). 시험자는 피 실험자에게 매 재생각도를 미리 구두로 주지시키고, 피 실험자는 고정된 각도에서 음상의 위치를 표시하도록 지시하였다. 실제로 사람이 실제 공간에서 음의 위치를 파악할 때 정확한 각도를 읽어 낼 수가 없다. 따라서 본 논문에서는 주관적 테스트시, 음상의 위치를 판단하는 것보다는 머리의 안과 밖

의 위치를 판단하도록 피 시험자에게 요구하였다. 각각의 각도를 알려준 상태에서 단지 음상의 안, 밖의 상대위치만을 고려하도록 하고, 앞, 뒤 음상 혼동이 되는 각도는 'o'로 표시했다. Fig.10 은 각 피 시험자의 주관적 음상궤적을 그린 것이다. 가운데 점선 원은 사람의 머리를 나타내고 실선이 음상이 맺힌 위치이다. 왼쪽 행은 원음재생필터를 고려했을 때의 결과이고 오른쪽 행은 고려하지 않고 재생했을 때의 결과이다. S 시험자의 원음재생필터를 고려했을 때의 결과인 왼쪽궤적을 보면 정면에서 앞뒤혼동 및 머리 내 음상정위를 보이고 옆과 뒤에서는 음상이 대체적으로 머리밖에 위치한 결과를 보이는 반면, 원음재생필터를 고려치 않은 경우는 음상이 측면을 제외한 전 방향에서 머리 내 음상정위를 경험했다. 또한 정면과 같이 음상이 머리 내에 위치한 경우(음상의 반경이 적다) 음상파악에 많은 어려움을 호소했다. 다른 피 시험자의 경우 원음재생필터 고려유무에 무관하게 정방과 후방음상의 경우 머리 안 음상정위를 경험했다. 좌우측면의 경우, 뚜렷한 머리 밖 음상정위를 보여주고 있다. 모든 실험 데이터에서, 원음재현 필터를 사용할 때는 단 한번(0.4%)의 혼동이 일어났고, 사용하지 않을 때 21 번(8%) 일어났다. Jean-marie Pernaux et.al.[6]에서도 원음재현 필터가 없을 때 혼동이 증가하는 것으로 보고하고 있다. 특히 각각의 각도를 알려주고 실험을 했음에도 불구하고 개개인 마다 차이가 있지만 약 $45^{\circ} \sim 75^{\circ}$ 와 $-45^{\circ} \sim -75^{\circ}$ 에서의 음상 위치가 양쪽 귀 쪽으로 이동됨을 언급하였다.

Fig. 12는 각 경우의 평균을 보인 것이다. 전반적으로는 원음재생필터를 고려한 것이 머리 밖 음상정위가 많고, 또한 이 경우 음상의 반경이 커져 앞뒤혼동이 크게 줄음을 볼 수 있다. 특히 측면에서의 음원 재현은 모든 피 시험자들이 머리 밖에 있다고 판단하였다.

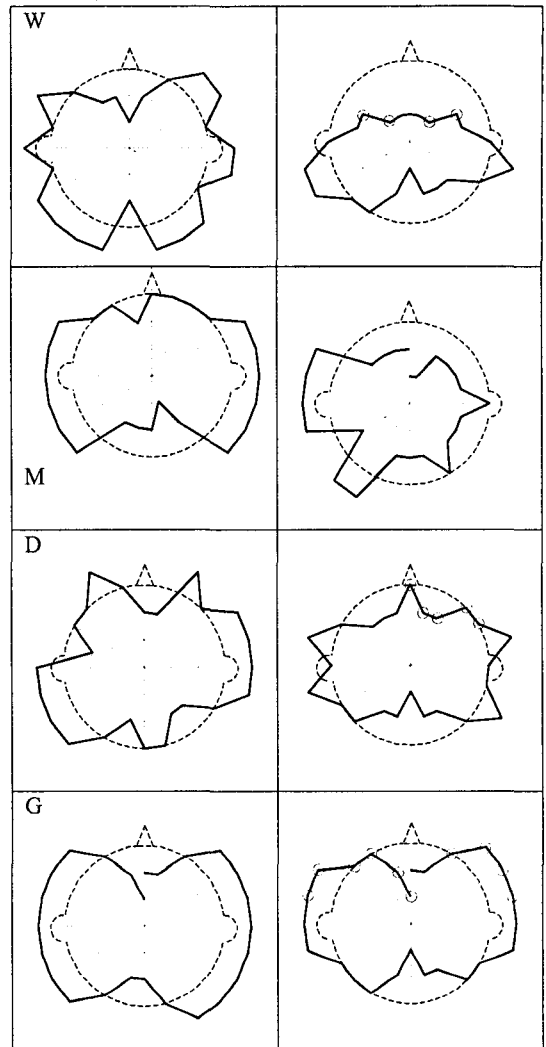
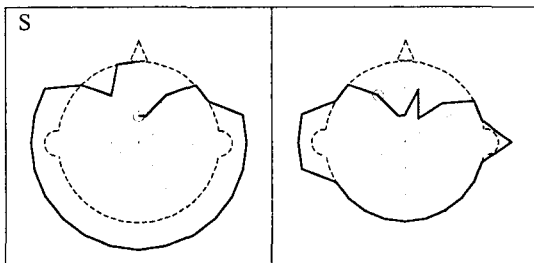


Fig 11 Subjective sound image trajectory with (left) and without (right) using the inverse filter. The capital alphabet in each row is subject's initial. The circles 'o' denote front-back confused positions.

Wightman et al.[1]은 HRTF 와 원음재현필터 모두를 피 실험자 자신의 것을 사용하였지만, 본 논문에서 보였듯이 HRTF 는 MIT HRTF database 를 사용하였다. 만약 본인의 HRTF 를 사용한다면 전/후방에서의 앞뒤혼동 및 머리 안 음상정위문제도 크게 향상 시킬 수 있을 것으로 사료된다. 시험결과는 가상음향의 효과적인 헤드폰 구현을 위해선 원음 재현 필터를 사용해야 함을 보여주고 있다.

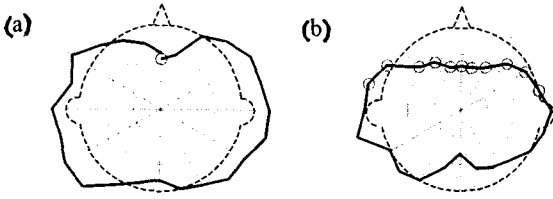


Fig 12 Averaged subjective sound image trajectory from the five subjects with (left) and without (right) using the inverse filter. The circles 'o' denote front-back confused positions.

3. 결론

헤드폰을 이용한 입체음향 재현에서 원음재생필터의 필요성을 입증 하였고, 정확한 필터의 설계를 위하여 필터의 정확성을 판단할 수 있는 객관적 테스트를 제시 하였다. 5 명의 테스트 결과, 비록 자신의 HRTF 를 사용하지 않더라도 테스트를 통해 정확한 원음 재생 필터를 사용하면 머리 밖음상 정위는 물론 앞,뒤 혼동문제도 탁월하게 개선됨을 확인할 수 있었다. 따라서 헤드폰 재현에서 원음 재생 필터가 꼭 필요함 보였다. 나아가 상업적 적용을 위해선 개개인의 HRTF 의 통계적 분석을 통한 일반화된 HRTF 에 대한 연구가 필요하다고 사료된다.

후 기

본 연구는 한국과학재단 목적기초연구(R05-2003-000-11999-0) 지원으로 수행되었음.

참고문헌

- (1) F.L.Wightman and D.J.Kistler, "Headphone simulation of free field listening II: Psychophysical validation," *J.Acoust.Soc.Am.*, vol.85, pp.868-878, (1989).
- (2) H.Moller, "Fundamentals of binaural technology," *Applied Acoustics*, vol.36, pp.171-218, (1992).
- (3) R.L.Martin, K.I.McAnally, and M.A.Senoia, "Free-field equivalent localization of virtual audio," *J.Audio Eng. Soc.*, vol.49, pp.14-22, (2001).
- (4) D.D.Rife and J. Vanderkooy, "transfer-function measurement using Maximum-Length Sequences". *J.Audio Eng.Soc.*, 37(6):419-444, June 1989
- (5) MIT Media Lab. "HRTF measurements of a KEMAR dummy-head microphone", May, 1994
- (6) Jean-marie Pernaux, Marc Emerit, Jerome Daniel & Rozenn Nicol, "Perceptual evaluation of static binaural sound synthesis" AES 22nd International conference, pp. 169 -178
- (7) Sang-Myeong Kim and Semyung Wang, "A Wiener Filter Approach to the Binaural Reproduction of Stereo Sound", Technical Paper No.1, Active Acoustic Lab. KJIST, January 2003
- (8) V.R.Algazi, R.O.Duda, D.M.Thompson and C.Avendano, "The CIPIC HRTF database". IEEE Workshop on Applications of Signal Processing to Audio and Acoustics 2001. pp. W2001-1 -W2002-4
- (9) Hartmann et al. "On the externalization of auditory images", *J.Acoust.Soc.Am.* 99(6), June 1996