

HMD를 이용한 오디오 재생 기술에서 Latency의 영향 분석

Effect on Audio Play Latency for Real-Time HMD-Based Headphone Listening

손상모† · 조현* · 김선민**

Sangmo Son, Hyun Jo, Sunmin Kim

Key Words : Head-Related Transfer Function(머리전달함수), Head Mounted Display(HMD, 헤드마운트 디스플레이), Head-Tracking(머리 움직임 정보)

ABSTRACT

A minimally appropriate time delay of audio data processing is investigated for rendering virtual sound source direction in real-time head-tracking environment under headphone listening. Less than 3.7 degree of angular mismatch should be maintained in order to keep desired sound source directions in virtually fixed while listeners are rotating their head in a horizontal plane. The angular mismatch is proportional to speed of head rotation and data processing delay. For 20 degree/s head rotation, which is a relatively slow head-movement case, less than total of 63ms data processing delay should be considered.

1. 서 론

심리음향(Psychoacoustics) 분야에서는 사람이 음원(Sound Source)의 3차원 위치정보를 인지함에 있어서 어떠한 인지요소(Sound Localization Cue)를 사용하는지, 그리고 그 인지한계(Threshold)는 어디까지 인지에 대한 많은 실험과 연구가 진행되어 왔다. 그 중에서 사람이 음원의 수평각도를 인지하기 위한 인지요소로는 사람의 양이간 시간차(Interaural Time Difference, ITD)와 양이간 레벨차(Interaural Level Difference, ILD)가 주파수 별로 서로 다른 역할을 하며 사용된다고 밝혀졌다⁽¹⁾. 또한, 음원의 고도각도 인지와 관련된 인지요소로는 Monoaural Spectral Cue라고

불리는 인지요소가 사용됨이 밝혀졌다⁽²⁾.

물리적인 스피커를 무향실에 두고 사람의 음원 방향인지 분해능에 대한 실험도 이루어졌다. 이 분해능은 Minimum Audible Angle(MAA)로 불리며, 특히 수평면 상에서의 분해능(약 1도)이 수직면(약 4도) 보다 매우 민감함이 밝혀졌다⁽³⁾. 물리적인 스피커가 움직일 때 사람이 얼마나 작은 움직임까지 인지할 수 있는지에 대한 인지한계도 실험되었다. Minimum Audible Movement Angle(MAMA)로 불리는 이 인지한계의 경우, 스피커가 수평면에서 10deg/s로 움직일 때 약 8deg 정도의 음원 움직임(MAMA)이 있어야 사람이 그 음원의 움직임을 인지함이 밝혀졌다⁽⁴⁾. 또한 음원의 속도가 증가할수록 MAMA도 커진다는 것이 밝혀졌다⁽⁴⁾.

디지털 신호처리 기술이 발전함에 따라 현실공간의 스피커가 방사하는 소리를 헤드폰이나 이어폰 청취 환경에서 똑같이 재현하려는 시도가 이루어졌다. 머리전달함수(Head-Related Transfer Function)를 측정하면 도입부에서 설명한 ITD, ILD, Spectral Cue 정보가 추출되며 측정된 머리

† 교신저자; 삼성전자
E-mail : sm21c.son@samsung.com
Tel : 010-6298-2638

* 제 2저자; 삼성전자

** 제 3저자; 삼성전자

전달함수를 원음에 Convolution 시킴으로 인해서 가상 음향공간 체험이 가능하다^(2,5). 이러한 Spatial Audio Rendering 기술은 최근, Head Mounted Display의 등장으로 Virtual Reality 혹은 Augmented Reality 시장에서의 그 필요성 또한 대두되고 있다.

관련된 대표적인 Spatial Audio Rendering 시나리오로는 Head Mounted Display(HMD) 등의 환경에서 사용자의 머리움직임 정보를 Head-Tracking Device 등으로부터 받고, 머리움직임에 연동하여 실시간으로 Rendering 위치정보를 바꿔주는 기능이다. 이를 통해 사용자 입장에서는 머리가 움직이더라도 음상의 위치가 바뀌지 않게 느껴지는(실제로는 실시간으로 Rendering 위치가 바뀌어도 불구하고), 마치 실제 공간에서 소리를 듣는 것 같은, 가상 음향공간 체험이 가능하다.

문제는 앞서 설명한 오디오 시나리오를 실제 구현 시에 Head-Tracker로부터 전해오는 Head-Tracking 딜레이와 머리전달함수 Convolution 등으로 인한 데이터 연산 딜레이로 인해 실제 물리적인 환경과는 다르게 수십ms 이상 딜레이가 발생하게 되고 이는 3차원 공간감에 이질적인 영향을 줄 수 있다는 점이다. 하여, 본 논문에서는 헤드폰/이어폰 가상청취환경에서 사람의 머리 움직임 속도(회전각도)와 프로세싱 딜레이와의 상관관계를 주관평가(Listening Test)의 관점에서 살펴보고, HMD 청취환경에서 3차원 음상 공간감을 유지하기 위한 최소한의 Delay가 몇 ms 이하로 유지되어야 하는지 그 한계를 분석하고자 한다. 또한 물리적인 스피커를 활용한 기존의 MAA 결과와 비교하여 본 논문에서 도출한 가상 음향공간에서의 결과가 기존의 결과와 어떠한 상관관계가 있는지 살펴보고자 한다.

2. 실험 구성 및 실험 결과

2.1 실험 구성

(1) 실험 장비 구성.

소리의 딜레이와 머리 움직임 속도 간의 상관관계를 주관평가로 확인하기 위해 아래와 같은 실험설계를 하였다. 먼저, 머리 회전각도 정보를 받기 위

해 헤드트랙커는 Polhemus 사의 FASTRAK SET를 활용하였다. 테스트 음원 청취를 위해서는 젠하이저 HD-650 헤드폰을 사용하였다. 그리고 실험을 진행할 Note-PC(삼성 NT-R520)로 헤드트랙커 장비를 구동하였으며 또한 같은 PC로 테스트 음원 송출하였고, S/W로 portaudio V.19를 활용하였다⁽⁶⁾. 실험자 정면에는 대형 스크린을 설치하였고, 사용자에게 머리움직임 속도 가이드를 주기 위한 동영상을 재생해 주었다. 동영상은 테스트할 속도에 맞게 붉은 공이 좌/우로 움직이도록 하여 사용자는 공의 속도에 맞게 머리를 좌/우로 움직이도록 하였다. 공의 움직임은 사용자의 정면(0도)을 기준으로 -30도 부터 +30도 까지 움직이도록 하였다. 전체 실험은 ITU-R BS.1116 규격 시청실에서 진행 하였다.

(2) 청취 평가 설계.

주관 청취 평가에 대한 이해가 있는 20대 후반부터 30대 중반까지의 총 4인으로 구성된 피 실험자가 실험에 참가하였다. 사용자의 실제 평상시 머리움직임 속도를 알고자 자연스럽게 머리를 좌우로 돌려보라고 한 뒤 10명에 대한 머리움직임 속도를 측정된 결과, 최저움직임 속도는 9.8deg/s 였고 최대는 76.7deg/s 였 으며, 평균 32.2deg/s의 속도를 보였다. 이러한 정보를 바탕으로 실험에 사용할 머리움직임 속도는 20deg/s, 40deg/s, 60deg/s 총 세 가지로 설정하였다.

음원의 재생 딜레이 값은 0, 20, 40ms 총 세 가지로 설정하였다. 따라서 피 실험자 1명당 총 실험 회수는 “머리회전속도 3가지(20, 40, 60 deg/s) X 음원 딜레이 3가지(0, 20, 40ms) X 10(회 반복)”이며, 따라서 총 90회의 테스트를 실시하였다. 90회 테스트를 위해서 약 50분이 소요되었다.

실험 순서는 아래와 같이 진행하였다. 시작으로는 정면을 바라보며 정면(0도)에 대한 음상 렌더링 결과를 헤드폰으로 재생한다. 음원이 재생이 되면 정면 동영상의 붉은 공의 속도에 따라 얼굴을 좌/우로 돌린다. 실험에서 머리를 움직이지 않고 눈만 돌려 붉은 공을 쫓아가는 경우를 방지하기 위해 실험 전 사용자에게 눈을 움직이지 않도록 가이드를 주었다.

20초 음원에 대한 1회 테스트가 끝나면, 실험 음

원과 음원 사이를 구분 해줄 Tonal Signal을 5초 동안 틀어주어 1회가 끝났음을 피 실험자에게 전달하였다. 피 실험자는 머리를 회전하는 동안 헤드폰에서 나오는 소리가 Stationary Sound Source처럼 들리면 O로 표기, Moving Sound Source처럼 들리면 X로 표기 하도록 하였다. 한 회의 실험이 끝난 후에는 다시 정면을 응시토록하고 피 실험자는 같은 테스트를 90회까지 반복하였다. 실험에 사용한 음원은 20초 길이의 백색잡음(Broadband white noise, 0~24kHz까지)을 사용하였다.

(3) 청취 평가용 알고리즘.

머리움직임에 따른 음원 위치 보상을 위해 CIPIC HRTF DataBase⁽⁷⁾의 azimuth DATA 중에 -30도 ~ +30도 까지를 이용하였다. HRTF의 간격은 5도 간격이었고, 실험에서 머리움직임 각도는 5도 간격으로 Truncation하였다. 테스트용 음원인 백색잡음을 HRTF 각도 별로 미리 필터링해두었다. 헤드트래킹 장비를 이용하여 실시간으로 머리 움직임의 각도 정보(x deg)를 받아 0도와 대칭되는 각도의 HRTF(-x deg 위치)로 필터링 된 음원을 매 2048 데이터 프레임 별로 실시간 재생해주었다. 머리움직임 각도를 추정 할 때부터 소리가 재생될 때 까지 약 43ms의 하드웨어적인 딜레이가 발생하였다. 실험에 사용한 샘플링 주파수는 48kHz다.

2.2 실험 결과

피 실험자는 머리를 회전하는 동안 헤드폰에서 나오는 소리가 Stationary Sound Source처럼 들리면 O로 표기, Moving Sound Source처럼 들리면 X로 표기 하도록 하였다. 그 결과를 전체 실험자를 대상으로 보게 되면, 아래 Table 1과 같다. 피 실험자들은 머리회전속도가 빨라질수록 Delay 증가량에 따른 부자연스러움을 좀 더 명확하게 느꼈다.

Table 1. Results from all four subjects are summarized. Percents represent that perceived direction stay at the front of subject (0 degree) in given test conditions.

	20deg/s	40deg/s	60deg/s
0ms	82.5%	60.0%	12.5%
20ms	75.0%	40.0%	22.5%
40ms	52.5%	22.5%	10.0%

개인별 편차가 있지만 전체적인 트렌드는 위의 결과와 동일하였다. 데이터 세로축 값은 43ms 딜레이를 포함하지 않은 수치이다.

실험 결과를 피 실험자 별로 살펴보면 다음과 같다. 피 실험자 1번의 경우 머리 회전속도가 40deg/s, Delay가 40ms 일 때 매우 부자연스러움을 느끼기 시작했으며, 머리회전속도가 60deg/s 일 때는 모든 delay에 대해서 부자연스러움을 느꼈다. 결과는 Figure 1과 같다.

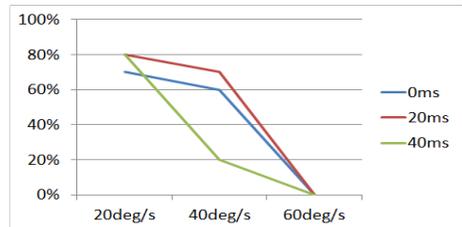


Figure 1. Results from subject1 is plotted with respect to the test conditions.

피 실험자 2번의 경우 머리회전속도 20deg/s, delay 40ms 일 때부터 매우 부자연스러움을 느꼈으며, 머리회전속도 40deg/s, delay 20ms일 때부터 급격하게 부자연스러움이 증가하였고, 머리회전속도 60deg/s 일 때에는 모든 delay에 대해 부자연스러움을 느꼈다고 답하였다. 결과는 Figure 2와 같다.

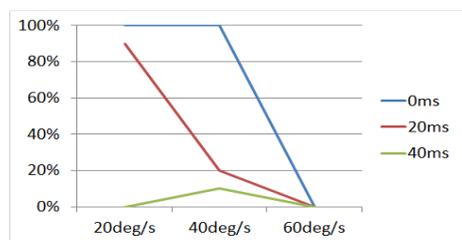


Figure 2. Results from subject2 is plotted with respect to the test conditions.

피 실험자 3번의 경우 delay 매우 민감한 반응을 보였으며 머리회전속도 20deg/s, delay 40ms부터 나머지 경우에 대해 모두 부자연스럽다고 답하였다. 결과는 Figure 3과 같다.

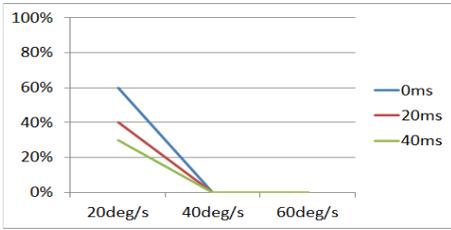


Figure 3. Results from subject3 is plotted with respect to the test conditions.

피 실험자 4번의 경우, 상대적으로 다른 피 실험자들에 비해 딜레이에 둔감한 양상을 보인다. 머리 회전속도 60deg/s, delay 20ms에서도 부자연스러움을 다른 실험자들에 비해 훨씬 덜 느꼈다. 결과는 Figure 4와 같다.

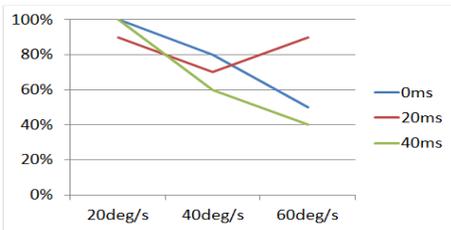


Figure 4. Results from subject4 is plotted with respect to the test conditions.

3. 고찰

본 장에서는 실험결과에 대한 물리적인 의미에 대해 고찰해보고자 한다. 실제로 피 실험자가 실험에서 느끼는 음원의 움직임은 음상의 위치 표현이 머리움직임 속도에 따라 즉각적으로 반응하지 못하기 때문이다. 예를 들어, 실험에서 피 실험자가 머리를 좌로 움직일 때는 음원의 딜레이로 인해 음원이 0도에 있다고 듣게 되는 것이 아니라 0도보다 x 각도 만큼 살짝 왼쪽에 위치하게 된다. 만일 이때 피 실험자가 머리를 정지하게 되면 딜레이 시간 이후에는 음상이 0도로 이동하게 된다. 반대로 머리를 오른쪽으로 움직일 경우에는 0도 보다 우측에 음상이 위치하게 되며, 실험에서처럼 머리를 좌/우로 움직일 경우에는 음상이 좌로 x만큼 그리고 우로 x만큼 머리움직임 방향에 따라 변화하게 된다. 이를 Minimum Audible Latency Angle(MALA)라 정의

하고 도식적으로 살펴보면 Figure 5.와 같다.

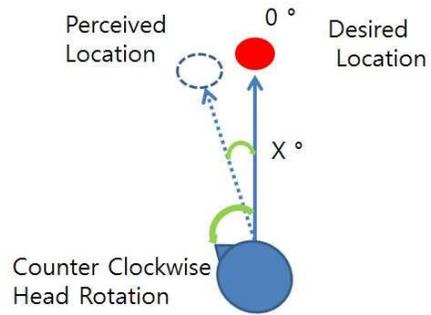


Figure 5. Conceptual drawing of perceived location mis-match when listener's head is rotated in counter-clock wise.

따라서 만일 x만큼의 각도 변화가 인지할 만큼의 각도변화라면 피 실험자는 음상의 위치가 부자연스럽게 움직이고 있다고 판단하게 되고, 만일 x만큼의 각도 변화가 인지할만한 수준이 아니라면 피 실험자는 실험도중에 계속 0도 위치에 음원이 정지하고 있다고 판단하여 부자연스러움을 느끼지 못한다.

Table 2.에서는 실험에서 사용한 딜레이와 머리움직임 속도로부터 음원의 MALA를 각도로 환산한 값이다. MALA 란, 딜레이와 머리움직임 속도의 곱을 의미한다. 여기서 각도로 환산 할 때 시스템 딜레이 43ms를 포함 하였다.

Table 2. Angle of tone balance with different test conditions.

	20deg/s	40deg/s	60deg/s
43ms	0.8	1.7	2.5
63ms	1.2	2.5	3.7
83ms	1.6	3.3	4.9

Table 1.의 결과에 70.7% 한계치를 적용해보면⁽⁸⁾, 전체적으로 약 1.2도 이하의 MALA 각도 이하의 경우에만 실험자들이 MALA를 인지하지 않음을 확인할 수 있다. 한편 실험에서 사용한 HRTF의 각도가 5도 간격이었던 점을 감안 한다면, 피 실험자들이 MALA를 느끼는 실제 각도는 다음과 같다. CIPIC DATA의 HRTF 간격이 5도 간격으로 Truncation되었기 때문에, 각도의 오차는 2.5도 와 실험결과 1.2도를 더한 3.7도 이상에서 발생함을 유

추할 수 있다. 즉, MAA의 관점에서 결과를 재해석해 보자면, 프로세싱 딜레이로 인해 발생한 MALA 각도가 MAA보다 작게 발생할 경우에, 사람들은 실제 스피커의 위치와 딜레이로 인해 변경된 위치와의 차이를 인지하지 못한다고 해석할 수 있다. Perrot 등이 밝힌 수평면에서의 MAA각도가 약 1도였고, Chandler가 밝힌 수평면에서의 스피커의 움직임을 인지하는 MAMA각도는 8도였다. 본 실험에서 검토한 가상 음향 공간에서의 수평면 Minimum Audible Latency Angle(MALA)이 3.7도이기 때문에 약간의 차이를 보인다. 본 논문의 가상 공간의 실험결과가 실제 사람이 물리공간에서 느끼는 MAA(혹은 MAMA) 실험결과 간의 상관관계에 대한 고찰을 향후 과제로 남긴다.

4. 결 론

본 논문에서는 헤드트래킹을 활용한 헤드폰/이러 폰 청취환경에서 사람의 머리 움직임 속도(회전각도)와 프로세싱 딜레이와의 상관관계를 주관평가(Listening Test)의 관점에서 살펴보고, 3차원 음향 공간감을 유지하기 위한 최소한의 프로세싱 Delay가 몇 ms 이하로 유지되어야 하는지 그 한계를 분석하였다. 실시간으로 수평면 음원의 각도를 표현함에 있어서 머리움직임 속도에 따른 각도 표현 오차가 최대 3.7도 이하로 유지되어야 음상의 자연스러운 위치표현이 가능하다. 이는 5도 간격의 HRTF를 사용하는 구현 환경에서, 머리움직임 속도가 20deg/s일 때 전체 소프트웨어와 하드웨어 프로세싱 Delay가 63ms 이하로 유지되는 수준이다.

참 고 문 헌

- (1) Rayleigh, L., 1907, On our perception of sound direction, *Philosophical Mag.* Vol. 13, pp. 214~232.
- (2) Begault, D. R. 1994, 3-D Sound for Virtual Reality and Multimedia, Academic Press Professional, Cambridge, MA, USA.
- (3) Perrott, D. R. and Saberi, K., 1990, Minimum audible angle thresholds for sources varying in both elevation and azimuth, *J. Acoust. Soc. Am.* Vol. 87, pp. 1728~1731.
- (4) Chandler, D. W. and Grantham, D. W., 1992, Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity, *J. Acoust. soc. Am.* Vol. 91, pp. 1624~1636.
- (5) Jo, H., Park, Y. and Park, Y.-S., 2010, A subjective listening test of specific subject's customized head-related transfer functions on various subjects, *Proc. of Spring Conf. of the Korean Society for Noise and Vibration Engineering.* pp.776~777.
- (6) Portaudio Install File
http://pkgs.fedoraproject.org/repo/pkgs/portaudio/pa_stable_v19_20071207.tar.gz/d2943e4469834b25afe62cc51adc025f/
 ASIO SDK 2.3 file,
<http://www.steinberg.net/en/company/developers.html>
- (7) CIPIC HRTF database files, CIPIC Interface Laboratory, U. C. Davis.
<http://interface.cipic.ucdavis.edu/>, released 1.1, August 21, (2001).
- (8) Levitt, H., 1971, Transformed up-down methods in psychoacoustics, *J. Acoust. Soc. Am.* Vol. 49, pp. 467~477.