Audio Engineering Society
Convention Paper 7882
Presented at the 127th Convention
2009 October 9–12 New York NY, USA

Reproduction 3D sound by measuring and construction of HRTF with Room reverberation

akira saji[1], Keita Tanno[1], Jie Huang[1]

[1] *The university of Aizu, Fukushima, 965-8580, Japan*

Correspondence should be addressed to akira saji (d8082102@u-aizu.ac.jp)

ABSTRACT
In this paper, we proposed a new method using HRTFs that contain room reverberations(R-HRTF). The reverberation is not added to the dry sound source separated with HRTF but contained at their measured process in the HRTFs. We measured the HRTFs in a real reverberant environment for directions of azimuth 0, 45, 90, 135 (left side) and elevation from 0 to 90 (step of 10 degrees) degrees then constructed a 3D sound system with the measured R-HRTF with headphones, examine if the sound reality is improved. As a result, we succeed to create 3D spatial sound system with more reality compared with traditional HRTFs sound system by signal processing.

1. INTRODUCTION

We have aimed to create 3D spatial sound system using HRTF based method, but such kind of the spatial sounds were less reality. Almost people judged the sound as unnatural, artificial and unfamiliar sound, so this judgment causes adverse effect about sound localization.
In this paper, we aim to solve this problem by application of reverberation.

In general, it is necessary to add reverberation effect to sound works for improve the quality. For example, to master CD, sound engineers add the reverberation to the sound that recorded in studio by reverberator devices in the last step. This step is called a reverberation process, and this is most difficult and important on sound creation so special training and the ability are necessary for this work. And then, the quality of this work defines sound engineer's ability.

Other example of good effects by reverberation, reverberations is very important for concert hall, it is

not an exaggeration to say that it decides the popularity of the hall. It is not uncommon that the performance of reverberations is considered hall construction process and adjusted after construction of the halls.
Moreover, we can find the sound reproduction equipments that can adjust reverberation like major digital audio players.

And, to compare 2ch stereo sound with 5.1 or more surround system, we feel more reality in the surround sysytem. This fact shows not only direct sounds come from side or back of listner, but reverberations come from side or back.

From these facts, reverberations always need to advance in quality of any sounds, and this fact seems to be applicable for 3D spatial sound systems.

Then, we will make sure the importance and effect of reverberations for 3D spatial sound using HRTF. Accordingly, to estimate that reverberations improve the reality of 3D spatial sound, and how to add the reverberations effectively.

The ultimate purpose is to realize 3D spatial sound with more reality creation by only signal processing and its reality is comparable to binaural sound system.

In this paper, we explain new way to create 3D sound, that is HRTFs with room reverberation(RHRTF) convolution method. The advantage of this method is reverberations was contained in HRTFs. It will be able to improve spatial sound localization on 3D sound system.

2. HRTF WITH ROOM REVERBERATION

In this paper, we focused the change of sound perception by HRTF included room reverberation(RHRTF) sounds. S. Yano, S. Shimada and H. Hokari stated in their paper that the combination of HRTF and reverberations allows listeners to perceive a distance of sound image [3].

According to past research, we have revealed the relationship between the length of reverberation and the distance perception by auditory experiments, and also showed that if the length of RHRTF is about 2000 points (with the sampling rate 44100 Hz,) listeners could perceived nearly identical distance between a loudspeaker and a HATS.

Room	ordinary echoic room
Audio Interface	EDIROL UA-1000 & UA-4FX
Amplifier	DENON UAVC-300
Loudspeaker	BOSE CS-6J satellite loudspeaker
Dummy Head	B&K Head and Torso Simulator Type 4128C
Ear Model	B&K Ear Simulator Type 4158 and 4159
Microphone	ACO condenser microphone × 2
Turntable Controller	B&K Type 5949

Table 1: Experimental Equipments of the RHRIR Measurement

From those results, RHRTF was expected to make the reduction of in-the-head localization and the improvement of sound localization, and the development of frontal localization is also provided.

And, the sound that added reverberation simply in reproducing process(HRTF+Reverb) was prepared for the comparison. This is for checking if the reverberations that added separately to the sound creation are effective enough to improve quality of 3D spatial sound with HRTF. The setting of the parameters of the reverberation is too difficult for us, only a professional engineer can adjust it accurately, therefore a so-called "room reverb" was used to match it to the following RHRTF measurement environments.

3. MEASUREMENT OF ROOM AND HEAD-RELATED TRANSFER FUNCTION

3.1. Experimental Equipments

The devices used in the measurement are shown in the Table 3.1 and the layout of the measurement room is described in the Figure 3.1.

In past reserch, N. Saito measured the room reverberation in the lecture room (high reverbration), however, listeners perceived sound image ambiguously because long reverb was an obstacle to localize sound images correctly [5]. Therefore, we selected the follow mentioned acoustical measurement room as a measurement room, where the resonant time

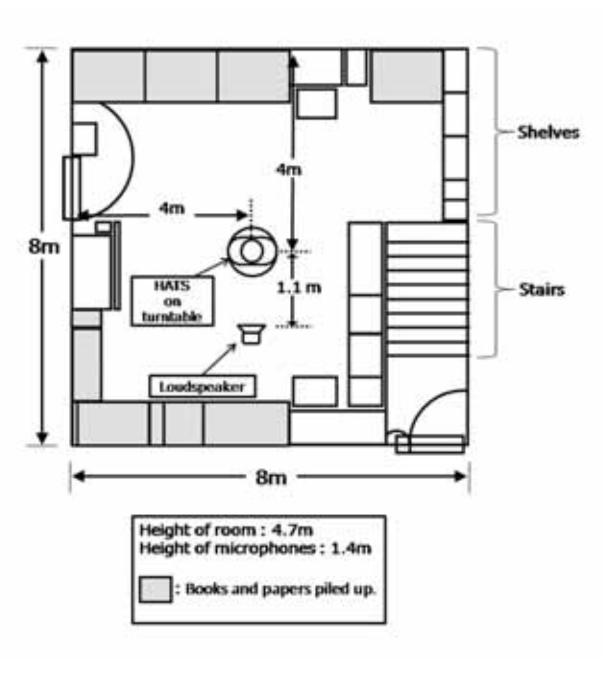

Fig. 1: Room Layout

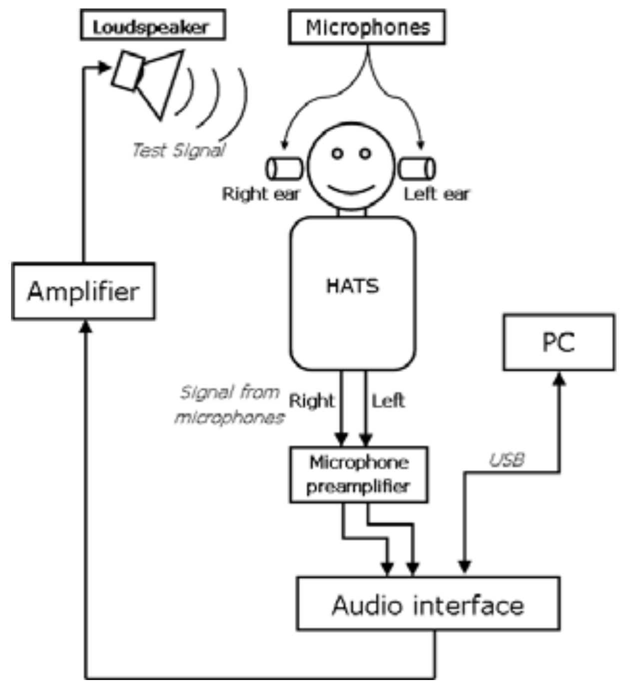

Fig. 2: Connection and Construction of Experimental Equipments

is shorter and many external sounds do not come. The basic dimensions of the room were width 8 m × depth 8 m × height 4.7 m and walls, the ceiling is made of gypsum concrete covered with sponge and fabric, and the flooring is covered with carpet.

HATS was set up in the center of the room and a loudspeaker was placed at a distance of 1.1 m from HATS.

Connection and construction of experimental equipments are shown in Figure 2.

The RHRIR measurement is achieved by recording an impulse signal reproduced from a loudspeaker in an ordinary reverberant room. However, obtaining the high-accurate RHRIR using an impulse signal is difficult because no loudspeakers can correctly reproduce an impulse signal that has the instantaneous rising edge and the large amplitude value. The measurement method using a TSP (Time Stretched Pulse) instead of the impulse signal is able to capture the more precise impulse response (IR) data than the method using an impulse signal.

An advantage of TSP is to improve higher S/N ratio than the conventional impulse signal, and a longer TSP has higher S/N ratio.

3.1.1. Details of Input Signals

Table 2 shows the details of input signals reproduced in the RHRTF measurement.

TSP Duration	2.97 s (65536 points)
TSP's Active Length	61604 points (94%)
Sampling Frequency	44100 Hz
Distance Between Loudspeaker and Microphones	1.1m
Recording Time	5.94 s (131072 points)
Numbers of Recordings	3 times
Binaural Recording	Glass breaking sound

Table 2: Details of Measurement Sounds

Reproduction and Recording of input signals was executed mechanically-synchronized. In order to raise the S/N ratio, the length of TSP was configured rather long value as 1.5 seconds to change frequency gradually and the recording was executed three times per one direction. In addition, since the reverberation sound is most important in this research, the recording during extra 1.5 s was continued after finishing the reproduction of TSP. Moreover, the binaural recording of glass breaking sound was also performed in parallel as a reference sound on auditory experiment.

3.2. Angular Resolution of RHRTF

RHRTF was measured at following angles: azimuth from 0 to 180 degrees by 10 degrees step, and elevation from 0 to 90 degrees by 10 degrees step (Figure 3.) Please notice about our expression of azimuth, azimuth starts from 0 (in front of listener) degree in counterclockwise rotation.

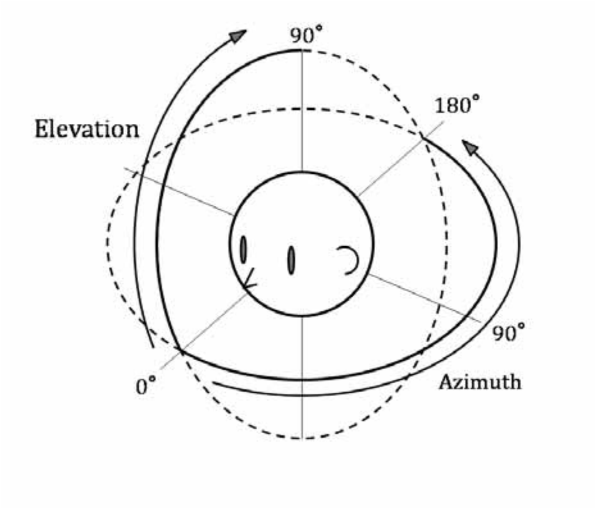

Fig. 3: Expression of azimuth and elevation

3.3. Extraction of RHRIR

Finally the extraction of RHRTF can be executed by convolving recorded TSP signal with ITSP which is inverse signal of TSP.

Figure 4 shows examples of recorded TSP, ITSP and an RHRTF derived from a convolution of TSP with ITSP. The upper figure describes about the time-domain form of the TSP signal that was recorded by a microphone of HATS, the middle figure expresses the time-domain of the ITSP signal, and the last shows the time-domain of the RHRTF, so called RHRIR, derived from a convolution of the recorded TSP with ITSP.

The RHRIR signal contains a frequency characteristic of loudspeaker used in the measurement. So an inverse filter of loudspeaker was generated by LS time-domain method, and the inverse filter was applied to the sound which was created by convolving a dry sound source with the RHRIR. Now, $S_{RHRIR}(t)$ denotes the time-domain of sound source convolved with RHRIR, then $S_{RHRIR}(t)$ was calculated as

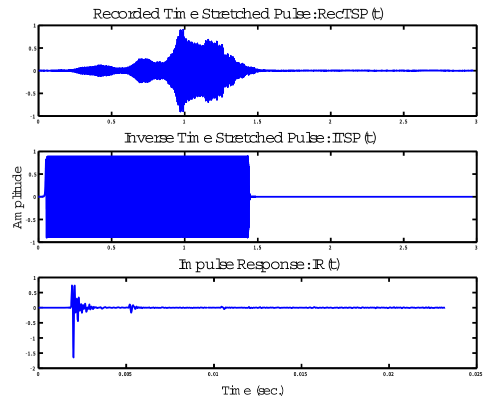

Fig. 4: Recorded TSP, ITSP and RHRTF: This figure describes about the time-domain form of the recorded TSP signal, the middle figure expresses the time-domain of the ITSP signal, and the last shows the time-domain of the RHRTF derived from a convolution of the recorded TSP with ITSP

$$S_{RHRIR}(t) = S_{Dry}(t) \times RHRTF(t) \times INV_{Loudspeaker}(t)$$

where, $S_{Dry}(t)$ denotes a dry sound source, $RHRIR(t)$ means the time-domain of RHRTF and $INV_{Loudspeaker}$ describes the time-domain inverse filter of loudspeaker that was used in RHRTF measurement.

4. AUDITORY EXPERIMENTS

In this section, we evaluate the performance of measured RHRTF with two kinds of auditory experiments. All experiments were conducted by a binaural reproduction system which sounds are played on a headphone.

4.1. Experimental Conditions

Table3 shows common conditions to two experiments.

Number of Subjects	10
Location	Anechoic Chamber
Direction of Sound Image	Azimuth: $45°, 90°, 135°$ ($fixed$)
	Elevation: $0° - 90°$ ($10° step$)
HRTF Type	HRTF, HRTF+Reverb, RHRTF, Binaural
Test Stimuli	Glass Breaking Sound
Headphone	SONY MDR-CD900ST

Table 3: Common Experimental Conditions

Two experiments were conducted to 10 people in an anechoic chamber. We selected the room for the experiments because of few external noises and convenience of in-place equipments although normal rooms suffice as a test room if the atmosphere is quiet and calm. Then, we refer to directions of sound image. From preliminary experiments, it was clearly shown that sound images on the median plane never localized correctly. In these experiments, horizontal directions where sound images were reproduced from were limited to three azimuths, 45, 90 and 135 degrees. Elevation changes between 0 degree and 90 degrees by 10 degree steps at each three azimuths.

In addition, there were four kinds of spatial sounds, the first is a sound convolved with HRTF which was measured by earlier study [6], the second is reverberations added separately to the first one(HRTF+Reverb), the third is a sound convolved with RHRTF measured by this research, and the last is a binaural sound recorded.

A glass breaking sound was used as a stimuli. Therefore in the experiments, (the number of HRTFs) $\times$ (the number of azimuths) $\times$ (the number of elevation) $= 4 \times 3 \times 10 = 120$ sounds were generated. Each directional sounds of above-mentioned 4 kind of spatial sound were adjusted to have same volume in order that sound localizations were not influenced by the difference of sound pressure of each sounds.

4.2. Experiment 1: Comparison of Absolute Perception for HRTF, HRTF+Reverb, RHRTF and Binaural Sounds

The experiment 1 was performed to investigate absolute perception of listeners for four kinds of 3-D sound at three azimuths by reproducing one sound at a time.

The rule of sound reproduction was that azimuth was fixed on an angle θ chosen out of 45, 90, 135 degrees in first and then elevation ϕ changed randomly between 0 and 90 degrees, after then ϕ changed sequentially from 0 to 90 degrees by 10 degree steps again. After finishing to reproduce all sounds at azimuth θ, then the next azimuth was chosen randomly and sound production to listeners continued in same way.

Listeners pointed out the direction of sound image using a stick, and experimenter did visual observation of angular information using angular scales located around listeners.

This sequence of sound reproduction was applied to HRTF, HRTF+Reverb, RHRTF and binaural sounds. The absolute perception of sound image at azimuth 45, 90, 135 degrees were investigated.

The experimental results are shown in Figure 5-7. In each figures,
(a) denotes the elevation perception at each azimuth in the case of random input
(b) expresses the azimuth perception at same azimuth. The data of figures are average value of all listeners. Because the results of sequential input of elevation are almost same as of random input, the experimental results of sequential input are shown in the appendix.

Totally, RHRTF sound localization on azimuth is better than binaural sound. However almost HRTF and HRTF+Reverb sounds were judged unstable. HRTF+Reverb sound localization about direction recognition was worse than we expected, in short, there are little differences between HRTF only and HRTF+Reverb. Therefore, HRTF and HRTF+Reverb sounds are almost the same for evaluation in this section.

From figures, it is shown that RHRTF sounds gave listeners a stable sound localization on vertical direction similar to binaural sounds at azimuth 45 and 135 degrees. Moreover, both perceptions of azimuth and elevation are quite ideal at azimuth 90 degrees.

Localizations of HRTF sounds are also clear at the side direction, however, the accuracy was improved by using RHRTF. These results declared the effectiveness of RHRTF. Also in high elevations at azimuth 45 and 135 degrees, elevation perceptions of RHRTF result in ideal without decreasing accuracy like HRTF sound. At diagonally forward or backward direction, interaural time difference(ITD) and interaural level difference(ILD) becomes smaller, but these differences are considered as basis of sound image localization at this points.

Moreover, it is conceivable that listeners are difficult to localize sound image at those points due to the incidence of in-head localization. Therefore it was obviously established that HRTF cannot allow people to give correct sound image at frontal and diagonal directions. In fact, azimuth and elevation perceptions of HRTF at diagonal directions are distinctly amphibolic although localized to some extent.

Thus, this cannot be apply in practical use. It seems to be caused by individual variation of HRTF mainly. HATS's HRTF is used a typical model of function in this experiment, naturally, it is different from the listener's HRTF. Since, the direction recognition was negatively affected.

On the other hand, RHRTF and binaural sounds those have accurate reverb factors can use not only ITD and ILD but also reverberation sounds to determine the position of sound image. At the same time, it seems that the reverberations equal to a real environment decreased the bad influence by the individual variation of HRTF and the accuracy of the direction recognition improved.

So it is considered that listeners could perceive RHRTF and binaural sounds correctly.
However, all sound images corresponding to frontal azimuth never localized in front direction. Listeners always recognized sound images in side or backward. Since a 3-D sound system using loudspeaker is able to give a listener the image of frontal localization, so this is a particular problem of headphone system. It will be necessary to improve this problem in the future.

4.3. Experiment 2: Comparison of Subjective Impression for HRTF, HRTF+Reverb, RHRTF and Binaural Sounds

The experiment 2 evaluated subjective impressions for HRTF, HRTF+Reverb, RHRTF and binaural sounds' localization and movement.

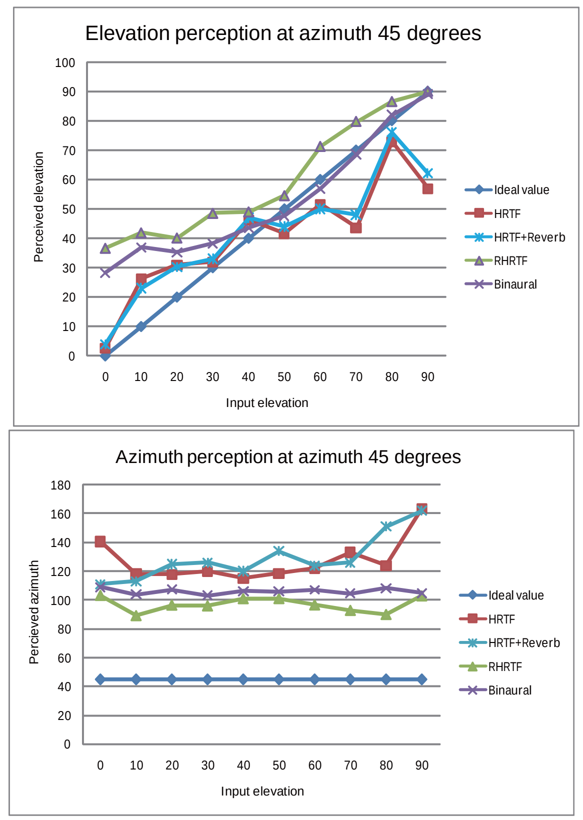

Fig. 5: Elevation and Azimuth Perceptions at Azimuth 45 Degrees

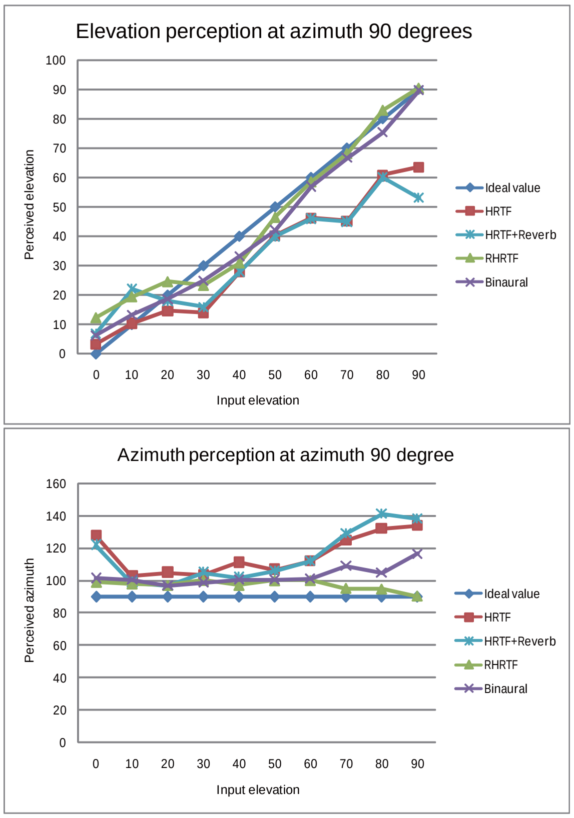

Fig. 6: Elevation and Azimuth Perceptions at Azimuth 90 Degrees

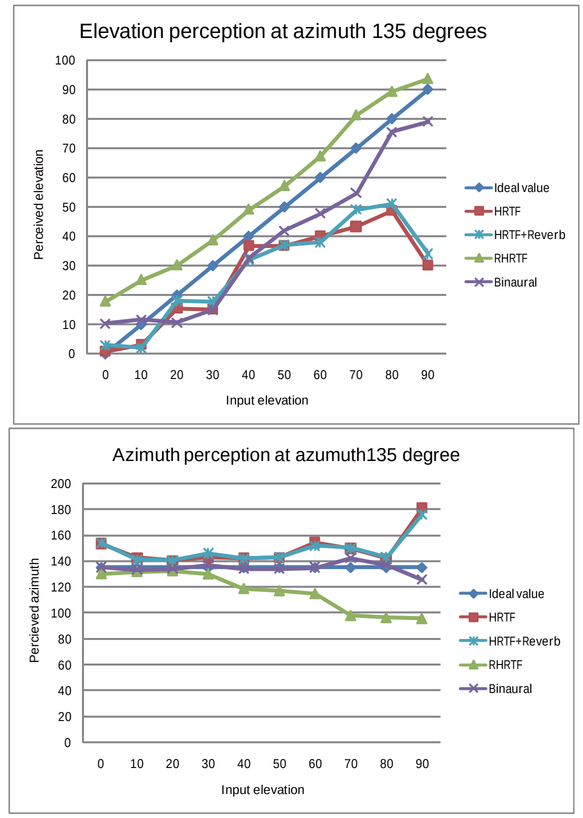

Fig. 7: Elevation and Azimuth Perceptions at Azimuth 135 Degrees

Motion	Could perceive clearly a motion of sound image?
Smoothness	Did a sound move at equal angular distance?
In-The-Head Localization	Did NOT a sound localize in the head?
Sharpness	Could you perceive a tight shape of sound image?
Azimuth Direction	Could hear a sound at azimuth known in advance?
Staggering	Did NOT a sound move with tottering steps?

Table 4: Experimental Questionnaires

In this experiment, each kinds of 3-D sounds were continuously reproduced from elevation 0 to 90 degrees by 10 degree steps at azimuth 45, 90 and 135 degrees. Azimuth was informed to listeners in advance before reproducing continuous sounds. After reproduction, each 3-D sounds were evaluated on an one-to-five scale according to the following Table 4.

The results of experiment for evaluating subjective IMPRESSION about moving sounds are shown in Figure 8-10. First, listeners had same impressions on RHRTF sounds and binaural sounds at all azimuths and it is not resemble to HRTF and HRTF+Reverb.

And, though it was expected that HRTF+Reverb became a good to some extent on evaluation, there is few improvement just like a result of experiment 1.

4.3.1. Motion

In the term of "Motion", impressions for HRTF, HRTF+Reverb, RHRTF and binaural sounds were almost the same. All method can create almost ideal sound image motion.

A sound reproduced at the last produces a lead for listeners to localize a new sound image at higher position when elevation changed continuously. So it is considerable that listeners could easily perceive the movement of sound image by an accumulation of the relative changes for 3D sound reproduction. Hence, there is no difference through all impressions for movements of HRTF, HRTF+Reverb, RHRTF and binaural sounds.

4.3.2. Smoothness

Next, the results of the term "Smoothness" is focused on. Figure 9 and 10 is showing that RHRTF and binaural sounds moved smoothly at azimuth 90 and 135 degrees, especially at azimuth 135 degrees.

From figure 7 shows HRTF and HRTF+Reverb sound images move little from elevation 40 degrees to 80 degrees.

And, focused in detail, smoothness score of HRTF+Reverb is better than HRTF. In Smoothness, simple addition of reverberation seems to be effective.

4.3.3. In-The-Head Localization

Then we state about the term of "In-the-head Localization." From the figures, a definite improvement was confirmed in this term, the sounds containing reverberation components achieved the out-of-head localization. On the other hand, HRTF sounds were localized in the head in many cases. And, the evaluation of HRTF+Reverb has improved at 45 degrees and 90 degrees.

These results are important evidence that any reverberations elements are essential to localize a sound image at out of the head.

In addition, many listeners answered that they could perceive the distance of sound image very clearly. Also from the answers, the advantage of RHRTF was demonstrated.

4.3.4. Sharpness

In the term of "Sharpness", HRTF sounds took an amazing turn and became a high-scorer. The reasons are assumed as follows. HRTF sounds are constructed by only the direct sound. So listeners could perceive the shape of sound image of HRTF sound very clearly even if the localization was not precise. Meanwhile, the other kind of sounds presented listeners many sound images of reverberation. Therefore, it can be considered that the sharpness of sound images with reverbration became blurry compared with HRTF only sounds.

4.3.5. Azimuth Direction

In the term of "Azimuth Direction", there is no remarkable differences among four sounds. From Figure 8, it was shown that listeners could not localize sound images of 3-D sounds at front direction.

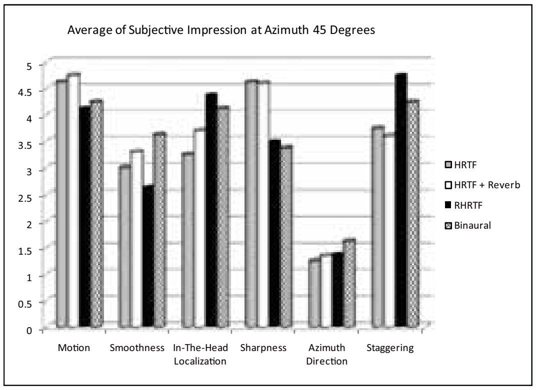

Fig. 8: Subjective Auditory Impression at Azimuth 45 deg.

Alternatively, sounds were always localized in the backward direction despite azimuth angles of sound image were informed in advance. This is a peculiar problem to the headphone system, therefore the evaluation of 45 degrees is worse.

4.3.6. Staggering

Finally, we are going to state about evaluations for the term "Staggering", which expresses about the straightness of movements of sound images. Figure 8 and 10 shows that HRTF sounds of azimuth 45 and especially 135 degrees tottered from elevation 0 to 90 degrees and HRTF+Reverb made little effect compared with HRTF.

On the other hand, auditory images of RHRTF and binaural sound rose up to higher elevations. These results point to a possibility that the reverberation components have an effect on the stabilization of azimuth localization. Meanwhile, all sounds were localized in a straight line at azimuth 90 degrees for the feature that headphone system reproduces sound right besides listener's ears.

5. CONCLUSIONS AND FUTURE WORKS

As a result, we concluded that the sound with reverberations especially RHRTF sounds was perceived more correctly than HRTF sounds and reverberation is also necessary for 3D spatial sound especially to improve reality.

However, simple addition of reverberation (HRTF+Reverb) cannot decrease an unstable

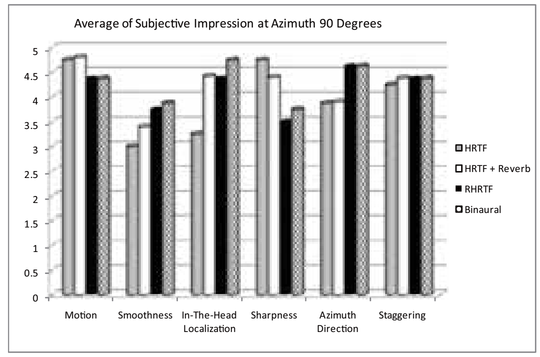

Fig. 9: Subjective Auditory Impression at Azimuth 90 deg.

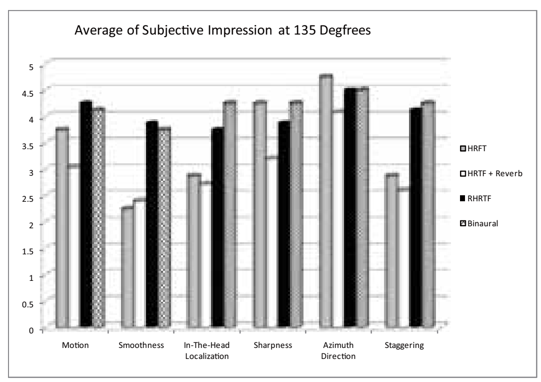

Fig. 10: Subjective Auditory Impression at Azimuth 135 deg.

location perception and unnaturalness. About this, better setting might exist, so we might have to consider in the future,

Though it is very effective only in the limited condition to have to construct a real environment for a binaural recording, RHRTF does not have this limit. So RHRTF is the most effective technique as a method of producing 3D sound image with the reality easily.

To apply RHRTF method, whatever dry sources, we can produced the 3D spatial sound images in more reality by only signal processing, not adjust complex reverberation elements or create construct real environment.

6. REFERENCES

[1] a. saji "3-D Sound Reproduction by HRTFs Combined with the Amplitude Panning Method", *Master's Thesis of the University of Aizu*, February 2004.

[2] K. Komura, "Measurement and Analysis of HRTFs", *Master's Thesis of the University of Aizu*, February 2008.

[3] S. Yano, S. Shimada and H. Hokari, "The Distance Perception of Sound Image by the Components of Reflection Wave", *Technical Report of IEICE EA*, vol. 96, no. 322, pp. 9-16, November 1996.

[4] K. Saito, Y. Iwaya and Y. Suzuki, "The Technique of Choosing the Individualized Head-Related Transfer Function Based on Localization", *Technical Report of IEICE EA*, vol. 104, no. 247, pp. 1-6 August 2004.

[5] N. Saito, "Reconstruction of Binaural Impulse Responses for 3-D Sound Generation", *Master Thesis of the University of Aizu*, February 2007.

[6] T. Nakamura, "High Resolution HRTF Measurement and Compression based on the Principal Components Analysis", *Master Thesis of the University of Aizu*, February 2005.

[7] H. Mano, "Evaluation of Correction Filters for Different Types of Earphones and the Influence on Spatial Sound Perception", *Master Thesis of the University of Aizu*, February 2005.

[8] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction", *J. Audio Eng. Soc.*, vol. 47, no. 7/8, pp. 583-595, July/August 1999.

[9] Bose Corporation, "BOSE Direct/Reflecting Technology", http://www.bose.co.uk/GB/en/learning-centre/direct-reflecting/4-applied-learning/index.jsp

[10] S. G. Norcross, G. A. Soulodre and M. C. Lavoie, "Subjective Investigations of Inverse Filtering", *J. Audio Eng. Soc.*, vol. 52, no. 10, pp. 1003-1028, October 2004.