A PRELIMINARY DEVELOPMENT OF HIGH DEFINITION VIRTUAL ACOUSTIC DISPLAY BASED ON *ADVISE*

Yôiti Suzuki†, *Shouichi Takane*∗, *Shusuke Takahashi*†¶ *and Tohru Miyajima*‡

† Research Institute of Electrical Communication
and Graduate School of Information Science, Tohoku University
2-1-1 Katahira, Aoba-ku, Sendai, Miyagi, 980-8577 Japan
E-mail: `yoh@ais.riec.tohoku.ac.jp`
∗ Faculty of Systems Science and Technology, Akita Prefectural University
84-4 Ebinokuchi, Tsuchiya, Honjo, Akita, 015–0055 Japan
‡ Institute of Technology, Shimizu Corporation
3-4-17 Etchujima, Koto-ku, Tokyo, 135-8530 Japan
¶Currently in Sony Corporation

ABSTRACT

A method for high definition virtual acoustic display (VAD) based on a model called the "Virtual Sphere (VS) model" is introduced [S. Takane *et al.*, Proc. ICAD2002(2002)]. This is named *ADVISE* (Acoustic Display based on the VIrtual SpherE model). In this paper, sound reproduction accuracy of *ADVISE* is first discussed via a computer simulation. Based on the simulation results, a practical way to determine necessary parameters for *ADVISE* is discussed, and these parameters are approximately given for the implementation under reasonable assumptions. Then, a real-time implementation of *ADVISE* is introduced. This system is realized as a binaural VAD with five floating-point DSPs (Texas Instruments TMS320C6701) hosted by a PC. Consideration of hardware performance and the programming in each DSP and the host computer is also mentioned. Results of a preliminary hearing experiment with this system showed a satisfactory performance.

1. INTRODUCTION

A method for high definition virtual acoustic display (VAD) based on a model called the "Virtual Sphere (VS) model" has been introduced [1]. This method is named *ADVISE* (Acoustic Display based on the VIrtual SpherE model)[2]. The Kirchhoff-Helmholtz boundary integral equation is applied as the theoretical basis of *ADVISE*. This equation asserts that a sound field generated by sound sources outside a certain closed boundary can be synthesized by phantom monopole and dipole sources distributed on the VS boundary. To implement this idea, a virtual boundary of spherical shape (named the VS boundary) is set around a listener, and sound transmission from the sound sources to the listener is divided by the VS boundary into two parts. One is the transmission system involving the transfer functions from the sound sources to the points on the VS boundary. These transfer functions reflects the acoustic properties of the sound field to be reproduced. They are called Room Transfer Functions (RTFs). The other part is the transmission system involving the Head-Related Transfer Functions (HRTFs) corresponding to the location of the phantom sources on the VS boundary. Sound pressure at the entrance of the listener's ears is reproduced by convolving the corresponding RTFs and HRTFs for each direction and then by summing the results for all directions around the listener. By this dividing, the change of the sound pressure at the listener's ears caused by the movement of the listener within VS, can be traced by changing only the HRTFs for the phantom sources distributed on the VS boundary. Moreover, the sound pressure in a certain sound field can be simultaneously presented to different listeners with different movement if the set of HRTFs for each listener may be provided. In this paper, practical implementation of *ADVISE* are discussed in three points. Firstly, sound reproduction accuracy is examined via a computer simulation. A real-time system based on *ADVISE* is then developed by using multiple DSPs hosted by a PC. Some approximations required in the development are additionally mentioned, and the results of a simple hearing experiment are also stated.

2. SOUND REPRODUCTION ACCURACY IN *ADVISE*

To examine the sound reproduction accuracy in *ADVISE*, a simple computer simulation was carried out with assuming the listener's head as an acoustically rigid sphere.

2.1. Conditions of the computer simulaton

The system to be simulated is illustrated in Figure 1, and the conditions of the computer simulation is depicted in Table 1. The sound field with a sound source located 2 m distant from the listener in free field was reproduced by using the method of *ADVISE*. Radius of the rigid sphere a was set to 8.5 cm refering to the size of the human head. The angle of incidence was 0, 45 and 90 deg., where 0 deg. indicates the sound source located in front of the listener, and 90 deg. means the sound source at the right of the listener. Radius of the virtual sphere boundary r was set to 0.25, 0.5 and 1.0 m, and the distance between adjacent nodal points D was 5 and 10 cm. The smaller the radius of the virtual sphere boundary is, the smaller the number of nodal points on the boundary becomes. These conditions are summarized in Table 1. The interval between $\boldsymbol{r}_i^+$ and $\boldsymbol{r}_i^-$ is indicated as δ. This interval was constant over every nodal point i and was set to 2, 4 and 8 cm. This parameter affects the accuracy in observing the normal derivative of the sound pressure at $\boldsymbol{r}_i$.

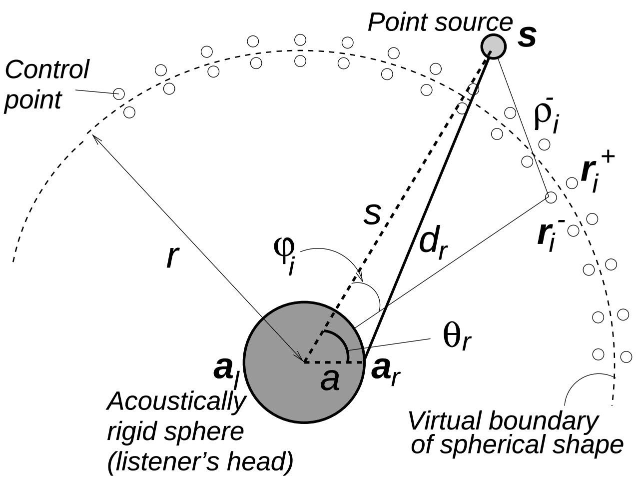


Figure 1: Setup of the computer simulation

Table 1: Conditions of the computer simulation

Condition	Value
Radius of sphere a	8.5 [cm]
Sound source distance s	2 [m]
Angle of incidence	0, 45, 90 [deg.]
Radius of VS boundary r	0.25, 0.5, 1.0 [m]
Distance between adjacent nodal points D	5, 10 [cm]
δ	2, 4, 8 [cm]

Under these conditions, an error level of the reproduced sound field at the points corresponding to the listener's ears were computed. This error levels at the listener's both ears are defined as follows:

$$E_{\mathrm{L}}(\omega) = 20\log_{10}\left|\frac{P_{\mathrm{L}}(\omega) - P'_{\mathrm{L}}(\omega)}{P_{\mathrm{L}}(\omega)}\right| \quad [\mathrm{dB}] \tag{1}$$

$$E_{\mathrm{R}}(\omega) = 20\log_{10}\left|\frac{P_{\mathrm{R}}(\omega) - P'_{\mathrm{R}}(\omega)}{P_{\mathrm{R}}(\omega)}\right| \quad [\mathrm{dB}] \tag{2}$$

where $P_{\mathrm{L}}(\omega)$ and $P'_{\mathrm{L}}(\omega)$ respectively indicate the sound pressure at the point corresponding to the listener's left ear in the original sound field and the reproduced sound field. The same relation stands for the suffix R as well.

2.2. Results and discussion

Figures 2 and 3 respectively show the frequency characteristics of $E_{\mathrm{L}}(\omega)$ and $E_{\mathrm{R}}(\omega)$ for various directions of the sound source when the radius of the VS is 25 cm. Solid line indicates the characteristics when the source direction is 0 deg., and the dash interval of the dashed lines becomes small as the direction angle of the source increases. These figures show that the error level of the reproduced sound pressure by *ADVISE* are less than about -20 dB for almost entire frequency range calculated. As the simulated sound source locates at the right side of the listener, the left ear locates contralateral, *i.e.* shadow side, to the sound source and the right ear locates ipsilateral. Hence only the sound diffracted by the rigid sphere arrives at the left ear, and the power of the observed response at this point is smaller than that at the right ear. These figures indicate that this affects the characteristics of error levels at each ear, that is, $E_{\mathrm{L}}(\omega)$ is higher than $E_{\mathrm{R}}(\omega)$ when the source direction is not 0 deg. Moreover, the abrupt increase of the error level at around 1950 Hz is observed in both panels when the angle of incidence is 90 deg. The reason for this phenomenon is that the element size of the divided VS boundary is no more small comparing with the wavelength corresponding to this frequency, and the error due to this is emphasized by the convolution of HRTFs.

Figures 4 and 5 show the frequency characteristics of $E_{\mathrm{L}}(\omega)$ and $E_{\mathrm{R}}(\omega)$ for various directions of the sound source when the radius of the VS is 50 cm. Overall tendency is the same as shown in Figures 2 and 3, except that the abrupt increase of the error level can no longer be observed even when the angle of incidence is 90 deg. Error level decreases, that is, the sound reproduction accuracy in *ADVISE* becomes higher than that when the radius of the VS is 25 cm. However, the enlargement of the VS leads to the increase in the number of the phantom sources and thus the increase in the processing power to convolve RTFs and HRTFs. There exists a tradeoff between the system size determined by the radius of the VS and the sound reproduction accuracy obtained.

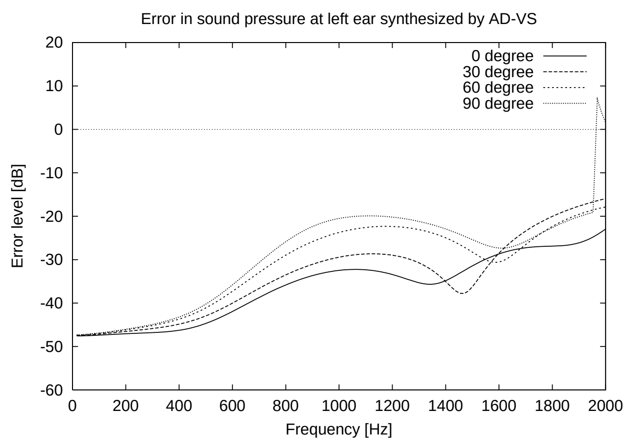


Figure 2: Frequency characteristics of the error level of the reproduced sound pressure at the point corresponding to the listener's left ear, for three angles of incidence; $D = 10$ cm, $\delta = 2$ cm, $r = 25$ cm.

3. DEVELOPMENT OF *ADVISE* WITH MULTIPLE DSPS

The results of the computer simulation in **2** indicate that the sound pressure at the listener's ears can be accurately synthesized by using the VS model in *ADVISE*. In the following subsections, development of a real-time system based on *ADVISE* is introduced.

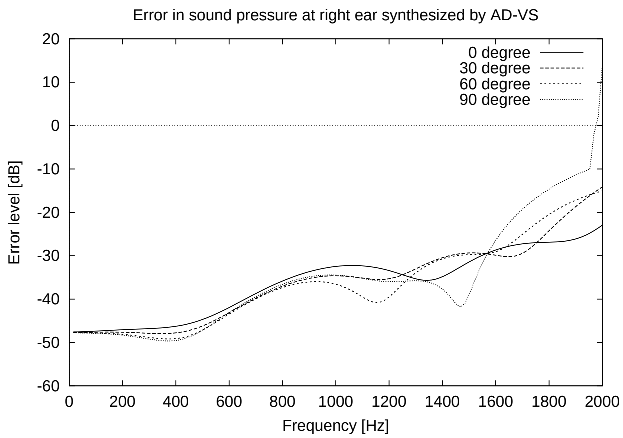


Figure 3: Frequency characteristics of the error level of the reproduced sound pressure at the point corresponding to the listener's right ear, for three angles of incidence; $D = 10$ cm, $\delta = 2$ cm, $r = 25$ cm.

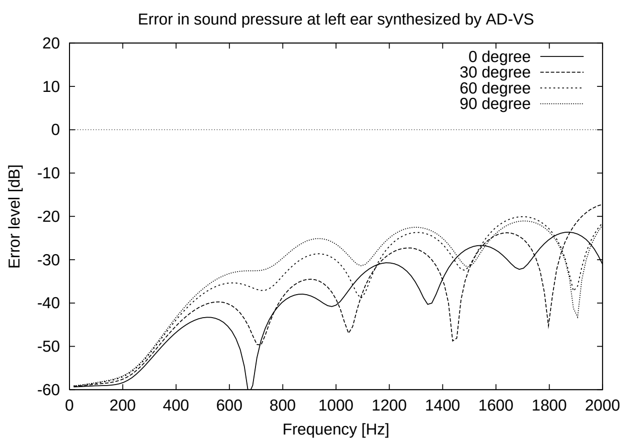


Figure 4: Frequency characteristics of the error level of the reproduced sound pressure at the point corresponding to the listener's left ear, for three angles of incidence; $D = 10$ cm, $\delta = 2$ cm, $r = 50$ cm.

3.1. Parameters required in *ADVISE* and their setting

Considering the practical implementation, the following parameters are required to develop the acoustic displays based on *ADVISE*:

1. Sound pressure and its normal derivative at the points on the VS
2. HRTFs for the points on the VS

Parameters of the first item can be synthesized from RTFs and the source signals, and the normal derivative of sound pressure can be approximated by the difference of sound pressure at two closely located points. Hence, RTFs and HRTFs for the points on the VS boundary as well as the source signals are required.

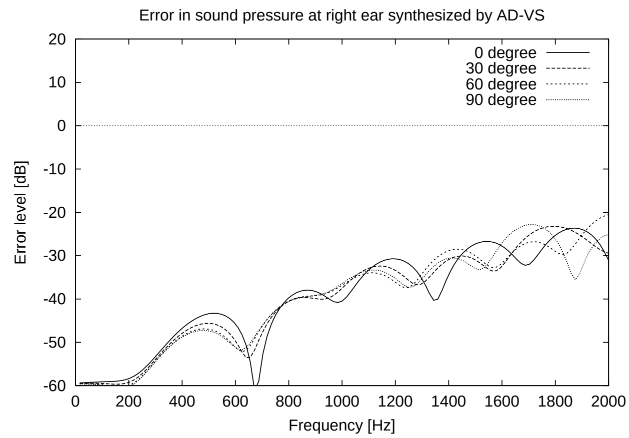


Figure 5: Frequency characteristics of the error level of the reproduced sound pressure at the point corresponding to the listener's right ear, for three angles of incidence; $D = 10$ cm, $\delta = 2$ cm, $r = 50$ cm.

As for HRTFs, some databases exist measured for a dummy-head or human subjects[3, 4]. HRTFs for the points on the VS boundary must be provided for the monopole sources. Since ideal point sources are hardly available in practice, possible differences between HRTFs measured by using a loudspeaker and those by using a point source is discussed via a simple computation. Figure 6

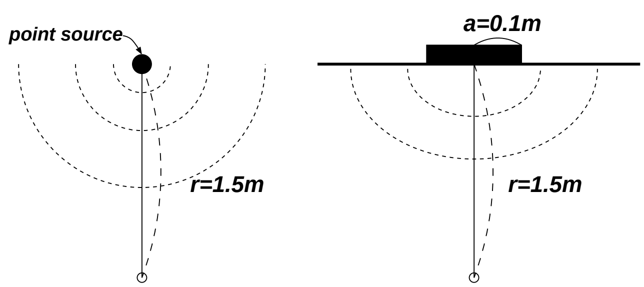


Figure 6: A point source and planar source on baffle

is an illustration of a point source and a planar source of which radius is 0.1 m on an infinite baffle. Sound pressures at the point of the same distance (1.5 m) from the point source and the planar source are computed, and the error level is plotted in Figure 7. The result shows that the error level is small in low frequency region, while it increases in high frequency region. In this condition, the error level is less than -20 dB when the frequency is less than 2 kHz. This error level increases as the distance from these sources are small. This implies that HRTFs measured by using ordinary loudspeakers may contain larger errors as the radius of the VS becomes smaller.

HRTFs for two closely located points at the VS boundary are also needed in *ADVISE*. When HRTFs can be approximated by regarding listener's head as acoustically rigid sphere, and the radius of the the VS boundary is sufficiently large, the following approximation is possible[5] for $H_{\mathrm{L}}(\boldsymbol{r}_i, \omega)$ (using the same notation as

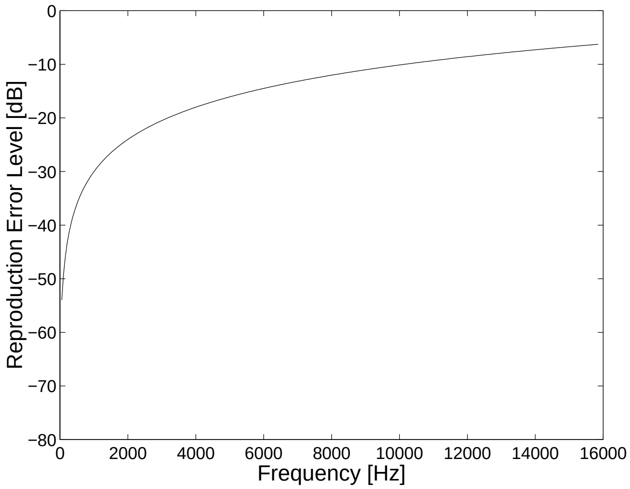


Figure 7: Error level between the sound pressure radiated by a point source and that by a planar source in infinite baffle.

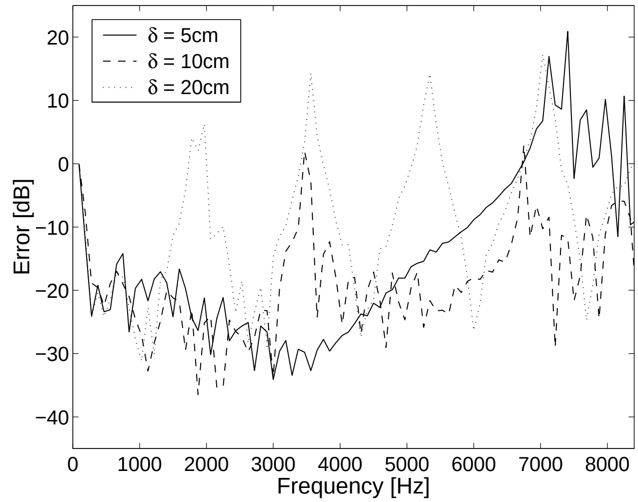


Figure 8: Error level of approximation of the left-ear HRTF difference for various δ_i (Azimuth: 45 degrees, distance: 1.5 m)

that in the accompanying paper by the authors[2]):

$$H_{\mathrm{L}}(\boldsymbol{r}_i + \Delta\boldsymbol{r}_i, \omega) \approx H_{\mathrm{L}}(\boldsymbol{r}_i, \omega) \cdot \frac{1}{1 + \delta_i / r_i} e^{-jk\delta_i}, \qquad (3)$$

where $r_i = |\boldsymbol{r}_i|$, $\delta_i = |\Delta\boldsymbol{r}_i|$, $k = \omega/c$ and c is the speed of sound. The same approximation is also possible for $H_{\mathrm{R}}(\boldsymbol{r}_i, \omega)$. Validity of this approximation is examined by computing the following equation with HRTFs measured in an anechoic room:

$$E_\Delta(\omega) = 20 \log_{10} \left| \frac{\Delta H_{\mathrm{M}}(\omega) - \Delta H_{\mathrm{A}}(\omega)}{\Delta H_{\mathrm{M}}(\omega)} \right|, \qquad (4)$$

where $\Delta H_{\mathrm{M}}(\omega)$ is the calculated difference from measured HRTFs and $\Delta H_{\mathrm{A}}(\omega)$ is that computed from the above approximation of a head as being a rigid sphere. Frequency characteristics of $E_\Delta(\omega)$ at the distance $r_i = 1.5$ m for various δ_i are shown in Figure 8. It is notable that the error levels are relatively large at very low frequency region for any δ_i. Moreover, several peaks due to the approximation of the derivative to the difference are observed when δ_i is as large as 10 cm and 20 cm. $E_\Delta(\omega)$ smaller than about -20 dB may be obtained from a few hundred Hz to 5 kHz if δ_i is set to 5 cm.

If listeners are allowed to translate inside the VS boundary, it may be difficult to obtain appropriate HRTFs by measurements. Furthermore, RTFs for the practical room may also be hard to be obtained with taking their change with the movement of sound source into account. In these cases, numerical methods such as Boundary Element Method (BEM) and Finite Element Method (FEM), etc. would be effective tools to provide even such HRTFs[6, 7, 8, 9].

3.2. Development of a real-time system based on *ADVISE*

Based on the theory of *ADVISE*[2], a real-time system has been developed by using five fast DSPs hosted by a PC. The parameters for *ADVISE* are determined by referring to the results of the previous section. Outline of the system is illustrated in Figure 9. Here, binaural presentation is chosen as the reproduction system. Transaural system needs real-time processing for the compensation of the change of the sound transmission systems in the reproduction sound field[10, 11], but such a processing is not needed in a binaural system. Two-dimensional sound field is assumed as a preliminary investigation. This system consists of a D/A board, three DSP boards and a magnetic sensor. Position and head rotation of the listener are sensed by a magnetic sensor (Polhemus FASTRAK), and the obtained data are transmitted in every 1/120 s to the DSP boards in serial line with the transmission rate of 38.4 kbps. Two of the three DSP boards (System Design Service PCI-DSP26701) are with two DSPs (Texas Instruments TMS320C6701) each, and the other one DSP board (System Design Service PCI-DSP6701F) is with one DSP and an interface to the D/A board (System Design Service PC-SAD2080D). Five DSPs in total are therefore used to execute the signal processing of *ADVISE* in real-time. Processings for the phantom sources distributed on the VS boundary are assigned to each DSP, according to their directions. They are indicated as the circles filled with different patterns in Figure 9. In each DSP, sound pressure and its derivative, HRTFs and their derivatives are stored in memory. HRTFs and their derivatives are interpolated from their stored data, and the sound pressure and its derivative at the points on the VS boundary is convolved to them in real-time. Results of the convolution are transmitted from two PCI-DSP26701 boards to PCI-DSP6701F. Interface between them is McBSP (Multi-channel Bufferd Serial Port) of which maximum transmission rate is 50 Mbps[12].

In the current development, the sampling frequency is 32 kHz and the length of FIR filters for HRTFs is set to 128. Moreover, only the head rotation of the listener is assumed, and all the sound sources are located outside the VS boundary without any movement. In this case, the source signals and the RTFs can be convolved beforehand, and the real-time change is needed only for the temporal change of the HRTFs caused by the listener's head rotation. TMS320C6701 is a floating-point processor with a maximum processing speed of 1 GFLOPS[13], thus the convolution for 6 phantom sources on the VS boundary is possible in a single DSP

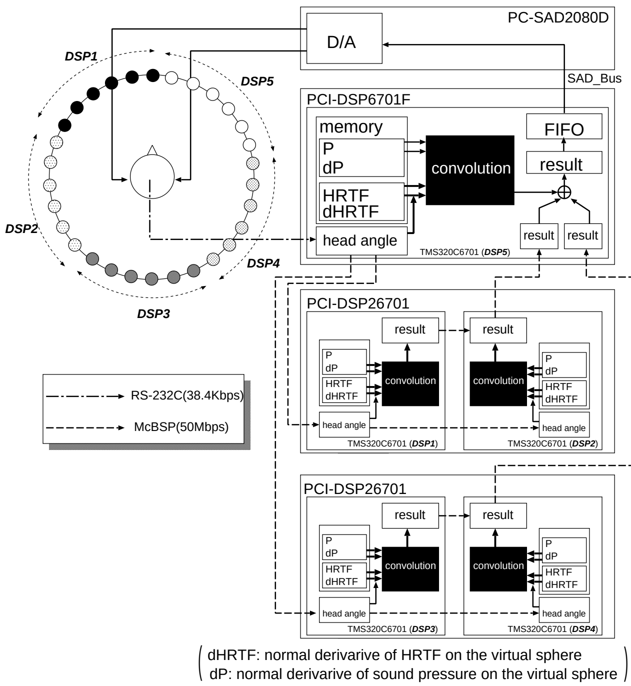


Figure 9: Outline of the developed real-time system based on *ADVISE*

with taking into account the overhead in each DSP. This means that 30 phantom sources may be distributed on the VS boundary by using five DSPs. If the radius of the VS is set to 1.2 m, distance between the phantom sources is around 0.6 m, which corresponds to about the twice of the wavelength of 1 kHz sound. Hence the frequency range is limited to 1 kHz at this stage.

4. HEARING EXPERIMENT

4.1. Conditions of the experiment

A simple hearing experiment investigating the fundamental performance of the developed system described in **3.2** is executed. The experimental setup is depicted in Figure 10.

Virtual sound sources in a free field located on the circle with 1.5 m radius with the interval angle of 30 deg. were simulated. Sound from either of the virtual sound sources was synthesized by using the developed system based on *ADVISE* with the radius of the VS set to 1.2 m. For comparison, conventional direct synthesis of the transfer functions from the virtual sound sources to the listener was also implemented. The sound synthesized by each system was presented to the listener through an open-ear type headphone with a magnetic sensor fixed on its top. Subject's head rotation was sensed only for yaw, and the horizontal change in HRTFs by his rotation was compensated.

HRTFs measured by the authors for a certain male subject and open in public via WWW[4] were applied to the system. Subjects are four males with normal hearing, but the used set of HRTFs in this experiment is corresponding to none of them, *i.e.*, the stimuli were presented to all subjects processed with the other's HRTFs.

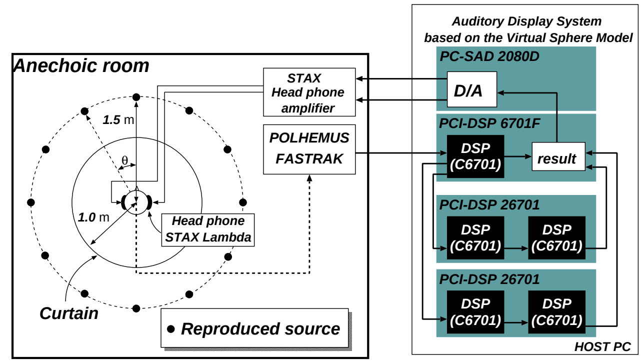


Figure 10: Illustration of the hearing experiment

Low-passed white noise with cut-off frequency of 800 Hz was used as a source signal to each virtual source. Duration of the source signal was 2 s, and interval between stimlui was 30 s. Each subject was asked to answer the perceived direction of the presented sound. Curtain was hung to avoid any visual information around the listener. Stimulus for one direction was presented 5 times for each synthesis method, hence a subject listens to 120 stimuli in the entire experiment, which consists of five sessions.

4.2. Results and discussion

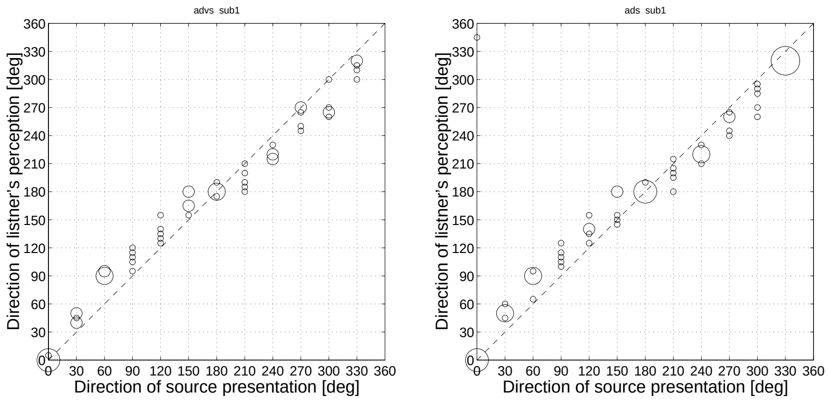
The presented directions and the perceived directions were compared for each method. Results for subjects #1 and #2 are respectively shown in Figures 11 and 12. It is found in Figure 11 that Sub. #1 can perceive the sound image in the same direction as that presented. In contrast, Sub. #2 has some front-back confusions when the presented source direction is near 180 deg., as seen from Figure 12. These phenomena might reflect how the HRTFs were close to the subject's ones. Since non-individual HRTFs were used for all subjects in this experiment, the front-back confusions may be attributable to some discrepancies between the subject's own HRTFs and those applied to the present system[16, 17].

Comparing two panels in each figure, overall tendency is almost the same, *i.e.*, both subjects perceived the sound image synthesized by using both methods in the same manner. This means that the developed system based on *ADVISE* has a good basic performance as an auditory display.

5. CONCLUDING REMARKS AND FUTURE WORKS

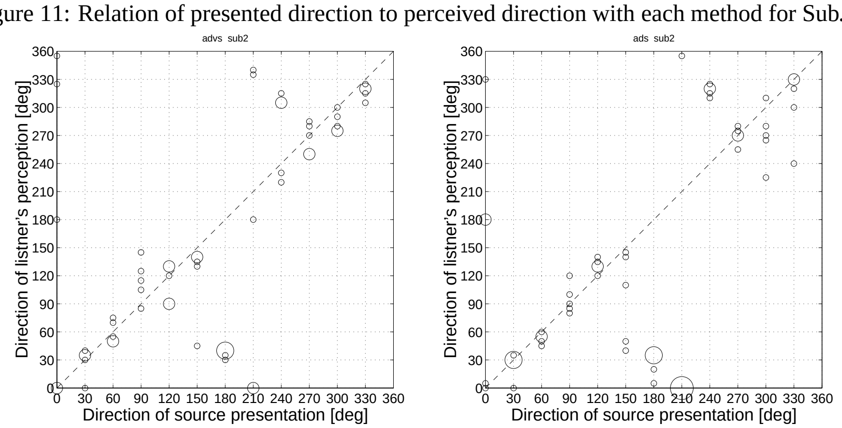
A preliminary development of a high definition virtual acoustic display based on *ADVISE* is introduced in this paper. In the current stage, two-dimensional sound field with a listener allowed to rotate the head but staying at the center of the VS can be synthesized by using a PC with five currently available fast DSPs.

Application of *ADVISE* to the more practical environments is possible if the processing power would be further improved. Moreover, the synthesis algorithm used in the implementation can also be improved. For example, the real-time convolution algorithm using FFT exists[14] while the conventional time-domain convolution is adopted in the current system. HRTFs can be coded more effectively than those using the FIR coefficients. One of such methods is proposed by Watanabe *et al.*[15].



(a) *ADVISE* (b) Conventional

Figure 11: Relation of presented direction to perceived direction with each method for Sub. #1



(a) *ADVISE* (b) Conventional

Figure 12: Relation of presented direction to perceived direction with each method for Sub. #2

6. REFERENCES

[1] S. Takane, T. Miyajima, Y. Yamada, D. Arai, Y. Suzuki and T. Sone, "An auditory display based on virtual sphere model," *Proc. ASVA*, 379-383(1997).

[2] S. Takane, Y. Suzuki, T. Miyajima and T. Sone, "*ADVISE*: A new method for high definition virtual acoustic display," to be presented in *2002 International Conference on Auditory Display*(2002).

[3] `http://sound.media.mit.edu/KEMAR.html`

[4] S. Takane, D. Arai, T. Miyajima, K. Watanabe, Y. Suzuki and T. Sone, "A database of Head-Related Transfer Functions in whole directions on upper hemisphere," *Acoust. Sci. Tech.*, **23**(3), 160-162(2002).

[5] *Handbook of mathematical functions*, edited by M. Abramowitz and I. A. Stegun (Dover Publications, New York, 1972).

[6] S. Takane, T. Sakurai, Y. Suzuki and T. Sone, The effect of assumption of local reaction to numerical analysis of sound field in an enclosure," *Proc. Inter-noise*(1999).

[7] M. Otani and S. Ise, "Numerical calculation of the Head-Related Transfer Functions by using the Boundary Element Method," *Proc. WESTPRAC VII*, **1**, 305-308(2000).

[8] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation," *J. Acoust. Soc. Am.*, **110**(5), Pt. 1, 2440-2448(2001).

[9] B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," *J. Acoust. Soc. Am.*, **110**(5), Pt. 1, 2449-2455(2001).

[10] W. G. Gardner, *3-D audio using loudspeakers* (Kluwer Academic Publishers, Boston, 1998).

[11] J. Garas, *Adaptive 3D sound systems* (Kluwer Academic Publishers, Massachusetts, 2000).

[12] *TMS320C6x periferals reference guide* (Texas Instruments, 1998).

[13] *TMS320C6701 digital signal processor data sheet* (Texas Instruments, 1998).

[14] Paul M. Embree, *C algorithms for real-time DSP* (Prentice Hall, New Jersey, 1995).

[15] K. Watanabe, S. Takane and Y. Suzuki, "A new interpolation method of HRTF based on the Common Pole-Zero model," *Proc. 17th International Congress on Acoustics*, 3D.04.03(2001).

[16] J. Blauert: *Spatial Hearing* (MIT Press, Cambridge, 1983).

[17] D. R. Begault: *3-D SOUND for virtual reality and multimedia* (AP Professional, Cambridge, 1994).