

On the simulation of sound localization

Masayuki Morimoto and Yoichi Ando

Faculty of Engineering, Kobe University, Nada, Kobe, 657 Japan

(Received 7 May 1979)

Sound localization both in the horizontal and the median planes was simulated by the aid of a digital computer through two loudspeakers located in the traverse plane. The simulation was performed by the use of the measured head-related transfer function (HRTF) for two ears. Results of localization tests of three subjects with different sized pinnae show that: (1) When the subject's own HRTF was used in the simulation, natural sound localization in both planes can be performed with the same accuracy as real sound sources. (2) The differences between individual HRTFs affect the accuracy of localization in the simulation. In particular, in the median plane, an HRTF with a high angle-dependency seems to be more advantageous than one with a low angle-dependency.

PACS number: 43. 66. Pn, 43. 66. Qp, 43. 85. Ta,
43. 88. Md, 43. 88. Vk

1. INTRODUCTION

In performing the subjective tests, the simulation of sound fields in an anechoic chamber is a powerful method. The directional informations of incoming sound signals, i.e., "sound localization", should be simulated because it includes an important effect on the subjective evaluations of sound fields.¹⁾

Schroeder *et al.*²⁾ synthesized locations of sound sources only in the frontal half of the horizontal plane by a symmetrical two-channel loudspeaker system, using the head-related transfer functions (HRTF). Ando *et al.*³⁾ generalized it to the asymmetrical two-channel loudspeaker system because a head and pinnae are in general asymmetric.

The purpose of this paper is to confirm the reliability of the simulation of sound localization at any position in a three dimensional space, by the use of the asymmetrical HRTF. Also, in order to discuss the effects of difference between individual HRTFs on localization, the measured HRTFs of three subjects with different sized pinnae are used in the simulation. The difference may be caused by the size of the pinna, its configuration and other boundary conditions of the head.

2. METHOD

To begin with, three special planes relating to a subject are defined as follows: A horizontal plane is the plane on which the left and the right tragus and the right orbitare lie. A median plane intersects the midpoint (the center of a head) of a straight line connecting the left and the right tragus at right angles to the line. A traverse plane intersects both the horizontal and the median planes at right angles and contains the left and the right tragus (Fig. 1).

2.1 Theory

As shown in Fig. 1, let the pressure impulse responses for the paths from a real sound source, $s(t)$, to the entrances of the left and the right auditory canals in a free field be

$$h_{l,r}(t) = h_{l,r}(t; \gamma, \theta, \psi),$$

where l and r indicate the left and the right ears, respectively, γ is the distance between a sound source and the center of a subject's head, θ is the elevation angle, and ψ is the azimuth angle.

The sound pressures, $p_l(t)$ and $p_r(t)$, at the entrances to be simulated are expressed by

$$\left. \begin{aligned} p_l(t) &= s(t) * h_l(t), \\ p_r(t) &= s(t) * h_r(t), \end{aligned} \right\} \quad (1)$$

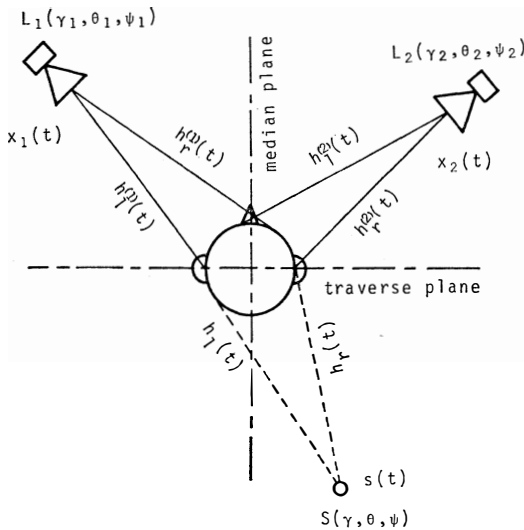


Fig. 1 Geometry and the pressure impulse responses to the entrances of the left and the right auditory canals. $s(t)$ is a sound source to be simulated and L_1 and L_2 are loudspeakers for reproduction.

where the asterisk signifies convolution.

Also, let the pressure impulse responses for the paths from the two loudspeakers, L_1 and L_2 , for reproduction to the entrances be

$$\begin{aligned} h_{1,r}^{(1)}(t) &= h_{1,r}(t; \gamma_1, \theta_1, \psi_1), \\ h_{1,r}^{(2)}(t) &= h_{1,r}(t; \gamma_2, \theta_2, \psi_2). \end{aligned}$$

Then the pressures, $p_l(t)$ and $p_r(t)$, being synthesized through the two loudspeakers are expressed by

$$\begin{cases} p_l(t) = x_1(t) * h_{1,l}^{(1)}(t) + x_2(t) * h_{1,l}^{(2)}(t), \\ p_r(t) = x_1(t) * h_{1,r}^{(1)}(t) + x_2(t) * h_{1,r}^{(2)}(t), \end{cases} \quad (2)$$

where $x_1(t)$ and $x_2(t)$ are the input signals to be fed into the loudspeakers, respectively.

Taking the Fourier transform on the right hand sides of Eqs. (1) and (2), we obtain

$$\begin{cases} S(\omega) \cdot H_1(\omega) = X_1(\omega) \cdot H_1^{(1)}(\omega) + X_2(\omega) \cdot H_1^{(2)}(\omega), \\ S(\omega) \cdot H_r(\omega) = X_1(\omega) \cdot H_r^{(1)}(\omega) + X_2(\omega) \cdot H_r^{(2)}(\omega). \end{cases} \quad (3)$$

The input signals in the frequency domain, $X_1(\omega)$ and $X_2(\omega)$, are obtained by

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = S(\omega) \cdot D(\omega)^{-1} \begin{bmatrix} H_r^{(2)}(\omega) & -H_1^{(2)}(\omega) \\ -H_r^{(1)}(\omega) & H_1^{(1)}(\omega) \end{bmatrix} \begin{bmatrix} H_l(\omega) \\ H_r(\omega) \end{bmatrix} \quad (4)$$

where

$$D(\omega) = \{H_1^{(1)}(\omega) \cdot H_r^{(2)}(\omega) - H_r^{(1)}(\omega) \cdot H_1^{(2)}(\omega)\}.$$

Taking the inverse Fourier transform on both sides of Eq. (4), we obtain

$$\begin{cases} x_1(t) = \{h_l(t) * h_r^{(2)}(t) - h_r(t) * h_l^{(2)}(t)\} * s(t) * d(t), \\ x_2(t) = \{h_r(t) * h_l^{(1)}(t) - h_l(t) * h_r^{(1)}(t)\} * s(t) * d(t), \end{cases} \quad (5)$$

where $d(t)$ is the inverse Fourier transform of $D(\omega)^{-1}$. Conditions for a unique solution or non-trivial solution are:

- i) $D(\omega) \neq 0$
- and
- ii) $H_{1,r}^{(1)}(\omega)$ and $H_{1,r}^{(2)}(\omega)$ have no zeros in the audio frequency range.

2.2 Subjects

In order to discuss the effects of the differences between individual HRTFs, three subjects were chosen for this investigation. All three are males, 21 years of age ± 1 year, with normal hearing sensi-

Table 1 Dimensions of pinnae in centimeter.

Dimensions ^{a)}	Subject (L)		Subject (M)		Subject (S)		Averaged (15 males)		Standard deviation	
	l	r	l	r	l	r	l	r	l	r
Ear length	7.40	7.26	6.54	6.63	4.98	5.44	6.53	6.56	0.45	0.40
Ear length above tragon	3.81	3.86	3.16	3.17	2.45	2.75	3.13	3.16	0.30	0.28
Ear breadth	3.08	3.16	3.30	3.30	2.55	2.75	3.12	3.10	0.28	0.29
Concha length	3.10	3.25	2.87	2.86	2.08	2.28	2.77	2.79	0.23	0.26
Concha length tragon to lower notch	1.96	2.03	1.71	1.73	1.42	1.46	1.77	1.77	0.15	0.16

^{a)} These dimensions are the same as those defined by Burkhard *et al.*⁴⁾

tivity.

As shown in Table 1, they have different sized pinnae; subject (L) has large-sized pinnae, subject (M) has medium-sized ones, and subject (S) has small-sized ones. The dimensions of subject (M) are almost equal to the average size of 15 males, whom we measured.

2.3 Measurement of the HRTF

In order to obtain both the amplitude and the phase parts of the HRTF accurately, the pulse technique^{5,6)} was used adding 100 repetitions. This yields an increase in the signal to noise ratio of 20 dB.

The subject was seated in an anechoic chamber, and his head was fixed. As the sound sources, thirteen loudspeakers were arranged at every 15 degrees both in the right half on the horizontal plane and in the median plane on a hemisphere with a radius of 1.5 m centered on the subject's head.

Tips of two probe-microphones (Brüel & Kjaer, Type 4170) were carefully set at the entrances of two auditory canals, that is the point of intersection of the center line of the auditory canal and the extension of the concha surface. A set of HRTFs for the paths from the sound source to the left and the right ears was measured simultaneously. A pulse of 20 μ s duration was fed through an amplifier into the loudspeaker. The probe-microphone outputs were led into a digital computer through an A-D converter (12 bit) with a sampling frequency of 50

kHz.

The HRTF, $H_{1,r}(\omega)$, are obtained

$$H_{1,r}(\omega) = \frac{G_{1,r}(\omega)}{F(\omega)},$$

where $F(\omega)$ is the Fourier transform of the wave form, $f(t)$, measured at the point corresponding to the center of the subject's head in a free field without a subject, and $G_{1,r}(\omega)$ are those measured at the two entrances of the auditory canals of the subject.

As an example, the measured amplitudes of the HRTFs in the median plane of the three subjects are shown in Fig. 2, since the amplitude of the HRTF is considered to play an important role for sound localization, especially in the median plane.⁷⁻⁹⁾ Taking note of the sharp notch-frequency (NF) in the frequency range above 4 kHz where localization cues in the median plane exist,⁷⁾ the NF at the left ear of subject (L) is not changed by the elevation angle, i.e., 9.3 ± 0.1 kHz, but the NF at the right ear of subject (L) is greatly changed in the range of 7 to 12.6 kHz. Such a high angle-dependency cannot be observed for the other subjects. The NF at the left ear of subject (M) is almost constant, i.e., 8.4 ± 0.4 kHz and the NF at the right ear of subject (M) is almost constant, i.e., 10.0 ± 0.3 kHz, except for $\theta = 165^\circ$ and 180° . Also, the NFs at the left and the right ears of subject (S) are almost constant, i.e., 10.0 ± 0.2 kHz and 10.0 ± 0.5 kHz, respectively. These differences are not solely

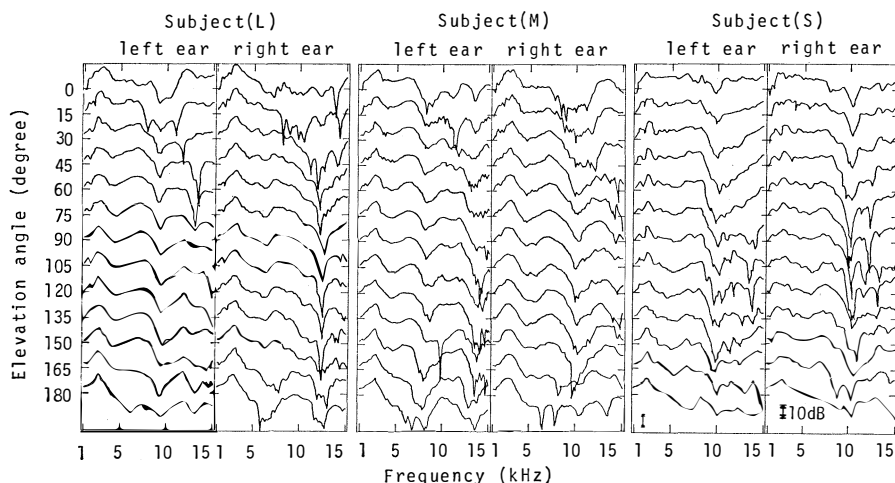


Fig. 2 Measured amplitudes of the HRTFs in the median plane of the three subjects.

related to the dimensions of pinnae as indicated in Table 1.

3. LOCALIZATION TESTS

3.1 Procedure

According to the above mentioned theory, the simulation was carried out by the aid of a digital computer, using both the amplitude and the phase parts of the measured HRTFs. White noise as a source signal ($s(t)$ in Eq. (5)) was put into a digital computer through a lowpass filter with a cutoff frequency of 13.6 kHz and an A-D converter (12 bit). After the computation, the output signals were recorded by a magnetic tape-recorder through a D-A converter (12 bit) and the lowpass filter.

Two loudspeakers, L_1 and L_2 , for reproduction were located at $(1.5 \text{ m}, +60^\circ, \pm 90^\circ)$ in the traverse plane, because the HRTFs for the paths from the positions to the left and the right ears have no remarkable notches for all subjects. Hence, the two conditions are satisfactory in obtaining the unique solution or the non-trivial solution of Eq. (5). The frequency characteristics of the two loudspeakers were flat within $\pm 5 \text{ dB}$, and they were equal to each other within $\pm 1 \text{ dB}$ in the frequency range of the stimulus. A frequency range from 0.3 to 13.6 kHz was chosen in order to also obtain a natural sound localization in the median plane.¹⁰⁾ The stimulus was of one-second duration, with the level of 50 dB (A) followed by an interval of six seconds. It was presented five times for each direction in random order.

When the subject's own HRTF was used in the simulation, the sound pressures at the entrances could be reproduced correctly, for example, as shown in Fig. 3 (A). The difference between the simulated sound pressure and a real sound pressure was within $\pm 4 \text{ dB}$ throughout this experiment and "over all patterns" as a function of frequency were similar as shown in the figure. Also, the notch frequencies of the simulated sound pressure and a real sound pressure nearly coincided. The interaural phase difference (IPD) in the low frequency range which plays an important role for sound localization in the horizontal plane could be reproduced correctly as shown in Fig. 3 (B), especially in the frequency range under 1.0 kHz. In the frequency range above 1.5 kHz, the difference between IPDs of the simulated sound source and a real sound source was also within 150 μs .

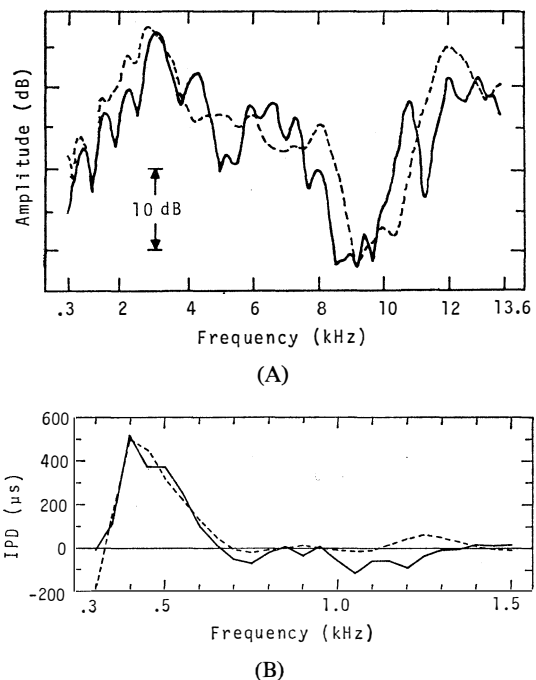


Fig. 3 Accuracy of the simulation. Sound pressures at the left ear entrance(A) and interaural phase differences(B). Solid lines indicate the simulated sound and dotted lines indicate the real sound (Subject (L), at $\theta=0^\circ$, $\psi=0^\circ$).

The subject was seated in a darkened anechoic chamber with his head fixed. Having 65 sheets of test-paper with a circle printed on it and a tiny light, he marked each sheet within an interval of six seconds, indicating his perceived direction.

Localization tests in the horizontal plane and the median plane were performed separately. Also, each subject responded to sound sources simulated by the use of his own HRTF and those simulated by the use of HRTFs of the other two subjects separately. In order to compare the accuracy of localization, the tests of real sound sources both in the horizontal and the median planes were also performed. The responses could be read with a protractor to an accuracy of one degree.

3.2 Results

In these localization tests, all subjects were able to localize all sound images at a distance from their heads.

An error of localization is defined by

$$E = \frac{1}{J} \cdot \frac{1}{K} \cdot \sum_{j=1}^J \sum_{k=1}^K |S_{jk} - R_{jk}| \quad (\text{degree}), \quad (6)$$

where S_{jk} indicates the real or the simulated direction, R_{jk} is the perceived direction, every j corresponds to a single direction ($J=13$), and every k is the stimulus for each direction ($K=5$).

Table 2 shows the error E in the localization of sound sources simulated by the use of the subject's own HRTF and real sound sources. It is found that all subjects could localize sound sources simulated

Table 2 Localization errors E (in degrees) of sound sources simulated by the use of the subjects' own HRTFs and real sound sources.

(A) In the horizontal plane.

Subject	Simulated sound source	Real sound source
L	18	8
M	12	7
S	10	9

(B) In the median plane.

Subject	Simulated sound source	Real sound source
L	15	10
M	19	19
S	14	10

Table 3 Localization errors E (in degrees) of sound sources simulated by the use of the subjects' own HRTFs and the two others' HRTFs.

(A) In the horizontal plane.

Subject	HRTF used in the simulation		
	L	M	S
L	18	36	42
M	21	12	26
S	19	9	10

(B) In the median plane.

Subject	HRTF used in the simulation		
	L	M	S
L	15	49	64
M	22	19	23
S	23	23	14

by the use of their own HRTFs with much the same errors as real sound sources. Table 3 includes errors E in the localization of sound sources simulated by the use of the HRTFs of the other subjects. It is found that the errors of subject (L) for sound sources simulated by the use of the other subjects' HRTFs are remarkably large, however, those of subject (M) and subject (S) for sound sources simulated by the use of the HRTFs of the other subjects are rather small.

A factor analysis of variance indicates that the HRTF factor has a significant effect ($p < 0.01$) on the localization error E in the simulation of localization in the median plane, but not in the horizontal plane.

In order to discuss the behaviors of the responses in details, the number of responses at every fifth degree angle are shown in Figs. 4 and 5. The area of the circle is proportional to the number of responses within a range of five degrees.

Figure 4 shows the results of localization tests in the horizontal plane. The results of real sound sources indicate that responses of all subjects almost corresponded to source directions. The results of sound sources simulated by the use of their own HRTFs indicate that all subjects could localize them with much the same accuracy as real sound sources, except for a few points indicated by subject (L) and subject (M).

On the other hand, the results of sound sources simulated by the use of the other subjects' HRTFs indicate the following: Most of the responses of subject (L) appeared in the rear, except for those to $\psi = 0^\circ$. The responses of subject (M) to sound sources simulated by the use of HRTF (L) and HRTF (S), and the responses of subject (S) to sound sources simulated by the use of HRTF (L) were for the most part correct, except for a few minor cases in which there was confusion as to whether the sources were in the front or rear. Also, the responses of subject (S) to sound sources simulated by the use of HRTF (M) were as accurate as real sound sources. After the localization tests, all subjects reported that all sound images appeared in the horizontal plane, in spite of sound sources being reproduced through loudspeakers in the traverse plane. Subject (L) was the only exception in his responses to sound sources simulated by the use of the other subjects' HRTFs.

Figure 5 shows the results of localization tests in

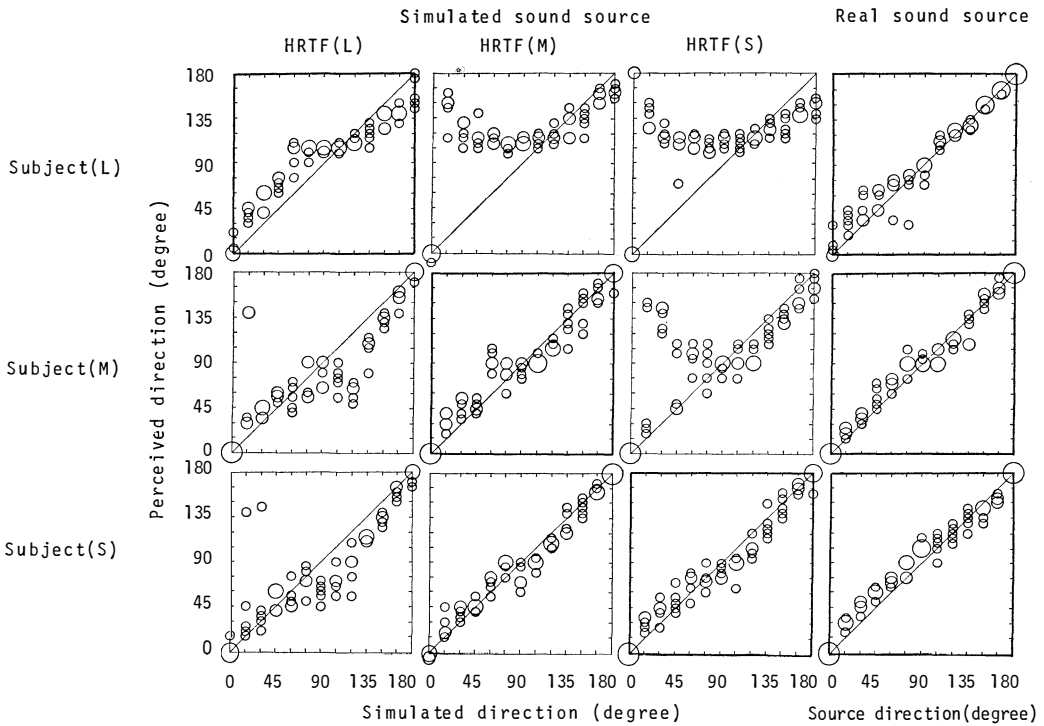


Fig. 4 Responses of the three subjects to simulated and real sound sources in the horizontal plane.

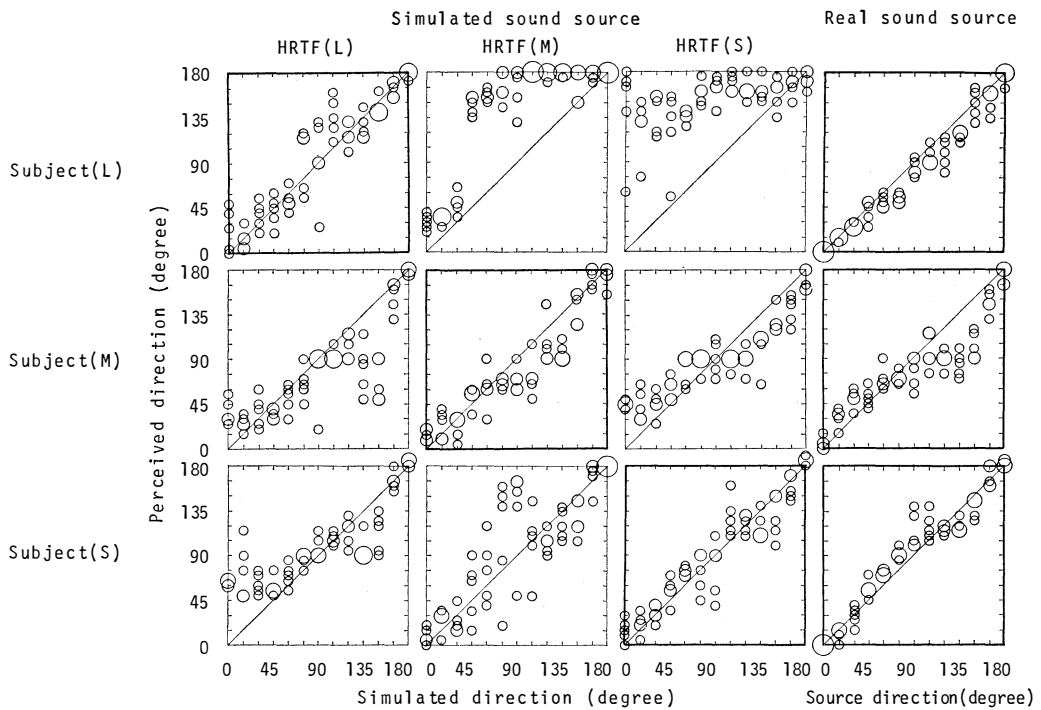


Fig. 5 Responses of the three subjects to simulated and real sound sources in the median plane.

the median plane. The results of real sound sources indicate that though the responses of subject (M) to $\theta=120^\circ$, 135° and 150° moved slightly upward, responses of all subjects almost corresponded to source directions. The results of sound sources simulated by the use of their own HRTFs indicate that all subjects could localize them with much the same accuracy as real sound sources. Only the variances of the responses of subject (L) to $\theta=0^\circ$, 75° , 90° and 105° , and those of subject (S) to $\theta=0^\circ$, 75° , 90° and 150° were somewhat larger than those of real sound sources.

On the other hand, the results of sound sources simulated by the use of the other subjects' HRTFs indicate the following: Most of the responses of subject (L) to sound sources simulated by the use of HRTF (M) appeared in the front or in the rear. Most of the responses of subject (L) to sound sources simulated by the use of HRTF (S) appeared in the rear. The responses of subject (M) and subject (S) to sound sources simulated by the use of the other subjects' HRTFs were almost correct, but their responses to the front moved upward slightly. The variances of all of their responses to all directions were larger than those of real sound sources. After the localization tests, all subjects reported that all sound images appeared in the median plane, even those of subject (L) to sound sources simulated by the use of the other subjects' HRTFs.

4. DISCUSSION

Butler *et al.*⁹⁾ have investigated the effects of the differences between individual HRTFs on the accuracy of localization in the median sagittal plane. In his experiments, each subject listened to sound sources recorded via microphones placed in his own auditory canals and those of the other three subjects, through headphones. The results of their localization tests show that good localization could be performed for sound sources recorded at the ears, when the notch frequency of the HRTF changes in an orderly way according to the elevation angle. But all subjects did not perform best localizations of sound sources recorded at their own ears.

In this paper, the results of the localization tests of simulated sound sources in the median plane, through the two-channel loudspeaker system, show that all subjects could localize sound sources simulated by the use of their own HRTFs with the same accuracy as real sound sources in the median plane,

even if the notch frequency is almost constant as in HRTF (M) and HRTF (S) (Fig. 2).

In the simulation, when a subject listens to sound sources simulated by the use of other subjects' HRTFs (for example, when the subject is (L) and the HRTF used is (M)), subject (L) does not listen to the sound with HRTF (M) as shown in Fig. 2, but to the complicated sound (Appendix I). However, there is a significant difference between the effects of individual HRTFs on the localization error E in the median plane. Comparing the results of localization tests of the three subjects (Table 3 (B)) and the sharp notch-frequency of their measured HRTFs (Fig. 2), it can be seen that subject (L) whose HRTF shows a high angle-dependency can not localize correctly sound sources simulated by the use of the low angle-dependent HRTF (M) and HRTF (S). On the contrary, subject (M) and subject (S) whose HRTFs show a low angle-dependency can localize correctly sound sources simulated by the use of the high angle-dependent HRTF (L).

5. CONCLUSIONS

The reliability of the simulation of sound localization at any position in a three dimensional space, by the aid of a digital computer through a two-channel loudspeaker system, was confirmed. The simulation was performed by the use of the measured head-related transfer functions (HRTF) of three subjects with different sized pinnae.

The results of localization tests of simulated sound sources both in the horizontal and the median planes show clearly that:

- (1) In spite of the sound sources reproduced through two loudspeakers located in the traverse plane, natural sound localization in both planes can be performed with the same accuracy as real sound sources, when the subject's own HRTF was used in the simulation.
- (2) The differences between individual HRTFs affect the accuracy of localization in the simulation. In particular, in the median plane, an HRTF with a high angle-dependency seems to be more advantageous than one with a low angle-dependency in order to obtain good performance in localization.

ACKNOWLEDGMENT

The authors wish to thank Prof. Z. Maekawa for

his useful discussions and encouragement. Also thanks to Mr. N. Johren and Mr. K. Nomachi for their cooperation in the experiments.

APPENDIX I

Suppose that subject (L) listens to sound sources simulated by the use of HRTF (M), the sound pressures at his left and right ears, $p_l^{(L)}(t)$ and $p_r^{(L)}(t)$, may be expressed by

$$\left. \begin{aligned} p_l^{(L)}(t) &= x_1^{(M)}(t) * h_l^{(1)(L)}(t) + x_2^{(M)}(t) * h_l^{(2)(L)}(t), \\ p_r^{(L)}(t) &= x_1^{(M)}(t) * h_r^{(1)(L)}(t) + x_2^{(M)}(t) * h_r^{(2)(L)}(t), \end{aligned} \right\} \quad (\text{A1})$$

where (L) and (M) mean subject (L) and subject (M), respectively,

$$\begin{aligned} x_1^{(M)}(t) &= \{h_l^{(M)}(t) * h_r^{(2)(M)}(t) \\ &\quad - h_r^{(M)}(t) * h_l^{(2)(M)}(t)\} * s(t) * d^{(M)}(t), \\ x_2^{(M)}(t) &= \{h_r^{(M)}(t) * h_l^{(1)(M)}(t) \\ &\quad - h_l^{(M)}(t) * h_r^{(1)(M)}(t)\} * s(t) * d^{(M)}(t), \end{aligned}$$

and $d^{(M)}(t)$ is the inverse Fourier transform of

$$\begin{aligned} &\{H_l^{(1)(M)}(\omega) \cdot H_r^{(2)(M)}(\omega) \\ &\quad - H_r^{(1)(M)}(\omega) \cdot H_l^{(2)(M)}(\omega)\}^{-1}. \end{aligned}$$

REFERENCES

- 1) Y. Ando, "Subjective preference in relation to objective parameters of music sound fields with a single echo," *J. Acoust. Soc. Am.* **62**, 1436-1441 (1977).
- 2) M. R. Schroeder and B. S. Atal, "Computer simulation of sound transmission in room," *IEEE Intern. Conv. Rec.* **11**, 150-155 (1963).
- 3) Y. Ando, S. Shidara, Z. Maekawa, and K. Kido, "Some basic studies on the acoustic design of room by computer," *J. Acoust. Soc. Jpn.* **29**, 151-159 (1973) (in Japanese).
- 4) M. D. Burkhard and R. M. Sachs, "Anthropometric manikin for acoustic research," *J. Acoust. Soc. Am.* **58**, 214-222 (1975).
- 5) V. Mellert, K. F. Siebrasse, and S. Mehrgardt, "Determination of the transfer function of the external ear by an impulse response measurement," *J. Acoust. Soc. Am.* **56**, 1913-1915 (1974).
- 6) S. Mehrgardt and V. Mellert, "Transformation characteristics of the external human ear," *J. Acoust. Soc. Am.* **61**, 1567-1576 (1977).
- 7) J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *J. Acoust. Soc. Am.* **56**, 1829-1834 (1974).
- 8) C. L. Searle, L. D. Braida, D. R. Cuddy, and M. F. Davis, "Binaural pinna disparity: another auditory localization cue," *J. Acoust. Soc. Am.* **57**, 448-455 (1975).
- 9) R. A. Butler and K. Belendiuk, "Spectral cues utilized in the localization of sound in the median sagittal plane," *J. Acoust. Soc. Am.* **61**, 1246-1269 (1977).
- 10) M. Morimoto and A. Saito, "Effects of frequency range of stimuli on sound localization in the median plane," unpublished.