Sound localization in headphone reproduction by simulating transfer functions from the sound source to the external ear

Jun’ichi Kawaura,* Yōiti Suzuki,** Futoshi Asano,** and Toshio Sone**

*Acoustics Laboratory, ONO SOKKI Co., Ltd.,
531–1 Hakusan-cho, Midori-ku, Yokohama, 226 Japan
**Research Institute of Electrical Communication, Tohoku University,
2–1–1 Katahira, Aoba-ku, Sendai, 980 Japan

(Received 7 December 1990)

Experiments simulating sound fields radiated from a loudspeaker by using headphones are described. Comparing the differences in characteristics between the sound reproduced through a loudspeaker and that through headphone, speaker listening conditions were simulated by a headphone listening using digital signal processing techniques on the following two aspects: the static state transfer function from the loudspeaker to the external-ear entrance of a subject (referred to as “head-related transfer function” here); and the dynamic change in interaural time difference (ITD) caused by head movement of the subject. When only the static transfer function was simulated, perceived sounds were mostly localized at the intended direction outside subject’s head, but a front-rear confusion was often observed when the intended direction was near the median plane. Furthermore, perceived distances of sounds are generally shorter than intended ones. When the dynamic changes in interaural time difference are simulated in addition to the static head-related transfer functions, however, the front-rear confusion remarkably decreased and natural sound localization was achieved.

Keywords: Localization, Lateralization, Head transfer function, Headphone reproduction, Simulation

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

1. INTRODUCTION

It is quite popular nowadays to listen to music through headphones because of the wide spread of stereo headphone sets and the prevention of neighborhood noise at night, etc. The sound reproduced through headphones, however, is somewhat different from that reproduced through loudspeakers. In particular, the sound reproduced with headphones is generally localized inside the listener’s head. This ‘in-head localization’ under headphone listening conditions could be caused by a shortage of cues for localization given to listeners through headphones.

If this is the case, then sound reproduced with headphones could be localized outside the listener’s head if enough cues were supplied.

The cues for sound localization can roughly be classified into two categories: cues related to sound pressure at the eardrum, and cues not so related. The interaural difference in sound pressure, the differential rate of sound pressure at the two ears caused by motion of a listener’s head, and the information carried by sound reflections are cues belonging to the former group. Bone-conducted sound, visual information and such are cues belonging to the latter group. The effect of the latter cues
have been investigated by some researchers. For example, Sone et al. showed that the addition of bone-conducted sound under headphone listening conditions is effective in achieving localization outside the head and in reducing front-rear confusion. Nakabayashi et al. showed that sound localization is heavily influenced by vision resulting in a coincidence of localized directions with the positions of visual images. Here we disregard factors other than hearing, however, and investigate precisely how the sound is localized under headphone-listening conditions by controlling the sound pressure given to the two ears.

Until a few decades ago, computers equipped with A/D and D/A converters, high performance signal generators and precision digital analyzers were not available. Thus, it was quite natural for researchers to investigate sound localization using simple parameters such as interaural level difference (ILD) and interaural phase difference (IPD) of simple signals such as pure tones and clicks. Sandel, Zwislocki et al. and Mills showed that the interaural phase difference is important in the low frequency range, the interaural level difference is important in the high frequency range, and the time difference is important for the localization of transient sound. Moreover, Sandel, Leakey and Toole pointed out that a sound is localized inside the listener's head when the sound is reproduced in phase with two loudspeakers located in front and at the rear of the listener.

The interaural differences are very small if the sound source is on the median plane. Therefore, cues other than interaural differences are required to explain median plane localization thoroughly. Roffler et al. and Blauert stated that particular patterns of a frequency spectrum at the ears as related to the direction of sound source were thought to be important cues in median plane localization. Moreover, such patterns of the frequency spectrum of a sound are now believed to play a very important role in its localization, especially for sounds with wide frequency spectrums which we often hear in our daily life.

Nowadays, these ideas of cues can be explained with the notion of the Head Transfer Function (or Head-Related Transfer Function, abbreviated HTF hereafter), which is the transfer function of the sound path from the sound source to the ear of the listener. The HTF's involve cues such as interaural differences and patterns of the frequency spectrum at the eardrum. Therefore, many researchers have investigated the directional dependencies of HTF's. Thus, the localization of sound could be controlled by simulating HTF's as HTF's are thought to contain all information required for sound localization. Schroeder et al. developed the basic theory for simulating HTF's and Ando et al. refined this formulation.

Laws and Blauert et al. reported that they could localize a sound outside the listener's head with headphone reproduction by simulating HTF's. However, these studies are in question with regards to the accuracy of the simulation of the transfer functions because of the analog filters used in them. Morimoto et al., on the other hand, conducted the simulation procedure with digital signal processing and reported that sound could be localized at any direction with two loudspeakers. They also reported that control of the perceived distance is very difficult for a frontal image, and even for a lateral image, if the distance is over 1 meter.

The studies mentioned above were concerned only with the simulation of static HTF's, disregarding movement of the listener's head. However, changes in sound pressure and tonal quality at the listener's ears produced by movement of the listener's head may be other important cues for sound localization. Thus dynamic variation of HTF's should also be taken into consideration.

For example, de Boer et al. considered the dynamic effect to be specially important when the interaural level difference is very small, such as in median plane localization. To confirm this idea, they reproduced a sound collected with a spherical dummyhead through headphones and asked subjects to move their head synchronously with the movement of the spherical dummyhead. They reported that the perceived direction coincided with the actual direction of the sound source when the subjects rotated their heads in the same direction as that of the dummyhead, while front-rear judgment is often reversed when the subjects rotated their heads in the opposite direction. Boerger et al. reported on the error rate of front-rear confusion for a sound recorded with a dummyhead. The error rates were significantly reduced when an interaural time difference (ITD) corresponding to the rotation angle of the subject's head was added to the signal. Leakey proposed a model using differential rates...
of ITD caused by head rotation in order to estimate the elevation angle of the sound source and to determine whether the sound is in front or behind. Thurlow et al. investigated the effect of spontaneous and forced movement of the head on sound localization in listening to an actual sound source. They reported that the error of front-rear judgment greatly decreased when the subject were allowed to move their heads freely. They also reported that the error also decreased when the subjects’ heads were forcibly rotated or pivoted, and that the effect of head rotation was far greater than that of pivoting.

Asahi et al. synthesized the sound image of a frontal incident by simulating HTF’s using digital signal processing techniques. They reported that the addition of an ITD corresponding to the angle of head rotation decreased the error of front-rear judgment and the rate of in-head localization.

If sound localization is dependent on the sound pressure at the eardrums, sound can be localized at any desired location by simulating HTF’s. In such simulations, whether the sound is reproduced by headphones or by loudspeakers would have no effect on sound localization so long as both the static and dynamic HTF’s were simulated accurately enough. In particular, consideration of the dynamic aspects of the transfer functions is very important in a simulation using headphones, where the reproduced sound images are apt to be localized inside the head. This is because in headphone reproduction, all sound images outside the head rotate along with head movement, an extremely irrational phenomenon which never occurs under non-headphone conditions.

In this study, we first investigated to what extent sound images on the horizontal plane can be controlled when only static HTF’s are simulated with digital signal processing techniques using 512-point FIR filters with a resolution of 16 bits and sampling frequency of 50 kHz.

Next, we approximated the dynamic changes in HTF’s induced by head movement and examined the effect of these changes. We considered only head rotation because we wished to focus on sound localization on the horizontal plane and because of the above-mentioned results of Thurlow et al. Moreover, when the rotation angle is small, the change in phase (or group delay) is more prominent than that in the power spectrum of the HTF’s. Thus, we simulated the change in HTF’s only by the change in group delay, while the power spectrum was kept constant.

In the following section, we will describe the method used to control sound localization based on the concepts mentioned above and then examined the results of psychoacoustic experiments in which perceived location of the simulated sound images on the horizontal plane were examined.

2. SIMULATION OF SOUND IMAGES ON THE HORIZONTAL PLANE THROUGH HEADPHONES

2.1 Method of Simulating Static HTF’s

Conditions existing when a loudspeaker is placed on the horizontal plane were simulated using headphones in this study. We simulated the sound image with a loudspeaker rather than an actual sound source because of the following reasons: (1) Headphones and loudspeakers can be treated equivalently by considering the transfer function of the path from the electrical input of the transducers to the listener’s ear; (2) Application of the technique to audio-equipments was thought to be easier. Here the transfer function of the path from the electric input terminal of the loudspeaker to the output terminal of the microphone at the eardrum of the right and left ear of the listeners are denoted by \( H_L(j\omega) \) and \( H_R(j\omega) \), respectively. Moreover, the transfer function of the path from the electric input terminal of the headphone worn by the listener to the output terminal of the microphones located at the eardrum of the right and the left ear of the listener are expressed by \( G_L(j\omega) \) and \( G_R(j\omega) \), respectively. Then if the source signal, \( S(j\omega) \), is reproduced with the loudspeaker, the sound pressure at the listener’s ears are given by

\[
P_L(j\omega) = H_L(j\omega) \cdot S(j\omega),
\]

\[
P_R(j\omega) = H_R(j\omega) \cdot S(j\omega).
\]

If source signals \( S_L(j\omega) \) and \( S_R(j\omega) \) are fed to the right and left headphones worn by the listener, on the other hand, the sound pressures at the eardrum of the left and right ear of the listener are given by

\[
P_L^{HF}(j\omega) = G_L(j\omega) \cdot S_L(j\omega),
\]

\[
P_R^{HF}(j\omega) = G_R(j\omega) \cdot S_R(j\omega).
\]

Thus, the sound pressure at the listener’s eardrums produced with the loudspeaker and the headphones will be the same if the following equations hold:
where $G_L^{-1}(j\omega)$ and $G_R^{-1}(j\omega)$ denote the inverses of $G_L(j\omega)$ and $G_R(j\omega)$, respectively.

Cues for horizontal sound localization mentioned in the previous section such as interaural differences and frequency responses of the ears are integrated into the transfer functions of the paths from an electric input terminal of a loudspeaker to the output terminals of microphones at both eardrums, $H_L(j\omega)$ and $H_R(j\omega)$. On the other hand, transfer functions of microphones and other equipment used in the measurement are automatically canceled because they are equally included in the transfer functions of both the speaker and the headphones.

To realize Eq. (3), we measured the transfer functions related to the loudspeaker, $H_{LBR}(j\omega)$, and those related to the headphones, $G_{LBR}(j\omega)$, by using a time-stretched pulse.\textsuperscript{25,26} We did not measure the sound pressure at the eardrum but at the external ear entrance for safety in measurement. Wiener et al. showed that the transfer function of the ear canal is only slightly dependent on direction.\textsuperscript{27} Moreover, Arai et al. showed that there is no great difference between the transfer functions of the ear canal with and without a headphone.\textsuperscript{28} If so, the transfer function of the sound path from the external ear entrance to the eardrum can be treated as a constant. Then the substitution of the measurement at the external ear entrance for that at the eardrum is a reasonable approximation because the transfer function of the ear canal is canceled in Eq. (3).

In the measurement of the transfer functions, synchronized averages were taken for 100 pulses. The sampling frequency was 50 kHz and the resolution of quantization was 16 bits. Then, impulse responses were calculated for 512 points. The impulse responses were used as the coefficients of the FIR digital filters. Moreover, inverse characteristics were calculated by use of the time-domain least-mean-squares method, which is often referred to as "deconvolution." More detailed explanation of the measurement and the method of simulation is given in our previous paper.\textsuperscript{29}

Figure 1 shows the maximum value of the cross-correlation between the impulse response of the sound path from the loudspeaker to the external ear entrance of the subject and the corresponding impulse response simulated by use of a digital filter and a headphone. The value would be 1.0 if the simulation is accurate. The actual values, however, are less than 1.0 as shown in the figure because of errors in measurement and calculation. Though it is unknown whether cross-correlation is an appropriate measure for evaluating the degree of precision of the simulation system, it could at least be a candidate for such measurement. In our previous paper,\textsuperscript{30} we showed an example of a simulated response and an original response when the correlation was 0.85.

Figure 1 shows that the simulation for the right ear was well-achieved in any direction, while that for the left ear becomes worse when the loudspeaker was located at the right side, i.e., at the opposite side of the ear. Though it is difficult to know the causes of this deterioration, the following speculations are possible: First, the degree of precision of the measurement of the impulse response of the sound path from the loudspeaker to the external ear entrance cannot be the cause of deterioration because the impulse response is also involved in the simulation system as a common component. Second, other possible causes include such things as accuracy of measurement of the impulse response of the headphone, the estimation of its inverse, and the deviation of the point of measurement (the tip of the probe microphone) from the expected position during measurement of the impulse response of the headphones and the loudspeaker. It is difficult to achieve further accuracy in the calculation of im-
pulse responses and its inverse. On the other hand, it would be possible to reduce the deviation of the point of measurement by using more sophisticated guides to position the probe microphone.

2.2 Simulation of the Change in HTF Caused by Subtle Head Rotation

Let \( H(j\omega, \theta) \) be the transfer function of the sound path from a loudspeaker located at azimuthal direction \( \theta \) on the horizontal plane to just in front of the eardrum of the listener. This transfer function can be written as

\[
H(j\omega, \theta) = H_0(j\omega, \theta) \exp(-j\omega(l/c)),
\]

where \( H_0(j\omega, \theta) \) is independent of the distance \( l \) between the loudspeaker and the listener, and \( \exp(-j\omega(l/c)) \) represents the time delay caused by distance \( l \). Then a small change in the transfer function \( \Delta H \) caused by a slight rotation of the head is denoted as

\[
\Delta H = \frac{\partial(H_0 \exp(-j\omega(l/c)))}{\partial \theta}.
\]

While the partial differential of \( H_0 \) in \( \theta \) could be expected to be almost zero, the time delay greatly changes since the distance between the loudspeaker and the ears of a listener changes. Thus \( \Delta H \) could be approximated using

\[
\Delta H = \frac{H_0 \partial \exp(-j\omega(l/c))}{\partial \theta}.
\]

This means that the simulation of the change in transfer function caused by slight head rotation can be tolerably achieved by compensating for only the change in delay time.

Accordingly, we measured and calculated the interaural time difference (ITD) when the azimuthal direction of the sound source is changed with the head fixed. The measured ITD was determined based on the time lag which maximizes the cross-correlation function of the impulse response from the sound source to both ears. The calculated ITD was derived from a geometrical distance from the sound source to ears under the assumption that the section of the head is a circle. Figure 2 shows the results. This figure can also be interpreted as showing the ITD's when the subject rotates his head while the sound source is fixed. This interpretation leads us to the conclusion that the change in ITD is approximately proportional to the rotation angle of the head when the sound source is located near the front or near the rear of the subject. The inclination, i.e., the rate of change of ITD vs. rotation angle of the head, is dependent on its diameter and shape. Thus the rate is greatly dependent on the subjects. Here we used an averaged rate of 6.7 \( \mu \)s/degree in the simulation.

A problem in the simulation is that the sign of the rate is different for a frontal sound source and a rear sound source. As Boeger\(^{21} \) and Asahi\(^{24} \) fixed the rate of change of ITD vs. rotation angle, inconsistency in the simulation would have occurred if they had examined both frontal and rear sound sources. The sign of the rate of change must at least be converted depending on whether the sound source is in front or at the rear.

Taking the above discussion into consideration, we fabricated the experimental equipment shown in Fig. 3. The head of the subject is fixed with an apparatus attached to the chair so that only rotation is possible. The direction of the head is linearly converted into electric potential by a potentiometer installed in the apparatus. Furthermore, the potential is converted into two clock signals by analog circuits and V/F converters. The period of a clock signal, \( f_{c1} \), is proportional to the direction...
while the period of the other clock signal, \( f_{c2} \), is inversely proportional to the direction. As the delay circuits inserted in the paths of the headphone signals cause delays proportional to the period of the input clocks, the delay time of the \( f_{c1} \) channel increases while the delay time of the \( f_{c2} \) channel decreases as the head rotates.

A switch, \( S_{w1} \), alters the relation between the direction of head rotation and the decrease or increase in delay time of the channels. If the switch is set to 'front mode,' the delay time of the left channel decreases while the delay time of the right channel increases as the head rotates clockwise. Therefore, the 'front mode' gives the subject ITD information so that he localizes the sound source located at the front. On the other hand, if the relation is altered by setting the switch to 'rear mode,' ITD information is given to the subject so that he judges the sound source as being placed at the rear.

It is impossible for the equipment to simulate the change in ITD if the sound source locates sideways, because the change in delay time is assumed to have a linear relation with the rotation angle. However, it is possible to simulate the change in ITD as long as the simulated sound source is near the median plane.

3. EXPERIMENT ON HORIZONTAL LOCALIZATION OF SIMULATED SOUND IMAGES

We conducted two experiments using the simulation method outlined in the above section: we simulated only static HTF's in the first experiment and then added the simulation of the dynamic change in group delay of the HTF's caused by the rotation of the subject's head in the second experiment.

Experimental conditions common to the two experiments are as follows: The subjects were three young male adults with normal hearing acuity. The directions of the simulated loudspeakers were twelve: 0 degree (in front) and every 30 degrees. The distance between the simulated loudspeakers to the center of the subject's head was 1.5 m. The height of the chair was so adjusted that the height of the listener's ears coincided with the center of the loudspeakers. Transfer functions of the path from the input of the loudspeakers and the headphones to the probe microphone located at the external ear entrances were measured both under loudspeaker- and headphone-listening conditions. We used jazz music and pink noise as source signals.

3.1 Horizontal Localization by Simulation of Static HTF

In this experiment, we produced a stimulus sound using the HTF’s of all three subjects with two source signals. Subjects listened to all the stimuli simulated by using the three subjects' HTF’s. Subjects listened to each stimulus three times. Thus, the total number of stimuli was 108 for each source signal, i.e., a combination of twelve directions, three pairs of HTF’s for the three subjects, and three repetitions. The stimuli were presented to the subject with headphones in random order. Each stimulus lasted 30 s. The subjects were requested to keep their eyes closed during the presentation and to write the perceived position of the sound stimuli on a sheet of paper after the presentation of each stimulus. A small headrest was used to support the subject's head.

Subjects were asked to list the perceived direction, distance and elevation of the sound images with the following instructions:

**Direction:**
Please indicate direction in degree, 0 degrees
indicating front and 90 degrees indicating the right side.

Distance:
Please rate the distance of sound from 1–6 according to the following scale.
1. near the center of your head
2. near the skin of the head but inside the head
3. outside the head but near the head
4. distance just within your reach
5. near the simulated distance (1.5 m)
6. farther than the simulated distance

Elevation:
Please rate the elevation of sound from 1–3 according to the following scale.
1. near the horizontal plane
2. slightly elevated
3. greatly elevated

Figures 4(a)–(d) show the experimental results for all the subjects. These figures show the results for jazz music; the results for pink noise show a very similar tendency.

Figures 4(a) and (b) show the perceived directions of the sound image. In the figures, responses within 15 degrees were combined for every 15 degrees. Figures 4(a) and (c) show the results when the subjects' own HTF's were used in the simulation, while Figures 4(b) and (d) show those when others' HTF's were used in the simulation. The simulated sound images for the loudspeaker were localized at almost all directions as intended through headphones. However, the subjects often perceived images in front of them as being located at rear positions symmetrical with respect to the line connecting the ears. Comparing Fig. 4(a) with 4(b), we see that whether their own HTF's or others' were used induced only small differences.

Figures 4(c) and (d) show the results of perceived distance and elevation. The figures show that most sound images simulating the loudspeaker by headphones were localized outside the subjects' heads. However, it seems that the perceived distance decreased and the perceived elevation decreases...
ed when other's HTF's were used instead of their own in the simulation. Moreover, irrespective of whether the HTF used was their own or not, decreases in perceived distance and increases in perceived elevation of the images were often observed for sound images near the median plane.

3.2 Horizontal Localization by Simulation of HTF Including Dynamic Change Induced by Slight Movement of the Subjects' Head

We conducted an experiment in which dynamic changes of group delay in HTF induced by head movement as well as the static HTF were simulated.

Fig. 5 Experimental results of sound localization through headphone reproduction using the transfer functions from the sound source to the subject's ear entrances. The sound sources simulated are loudspeakers distributed around the listener in an anechoic room. ITD was given in accordance with the rotation of the subject's head. Filled circles indicate perceived positions of sound images. The area of the circles is proportional to the frequency of the subjects' responses. Figures at the top indicate the average perceived elevations.
In the experiment, HTF's of only one of the three subjects, *i.e.*, those of subject 2, were used. Only jazz music was used as source signal.

When the direction of the simulated sound images is lateral, the effect of head movement is small and the relation between changes in delay time and angles of head rotation is non-linear. We simulated, therefore, only sound images whose direction is 0, ±30, ±150 and 180 degrees. The simulated sounds were repeated three times as stimuli. The stimuli for the six directions were presented four times each in random order via headphones. In each session, the mode was fixed at either ‘front’ or ‘rear.’ Thus, each subject listened to 24 stimuli (6 directions times 4 repetitions) per session. The subjects judged the perceived direction, distance and elevation as in the previous experiment. They were requested to close their eyes and not to rotate their heads more than ±30 degrees while the sound was presented, and to respond after the perceived position of the sound is established.

Figures 5(a)–(c) show the responses of the three subjects. In the figures, the upper two rows correspond to the ‘front mode’ while the lower two rows, the ‘rear mode.’ For each of the two rows, the upper one shows results for sound images simulated by the use of HTF’s of frontal incidence while the lower one shows results for rear HTF’s. The numbers at the lower right and the dashed lines radiating from the center of the head in each panel show the direction of the static HTF used in the simulation. Filled circles show the perceived directions and distances of the sound images. The area of the circles is proportional to the frequency of subjects’ judgments. Moreover, the number at the upper right in each panel shows the simple average of the subjects’ responses for the elevation angle of the sound images. As stated in the previous section, “1” represents a point near the horizontal plane, “2” indicates a slightly elevated point, and “3” represents a greatly elevated point.

In this experiment, two cues for a sound image in front or at the back were given to the subject: one is static information from static HTF’s and the other is dynamic information from dynamic changes in HTF’s arising from head rotation. The two cues could be set independently, *i.e.*, the highest and the lowest rows in Fig. 5 show the results when cues were consistent, while the second and third rows show those when two were inconsistent. The results of the experiment in which the two cues were consistent show smaller error in front-back confusion and lesser elevation of the localization than those when only static HTF’s were simulated.

On the other hand, if the two cues were inconsistent, in-head localization often occurred and the elevation of the sound image was larger for all the subjects. The subjects reported that the sound image was often localized as if stuck to the head. The perceived direction of the sound image was different among the subjects. The localization was determined by dynamic cue for subjects 1 and 3, but was always at the rear for subject 2.

4. DISCUSSION

It was shown in the experiments that the sound image of a loudspeaker simulated through headphones can be localized in almost any direction on the horizontal plane though front-back confusion often occurs when the simulated direction is near the median plane. Moreover, comparing Fig. 4(a) with (b), we see that whether the HTF’s used in the simulation were their own or not did not have much influence on the perceived direction.

A similar experiment using analog filters to simulate HTF’s was reported by Laws[16] and by Blauert et al.[17] They could localize a sound outside a listener’s head with headphone reproduction by simulating HTF’s. Their studies, however, focussed only on the control of the distance of the sound images and did not consider the direction of the localization at all. Moreover, they reported that the sound image is often localized outside the subjects’ heads with headphone reproduction without any signal processing. Thus it is difficult to directly compare the performance of our simulation with theirs.

Examining Figs. 4(c) and (d), which show the results of perceived distance and elevation, we see that almost all the sound images were localized outside the subjects’ head. We also see, however, that the decrease in the perceived distance and the elevated localization were often observed when other’s HTF’s instead of their own were used and when the simulated direction was near the median plane even if the HTF used was the subject’s own. Such a decrease in the perceived distance and elevation of images is often reported in experiments on binaural reproduction using a dummyhead and
headphones. It is usually stated that this phenomenon is caused by some discrepancies between the simulated HTF's and the subject's HTF's.

Morimoto et al., who simulated HTF's with two loudspeakers, also reported front-rear confusion and elevation of images. Moreover, they reported that the control of distance was very difficult for frontal incidence and for distances beyond 1 m, even for lateral incidence.

It is probable that the phenomenon stated above is mainly caused by some differences between the simulated and the subject's HTF's. However, some quantitative indices to evaluate the discrepancies among HTF's rather than whether the HTF's used are the subject's own or not should be introduced in order to examine the effect of the difference of the HTF's on simulation performance. The index must be a function of the original and the simulated HTF's. This function should connect the perceptual difference reported by the subjects with the physical difference of transfer functions. It is quite difficult, however, to devise such a function at present. Thus, here we base our considerations on a correlation coefficient of the impulse responses, sometimes used to explain physical similarity between transfer functions.

The maximum values of the cross-correlation function of the impulse responses of the HTF's between two of the three subjects are shown in Fig. 6 as a function of the direction of the sound source. The values shown in this figure are averages of the maximum values for the right and left ears. The cross-correlation of the HTF's is high between Subjects 2 and 3 for all directions, while that between Subjects 1 and 2 and that between Subjects 1 and 3 are relatively low and especially low for the frontal direction. If the cross-correlations of transfer functions and the similarities among them are closely related in a sense, possible causes of the fall of the cross-correlation in the front position might be as follows: First, the precision of the measurement of HTF's or of their simulation for Subject 1 for frontal incidence might have deteriorated. However, Fig. 1 suggests that such deterioration was more likely for the range of 90–150 degrees. Thus this reason seems to be improbable. Second, the HTF's of Subjects 2 and 3 might have been similar and those of Subject 1 much different from those of Subjects 2 and 3. If so, it is very interesting that the cross-correlation dips at the frontal incidence. The individual variation of HTF's might be generally great for frontal incidence.

Figure 7 shows an example of the effect of the HTF's used in the simulation. In this figure, the experimental results of Subject 2 are shown for different HTF's used in the experiment. The results shown in the figure are part of those shown in Fig. 4. Figures 7(a), (b), and (c) show the perceived directions when the HTF's of Subjects 1, 2 and 3 were used in the simulation, respectively. This figure is drawn in a slightly different manner from that in Fig. 4. Here the case of in-head localization is indicated separately from other cases. Figures 7 (d), (e) and (f) show the perceived distances and elevations when the HTF's of Subjects 1, 2 and 3 were used, respectively. These figures are drawn in the same manner as Fig. 4. This shows that localization was very good when the HTF's of Subjects 2 and 3 were used in the simulation. Front-rear judgment error and elevation of perceived sound images seldom occurred even for frontal incidence. On the contrary, when the HTF's of subject 1 were used, elevation of perceived sound images was often observed through the perceived directions were as accurate as in the former cases. Taking the above consideration as well as the cross-correlation of the impulse responses of HTF's (Fig. 6) into consideration, we could speculate that the cross-correlation of HTF's or the similarity of HTF's has a certain relationship with the quality of the sound localization of simulated sound images.

Next, we consider the effect of head rotation. If the two kinds of information concerning front-rear
judgment, one from static transformation of the frequency spectrum of the sound source by the HTF's and the other from dynamic changes of the HTF's induced by head rotation, are coincidentally given to a subject, the rate of front-rear confusion and frequency of elevation of the sound image are far smaller than when only static information is given to the subject. This result indicates that information from static transfer function is not sufficient and that the information due to the change in sound pressure levels at both ears induced by head motion is needed in order to localize a sound image in front or at the rear of a subject correctly. De Boer et al. examined the effect of head rotation using a spherical dummyhead. They reproduced a sound collected with the spherical dummyhead through headphones and asked subjects to move their heads synchronously with the movement of the spherical dummyhead. They reported that the perceived direction coincided with the actual direction of the sound source when the subjects were asked to rotate their heads in the same direction as that of the dummyhead, while the front-rear judgment was often reversed when the subjects rotated their heads in the opposite direction. Thurlow et al. investigated the effect of spontaneous and forced motion of a head in listening to an actual sound source. They reported that the error of front-rear judgment greatly decreased when subjects moved, especially rotated, their heads. Boerger reported the error rate of front-rear confusion for sounds recorded with a dummyhead. The error rates significantly decreased when ITD (interaural time difference) corresponding to the rotation angle of the subjects' head was added to the reproduction of the sound. Though they did not simulate a subject's HTF's as strictly as we did, but rather substituted the subject's HTF's for the HTF's of the dummyhead, their results also showed that consideration of the dynamic change of the transfer functions was effective for improving the performance of the simulation. Moreover, Asahi et al. simulated static HTF's for a frontal sound image using digital signal processing techniques and ITD by an analog circuit; they re-
ported similar results for the frontal sound image. 24)

The above results show that the changes in sound pressure at the ears induced by head motion should be taken into consideration to achieve 3D-stereo reproduction using headphones whether a technique similar to ours or a binaural technique using a dummyhead is applied. Otherwise, it would seem to be quite difficult to avoid front-rear confusion and elevation of perceived images.

Subject 3 showed higher elevation of the sound images than the other two subjects. This might be attributable to the size of the subjects' heads, Subject 3's head being larger than the others. Though this leads to a larger rate of ITD vs. rotation angle (Fig. 2), an averaged rate was used in the experiment. The rate is small if the sound source is apart from the horizontal plane. It is natural that Subject 3 localized sound images at the upper position because the ITD was smaller than what he experiences in his daily life.

When the front-rear information from static HTF's and those from the dynamic change in HTF's induced by head rotations were inconsistently given to a subject, in-head localization often occurred, the elevation of the sound image became greater for all subjects, and the sound image was often localized as if stuck to the head. It is interesting that the way the perceived incidence was judged was different among the subjects under this condition. For Subjects 1 and 3, the perceived direction was almost entirely based on dynamic information, while always rear for Subject 2. This phenomenon might have been caused by the fact that the HTF's used in the simulation were those of Subject 2. Thus dynamic information not real for all the subjects would be relatively more important for Subjects 1 and 3 because the HTF's used in the experiment were not their own. On the other hand, static and dynamic cues would be competitive for Subject 2.

The experimental results discussed above suggest that the information on the position of the sound source derived from the changes in the transfer functions is at least as important as information derived from the static transfer functions. Recently, static HTF, whether provided by a dummyhead or by digital signal processing techniques, is often treated as the main cue to control sound localization. The present study shows, however, that the influence of head motion should be further examined.

5. CONCLUSION

We proposed a method of localizing a sound image on the horizontal plane at any desired location through headphones by simulating loudspeaker listening conditions. Moreover, we examined the applicability of this method using psychoacoustical experiments.

Sound localization on the horizontal plane could be achieved to some degree by simulating only static HTF's of the sound path from the sound source to external ear entrances. Under this condition, however, decrement in perceived distance, elevation of sound images, and in-head localization were sometimes observed.

Next, we simulated both dynamic changes of HTF's induced by rotation of the head approximately as a change in ITD and static transfer functions. The results of the experiment for this kind of simulation show no front-back confusion, small elevation of the localization, and low frequency of in-head localization.

In the experiment, dynamic cues are given to the subjects either consistent with or inconsistent with the static cues. From the comparison of the results for the two conditions above, it is suggested that the amount of positional information of sound source involved in the dynamic change of HTF may weigh at least as much as those involved in the static HTF.

It seems important in the future to clarify the limit of the sound localization control by simulating the sound pressure at the eardrums. Furthermore, additional effects of bone-conduction and/or visual images should be further investigated.

ACKNOWLEDGEMENTS

The latter part of this study was supported by the Grant-in-Aid for the Scientific Research No. 61460146. The authors wish to thank Mr. Jiro Soma (Matsushita Communication Co.) and Prof. Toshiyuki Gotoh (Shizuoka University) for their helpful comments on the experimental results. It should be also noted that almost all the experiments mentioned in this paper were conducted while one of the authors, Jun'ichi Kawaura, was a graduate student at Tohoku University.

REFERENCES

1) T. Sone, M. Ebata, and T. Nimura, "On the difference between localization and lateralization,"
J. KAWAURA et al.: SOUND LOCALIZATION IN HEADPHONE REPRODUCTION

6th I. C. A. A-3-6 (1968).


Jun’ichi Kawaura was born in Hokkaido, Japan on May 21, 1961. He received the B.E. and M.E. degrees in electrical communication engineering from Tohoku University, Sendai, Japan in 1984 and 1986, respectively. He has been working at the ONO SOKKI Co., Ltd., Acoustics Laboratory from 1986.

Yōiti Suzuki was born on 11th January 1954 in Japan. He received the B. Eng. degree in electrical engineering from Tohoku University, Sendai, Japan in 1976. He received the M. Eng. and Dr. Eng. degrees in electrical and communication engineering from Tohoku University in 1978.
and 1981, respectively. From 1981 to 1987 he was at Research Institute of Electrical Communication, Tohoku University as a research associate. From 1987 to 1989 he was an associate professor at the Computer Center, Tohoku University. He is currently an associate professor at the Research Institute of Electrical Communication, Tohoku University. His research interests include loudness, sound localization, timbre, hearing aids and digital signal processing of a sound. Dr. Suzuki is a member of the Acoustical Society of Japan, the Institute of Noise Control Engineering of Japan, the Institute of Electronics, Information and Communication Engineers, Japan, and the IEEE. He received the Awaya Kiyosi Award from the Acoustical Society of Japan in 1986.

Futoshi Asano was born in Fuku-shima, Japan on October 13, 1962. He received the B.E. degree in electrical engineering, M.E. and Dr. E. degrees in electrical and communication engineering from Tohoku University in 1986, 1988, 1991, respectively. Since 1991, he has been a research associate at Research Institute of Electrical Communication, Tohoku University.

Toshio Sone was born on 14 May 1935. A graduate in electrical engineering at Tohoku University, Japan in 1958, Sone did his graduate work at the same university, where he was awarded his Ph. D in electrical and communication engineering in 1963. He joined the Faculty of Engineering, Tohoku University as a research associate in 1963, becoming an associate professor in 1964 and a full professor in 1979. He is now a professor of the Research Institute of Electrical Communication, Tohoku University. He has been engaged in researches on psychological acoustics, electroacoustics, room acoustics and noise control over thirty years, and is a fellow member of the Acoustical Society of America, a corresponding member of the Institute of Noise Control Engineering of the USA, Vice-President of the Acoustical Society of Japan, Vice-President of the Institute of Noise Control Engineering of Japan, a member of some other academic societies. He was a member of the Board of Directors of the International Institute of Noise Control Engineering from 1985 to 1990 and also acts as Secretary-General of the Western Pacific Commission for Acoustics from 1989 to 1991.