Applied Acoustics **38** (1993) 303–317

Basic Study on Sound Field Simulation Based on Running Interaural Cross-Correlation

K. Iida* & M. Morimoto

Environmental Acoustics Laboratory, Faculty of Engineering, Kobe University, Rokko, Nada, Kobe 657 Japan

(Received 10 December 1991; revised version received 18 February 1992; accepted 22 February 1992)

ABSTRACT

The purpose of sound field simulation is to make a listener perceive the same characteristics of sound image as those of sound image which he perceives in an original (to be simulated) sound field, in any other space. This paper investigates the possibility of simulating them based on the running iteraural cross-correlation (RCC) model, which is regarded as one of mechanisms in binaural signal processing. Basic experiments of dissimilarity judgement and some subjective evaluation concerning sound image are carried out between an original sound field and simulated ones with different durations of temporal window. The results show that the sound field based on the RCC model can simulate a sound image close to the original one, although they are different from each other in a physical aspect.

1 INTRODUCTION

The purpose of sound field simulation is to make a listener perceive the same characteristics of sound image as those of a sound image which he perceives in an original (to be simulated) sound field. A sound image is defined as an

* Present address: Matsushita Communication Industrial Co., Ltd, Saedo, Midori, Yokohama 226, Japan.

auditory event which a listener perceives in hearing a sound stimulus. A listener perceives three attributes of a sound image in a sound field. They are the temporal (reverberant, rhythm, durability, etc.), the spatial (direction, distance, broadening, etc.) and the qualitative (loudness, timbre, pitch, etc.) attributes. It is clear that this purpose can be achieved by simulating temporal, spatial and spectral structures of the original sound field accurately.

Here, let us consider the psychological evaluation system of a sound field. Figure 1 shows the evaluation system. The acoustic signal radiated from a sound source is affected by the room transfer function and arrives at the position of a listener. Then this signal is affected by the head-related transfer function and arrives at the entrances of both ears of a listener as the input signals to the auditory organ. These signals are processed by the auditory system, and then the listener perceives a sound image which has various characteristics, such as loudness, timbre, broadening and so on. Then the listener has his overall emotion response to the sound with weighting subjective to each characteristic.

According to this evaluation system, three approaches for the sound field simulation can be derived:

(1) simulation of room impulse response,
(2) simulation of binaural impulse response, and
(3) simulation of characteristics of sound image.

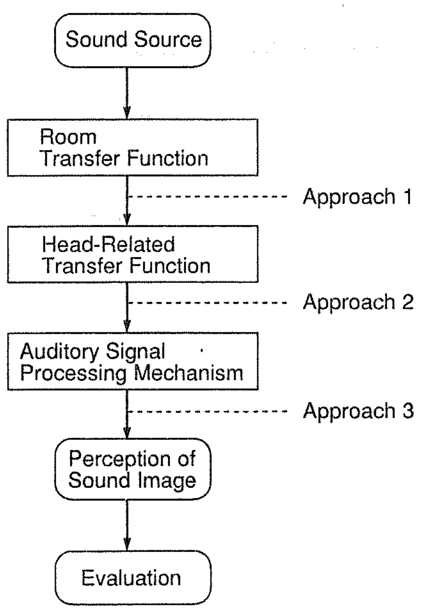

Fig. 1. Evaluation process of a sound field and three approaches of sound field simulation.

The target of the first approach is to simulate the transfer function between a sound source and a listening point in an original sound field. Most of the existing sound field simulation are of this type.[1,2]

The target of the second approach is to simulate the transfer functions between a sound source and both ear entrances of a listener.[3] TRADIS (True Reproduction of All Directional Information by Stereophony),[4] OSS (Ortho stereophonic system),[5] partition-stereophony[6] and the binaural simulation[7] are well known types of this simulation.

In these cases, the time delay, the amplitude and the direction of each reflection must be equal to the original one, as a rule. On the other hand, the target of the third approach is to simulate the output of the binaural signal processing mechanism. Namely, the simulation is considered to be achieved if the characteristics of sound image perceived in the simulated sound field are the same as those perceived in the original sound field, even if the simulated sound field is not physically equal to the original one. To the authors' knowledge there is no investigation on this type of simulation.

This paper investigates the possibility of the third type of the sound field simulation. The suggested method is based on the running interaural cross-correlation (RCC) model, which is regarded as a mechanism of spatial hearing, because the authors main interest is the simulation of spatial attributes of sound images.

2 SIMULATION OF A SOUND FIELD BASED ON BINAURAL SIGNAL PROCESSING MODEL

2.1 Binaural signal processing model

A lot of models of binaural signal processing mechanisms have been reported. Figure 2 shows a binaural signal processing model[8–10] which has been developed to explain the perception of spatial attributes of a sound image.

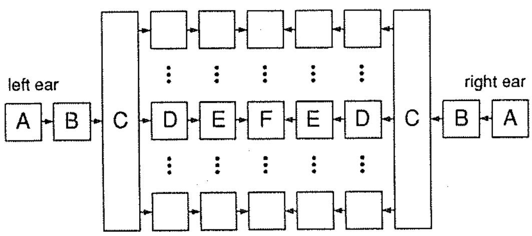

Fig. 2. A model of binaural signal processing. A—external-ear filters; B—middle-ear filters; C—cochlea filters; D—hair cell; E—synapse; F—running interaural cross-correlation mechanism.

A simulates the head-related transfer function; B represents the filters that simulate the transfer behaviour of the middle ear; C simulates the critical band-pass filters; D represents the hair cells that act as the half-wave rectifiers; E represents synapses that act as low-pass filters; and F is the RCC mechanism. It is considered that several spatial characterististics of sound image—direction, auditory source width, envelopment—are explained by F.

2.2 Procedure of simulating a sound field based on the RCC model

In this paper, the sound field simulation based on only the RCC model was investigated, neglecting B, C, D and E in Fig. 2 as the first step of the investigation. The procedure of the RCC method is as follows.

First, impulse responses from a sound source to both entrances of the ear canals in an original sound field are obtained by measurement with a dummy head or by predictive calculation (for instance, the image method).

Next, the RCC function $\Phi_{lr}(t)$ is calculated by eqns (1) and (2).

$$\Phi_{lr}(t, \tau) = \int_{-\infty}^{t} p_l(\xi) p_r(\xi - \tau) G(t - \xi) \, d\xi \tag{1}$$

where $p_l(\xi)$ and $p_r(\xi - \tau)$ are the input signals to the left and the right entrance of ear canal, respectively; τ is a time lag between left and right ear; $G(t - \xi)$ is a temporal window. An exponentially shaped temporal window has been used in the model. In the present paper, as the first step of the investigation, a simple rectangular temporal window was adopted to first approximation.

Then the degree of interaural cross-correlation for a temporal window at t, RCC (t), is defined by

$$\mathrm{RCC}(t) = |\phi_{lr}(t, \tau)|_{\max}, \quad \text{for } |\tau| \leq 680\,\mu\mathrm{s} \tag{2}$$

where $\phi_{lr}(t, \tau)$ is the normalised interaural cross-correlation function for a temporal window at t.

The interaural time difference for a temporal window at t, RTD (t) is obtained by eqn (3).

$$\mathrm{RTD}(t) = \tau, \quad \text{for } |\phi_{lr}(t, \tau)| = \mathrm{RCC}(t) \tag{3}$$

The equivalent sound pressure level for a temporal window at t, RESPL (t) is obtained by eqn (4), which is a formula of binaural summation of loudness.[11]

$$\mathrm{RESPL}(t) = 6 \log_2 (2^{L_l(t)/6} + 2^{L_r(t)/6}) \tag{4}$$

where $L_l(t)$ and $L_r(t)$ are the sound pressure level at the left and the right entrance of ear canal for a temporal window at t, respectively.

One reflection of the simulated sound field is derived from each temporal window. The amplitude of the reflection is the RESPL (t). The direction of the reflection is the direction where the interaural time difference corresponds to the RTD (t). The time delay of the reflection is the time when the amplitude is maximum in the temporal window.

3 PSYCHOLOGICAL EXPERIMENT I

The purpose of this experiment is to examine how similar the sound field simulated by the RCC method is to the original sound field.

3.1 Method

The dissimilarity between the original sound field and the simulated one was investigated using Kruskal's multidimensional scaling.

3.1.1 Music motif

The music motif used in the experiment was a 7 s section of Saint-Saens's *Introduction et Rondo Capriccioso* for violin, bars 4–6, recorded in an anechoic chamber.

3.1.2 Original sound field

Figure 3 shows the original sound field which consists of 67 discrete reflections (also including a direct sound). They were emitted from eight loudspeakers which were placed at a horizontal angle of 0°, ±15°, ±30°, ±45, 60° from the median plane. The amplitude and the time delay of reflection was referred to the discrete early reflections measured by Vienna Musikvereinssaal. The direction of incidence of reflection was chosen at random among eight loudspeakers. The direct sound comes from the loudspeaker placed at horizontal angle of 0°.

3.1.3 Simulated sound fields

Table 1 shows eight kinds of simulated sound fields used in the experiment. Four of these sound fields were created by the RCC method as a parameter of duration of the temporal window. The shape of the temporal window was rectangular. The adjacent temporal windows did not overlap, but were in contact with each other; these sound fields were named RCC40, RCC20, RCC10, and RCC5. Another four sound fields were also created by choosing the greatest reflections among those of the original sound field (CGR method); they were named CGR40, CGR20, CGR10, and CGR5. The numeric in both names represents the number of reflections.

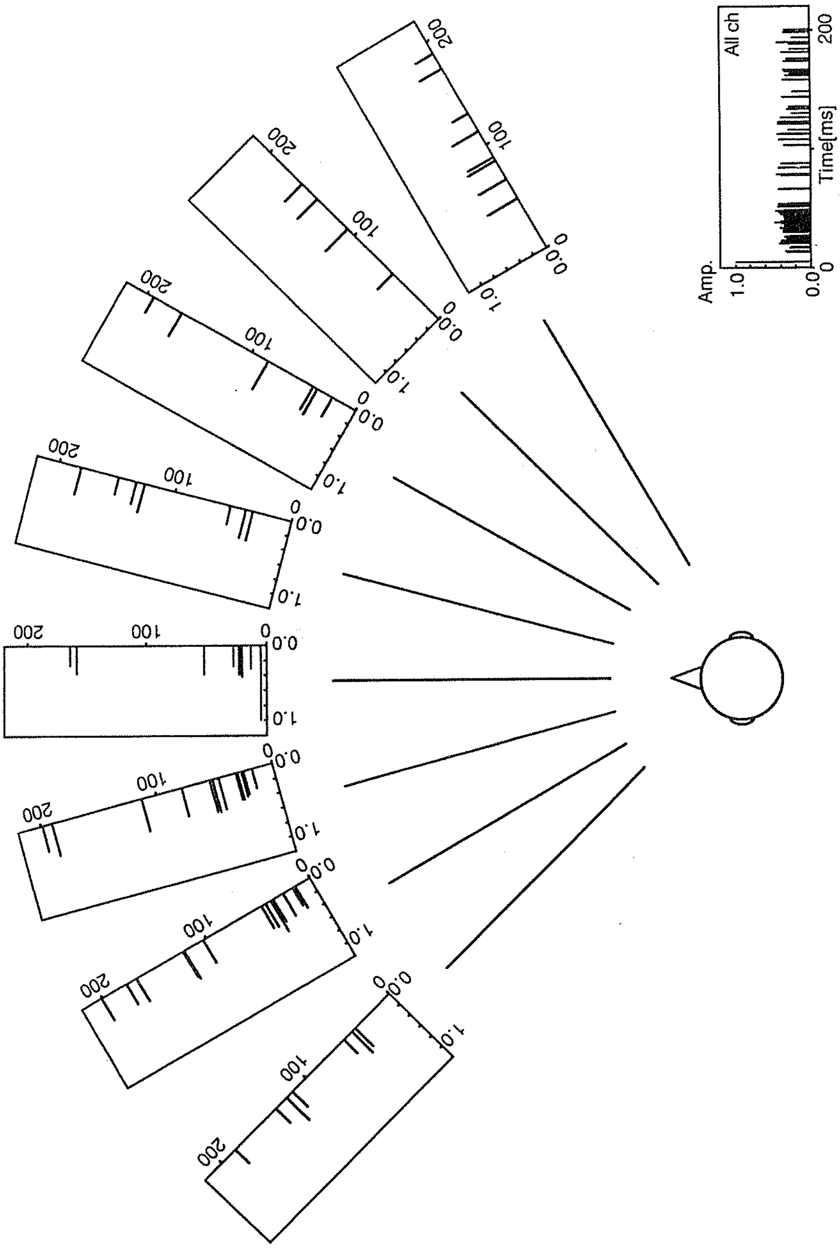

Fig. 3. Temporal and spatial structures of an original sound field which consists of 67 discrete reflections (also including a direct sound).

TABLE 1
Eight Kinds of Simulated Sound Fields

	Running cross-correlation				*Choose the greatest reflections*			
	RCC40	*RCC20*	*RCC10*	*RCC5*	*CGR40*	*CGR20*	*CGR10*	*CGR5*
Number of reflections	40	20	10	5	40	20	10	5
Duration of temporal window (ms)	5	10	20	40	—	—	—	—

Figure 4 shows the temporal structure of the original and the simulated sound fields. Apparently, the sound fields simulated by the CGR method are similar to the original one, but the sound fields simulated by the RCC method are not at all.

3.1.4 Procedure

The 36 pairs of nine sound fields (an original sound field and eight simulated ones) were presented to the subjects. The task of the subjects was to judge the degree of dissimilarity of each pair, concerning sound images on a five-point scale, as shown in Table 2. The interval between two stimuli was 1 s. All pairs of stimuli were arranged, followed by an interval of 4 s, in random order. The length of one series of evaluations was about 16 min. The

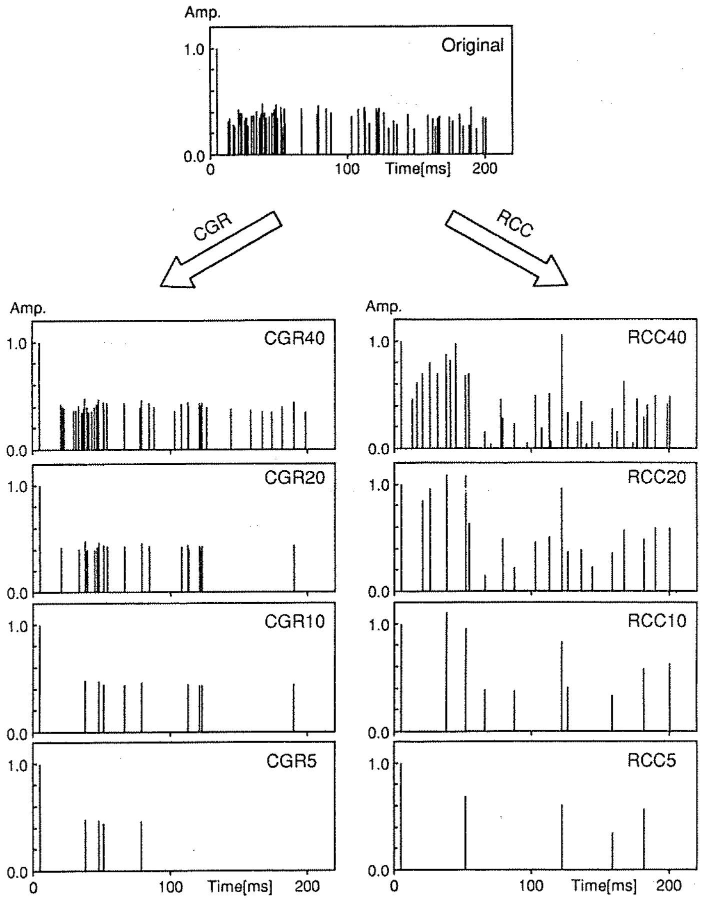

Fig. 4. Temporal structure of the original sound field and the simulated sound fields created by (right) the RCC method and (left) the CGR method.

TABLE 2
Categories for Dissimilarity Judgements

Two sound images are
1—Not different
2—A little different
3—Different
4—Fairly different
5—Extremely different

binaural summation levels[11] of all sound fields were constant at 83·0 dB slow peak.

Each subject was tested individually, while seated, with head fixed in an anechoic chamber. A spotlight was used, such that it provided just enough lighting for the subjects to mark their judgement on the answer sheet. The subjects judged the dissimilarity of three particular pairs of sound fields as a training before their judgements were obtained. Each subject judged the degree of dissimilarity twice for each pair.

3.1.5 Subject

Twelve male students, 22–26 years of age with normal hearing sensitivity acted as subjects in the experiment.

3.2 Results and discussion

At first, the dissimilarity judgements of subjects were examined by three criterion.

(1) *Reliability of judgement.* The average dissimilarity judgement on the pair of identical sound fields is less than 2 (a little different) on a five-point scale in Table 2.

(2) *Reproducibility of judgement.* The correlation coefficient between judgements of the first series and the second series must be more than 0·7 to consider the judgements of the two series to be highly correlated, in general.[12]

(3) *Agreement of judgement.* The correlation coefficient between judgements of any two subjects must be more than 0·7 to consider the judgements of the subjects to be highly correlated, in general.[12]

As a result of the examination, four subjects' dissimilarity judgements passed the criterion. They were averaged and analysed by Kruskal's multidimensional scaling.[13,14]

Figure 5 shows the stress curve. It drops sharply at two-dimensions and

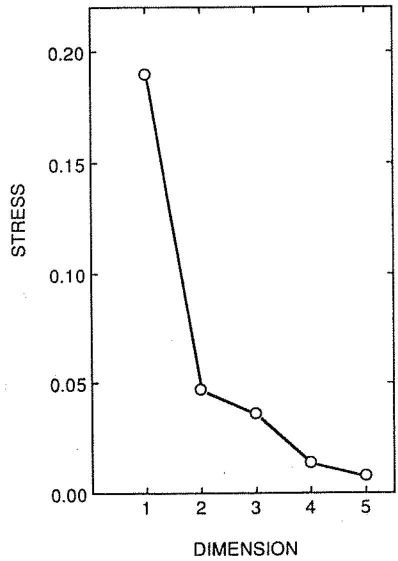

Fig. 5. Stress curve analysed by MDS.

then it decreases slightly. The stress for the two-dimensional configuration was about 4·7%. Kruskal suggests the following idea of how stress values may be used to determine goodness of fit:

20% poor	2·5 excellent
10% fair	0% perfect
5% good	

This means the goodness of fit is 'good' and a two-dimensional solution explains about 95% of psychological space, in this case.

Figure 6 shows the two-dimensional configuration of the nine sound fields. The original sound field is plotted at the origin. The coordinate axes have no meaning and can be set arbitrarily. The times in parentheses are the duration of temporal window. The distance between any two points in space indicates the degree of dissimilarity of two sound fields.

The sound field RCC20 is the most similar sound field to the original one. Namely, the sound field RCC20 which consists of 20 reflections is more similar to the original one than the field CGR40 which consists of 40 reflections. The average of the judgements of the degree of dissimilarity between the sound field RCC20 and the original sound field was 1·63, which lies between 'not different' and 'a little different' in Table 2. This means that the sound image perceived by 67 reflections can almost be simulated by only 20 reflections.

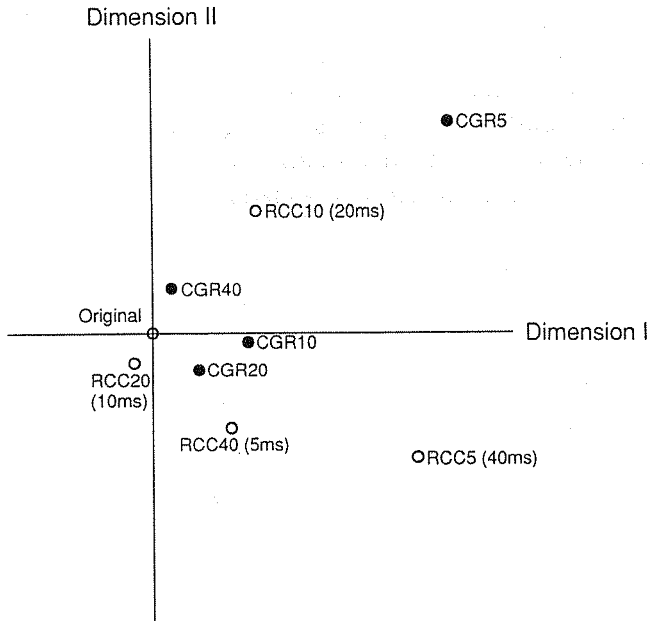

Fig. 6. Two dimensional configuration of nine sound fields. The number represents the number of reflections, and the times in parentheses are the duration of the temporal window.

The similarity between the sound field simulated by the RCC method and original one is not correlated to the number of simulated reflections. From the results of this experiment, it seems to suggest that an appropriate duration of the rectangular temporal window is 10 ms in calculating the RCC.

On the other hand, the similarity between the sound fields simulated by the CGR method and original one is correlated to the number of simulated reflections. Therefore, it is difficult to decide scientifically how many reflections are necessary for the simulation of sufficient accuracy.

4 PSYCHOLOGICAL EXPERIMENT II

The purpose of this experiment is to interpret the psychological configuration obtained in experiment I.

4.1 Method

In experiment I, the subjects reported that they perceived the difference between the sound image of original sound field and that of the simulated one in five characteristics as follows: (a) auditory source width (ASW),[15,16] (b) direction, (c) distance, (d) reverberance, and (e) softness. Auditory source width, direction and distance belong to the spatial attribute of a sound image. Reverberance belongs to the temporal attribute. Softness belongs to the qualitative attribute. In experiment II, the paired comparison test concerning the five characteristics was performed.

4.1.1 Music motif
The same music motif as experiment I was used.

4.1.2 Original sound field
The same original sound field as experiment I was used.

4.1.3 Simulated sound fields
The same eight sound fields as experiment I were used.

4.1.4 Procedure
All pairs of the nine sound fields (36 pairs) were presented to the subjects, as well as experiment I. The task of subjects was to choose which sound field was (a) wider (ASW), (b) righter (direction), (c) further (distance), (d) more reverberant (reverberance), (e) softer (softness), respectively.

Each subject was tested individually, while seated, with head fixed in an anechoic chamber. The subjects judged each characteristics individually. Each subject judged each pair twice, concerning the five characteristics.

4.1.5 Subject
The same four subjects whose judgements were adopted in experiment I participated in this experiment.

4.2 Results and discussion

The psychological scales of the five characteristics for the nine sound fields were obtained using the Thurstone Case V model.[17] The psychological configuration obtained in experiment I was interpreted by multiple regression analysis. In analysing, a dependent variable was the psychological scale values of each characteristic obtained in experiment II and an independent variable were the coordinates obtained in the experiment I.

Figure 7 shows the psychological configuration of the nine sound fields and the axes of the five characteristics. The direction of each arrow indicates 'less wide', 'less reverberant' and so on. The accuracy of simulation of all characteristics is correlated to the degree of similarity in both the RCC and the CGR method. In other words, it cannot be concluded that it is easy or difficult to simulate some particular characteristics.

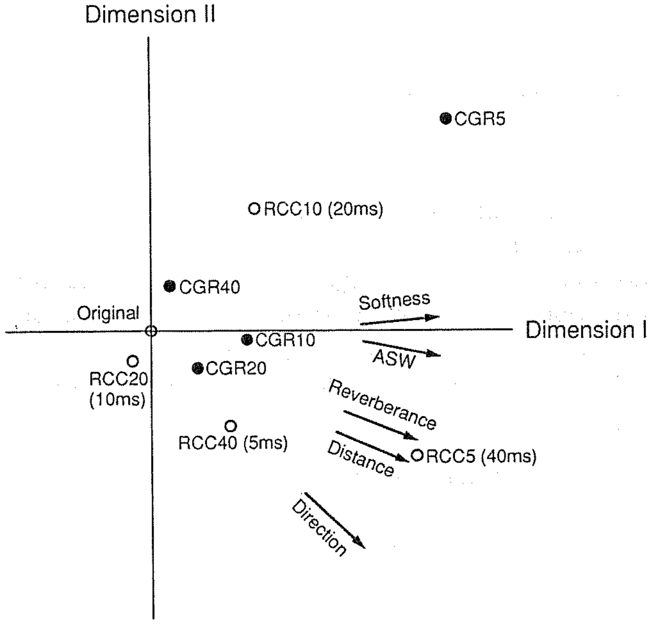

Fig. 7. Two-dimensional configuration of nine sound fields and psychological axes obtained by multiple regression analysis. The arrows indicate the directions that the subjects perceived less concerning each characteristics of sound image.

5 CONCLUSIONS

Two hearing tests were carried out in order to pursue the possibility of the sound field simulation based on the running interaural cross-correlation model (RCC method), comparing with the simulation by means of choosing the greatest reflections (CGR method). The results infer the following.

(1) There is a possibility of simulating almost the same characteristics of sound image as those perceived in an original (to be simulated) sound field, by the RCC method. Using the RCC method, the necessary reflections for the simulation is significantly less than reflections included in the original sound field.

(2) The accuracy of simulation by the CGR method is correlated to the number of reflections, but that by the RCC method is not correlated to the number of reflections. This suggests that there is an appropriate temporal window (shape, duration, and so on) for the simulation by the RCC method.

(3) The accuracy of simulation of five characteristics (auditory source width, direction, distance, reverberance and softness) is correlated to the accuracy of simulation obtained by the dissimilarity judgement

between an original sound field and sound fields simulated by the RCC method and the CGR method. It cannot be concluded that it is easy or difficult for both methods to simulate a certain particular characteristic.

There still remains technical problems to be solved in practical application. In principle, the RCC method described in this paper can estimate the horizontal angle of a sound image. But it cannot distinguish between front and back directions of a sound image nor can it estimate vertical angle. A tentative method is now being investigated to settle these problems. It is the RCC method which uses the impulse responses at both entrances of the ear canals for different directions of the aural axis.

ACKNOWLEDGEMENT

The authors wish to thank Mr E. Koyama for his cooperation in the experiments.

REFERENCES

1. Camras, M., Approach to recreating a sound field. *J. Acoust. Soc. Amer.*, **43** (1967) 1425–31.
2. Kleiner, M., Svensson, P. & Dalenbäck, B.-I., Auralization: experiments in acoustical CAD in *Proceedings of the Audio Engineers Society 89th Convention*, 1990, I-I-3.
3. Morimoto, M. & Ando, Y., On the simulation of sound localization. *J. Acoust. Soc. Jpn*, **(E)1** (1980) 167–74.
4. Damaske, P., Head-related two channel stereophony with loudspeaker reproduction. *J. Acoust. Soc. Amer.*, **50** (1971) 1109–15.
5. Hamada, H., Ikesyhoji, N., Ogura, Y. & Miura, T., Relation between physical characteristics of orthostereophonic system and horizontal plane localization. *J. Acoust. Soc. Jpn*, **(E)6** (1985) 143–54.
6. Kimura, Y., Watanabe, K. & Gotoh, T., Partition-stereophony to reproduce one's spatial hearing events. *J. Acoust. Soc. Jpn*, **42** (1986) 387–95.
7. Posselt, Ch., Schroter, J., Opitz, H., Divenyi, H. & Blauert, J., Generation of binaural signals for research and home entertainments. In *Proceedings of the 12th International Congress on Acoustics*, Toronto, 1986, B1–6.
8. Blauert, J., *Spatial Hearing: The Psychophysics of Human Sound Localization.* The MIT Press, Cambridge, MA, USA, 1983.
9. Blauert, J., Morimoto, M. & Gotoh, T., *Kuhkan Onkyo* [Spatial Hearing], Kajima-Syuppankai, Tokyo, Japan, 1986 (in Japanese).
10. Blauert, J. & Cobben, W., Some consideration of binaural cross correlation analysis. *Acustica*, **39** (1978) 96–104.
11. Robinson, D. W. & Whittle, L. S., The loudness of directional sound field. *Acustica*, **10** (1960) 74–80.

12. Hitano, S., Seya, M. & Okawa, N., *Tokeigaku* [Statistics], Baifukan, Tokyo, Japan, 1961 (in Japanese).
13. Kruskal, J. B., Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica*, **29** (1964) 1–27.
14. Kruskal, J. B., Nonmetric multidimensional scaling: A numerical method. *Psychometrica*, **29** (1964) 115–29.
15. Morimoto, M. & Maekawa, Z., Auditory spaciousness and envelopment. In *Proceedings of the 13th International Congress on Acoustics*, Vol. 2 (Belgrade, 1989), pp. 215–18.
16. Morimoto, M., Fujimori, H. & Maekawa, Z., Discrimination between auditory source width and envelopment. *J. Acoust. Soc. Jpn*, **6** (1990) 449–57 (in Japanese).
17. Thurstone, L. L., A law of comparative judgement. *Psycol. Rev.*, **35** (1927) 273.