Relationships between interaural cross-correlation function and spatial impression of sound reproduced with five loudspeakers

Ryo Isobe and Takashi Nishi*

Graduate School of Environmental Engineering, Kitakyushu University, 1–1 Hibikino, Wakamatsu-ku, Kitakyushu, 808–0135 Japan
(Received 24 August 2011, Accepted for publication 21 November 2011)

Keywords: Spatial impression, Interaural cross-correlation function, Sound localization

PACS number: 43.60.Pn, 43.66.Qp [doi:10.1250/ast.33.193]

1. Introduction

In this paper, we describe the evaluation of a five-channel sound reproduction system. Multichannel home theater systems have become increasingly common. It has become important to evaluate the spatial impression of sound images, such as their position, number and width, and the extent of the ambience. The physical parameters, interaural time difference (ITD), interaural level difference (ILD), and interaural cross-correlation coefficient have been used to evaluate the spatial impressions of sound images [1–3]. However, so far, no method has been established to systematically evaluate the spatial impressions reproduced by a multichannel sound system.

To address this problem, the interaural cross-correlation function (IACF) is proposed as the key parameter for evaluating spatial impressions. Typical sound scenes are simulated using a five-channel reproduction system. Five pink noise signals, which are stationary and independent, are used as the sound source for the simulation study. Using the IACF patterns, the typical sound scenes are classified.

2. Evaluation parameters for spatial sound images

2.1. Sound source

The Radiocommunication Sector of the International Telecommunication Union (ITU-R) recommends that the loudspeaker configuration shown in Fig. 1 be adopted to reproduce five sound channels.

Five channel signals (L, R, C, LS, RS) were created using 5 independent signal of pink noise (p1–p5):  
\[
\begin{align*}
L &= p_1 \cos \theta_1 + p_2 \sin \theta_1 \\
R &= p_1 \sin \theta_1 + p_2 \cos \theta_1 \\
C &= p_3, \\
LS &= p_4 \cos \theta_2 + p_5 \sin \theta_2 \\
RS &= p_4 \sin \theta_2 + p_5 \cos \theta_2
\end{align*}
\]

where \( \theta_1 \) is the parameter that controls the correlation \( \rho_1 \) between the front pair of loudspeaker signals (L and R), and \( \theta_2 \) controls the correlation between the rear pair of loudspeaker signals (LS and RS). The relationship between \( \theta \) (\( \theta_1 \) or \( \theta_2 \)) and the cross-correlation coefficient \( \rho \) is shown in Table 1.

2.2. Creation of IACF pattern

Figure 2 shows the block diagram for obtaining the IACF pattern. The binaural L and R signals were obtained by convolving the head-related impulse responses (HRIR) of the left and right ears to pink noise sound sources from 5 loudspeakers. The database of HRIR provided by MIT [4] was used. The peripheral processing of the human auditory system was simulated by a GammaTone Filter Bank (GTFB) [5] followed by inner-hair-cell processing. The GTFB had 50 channels of sound ranging from 50 Hz to 20 kHz equally distributed on an equivalent rectangular bandwidth scale. The neural transduction process was approximated by envelope modulation and half-wave rectification followed by the square root envelope compression proposed by Bernstein and coworkers [6,7]. Using the binaural signals \( SI \) and \( SR \), IACF is defined as

\[
IACF(f, \tau) = \frac{\left[ \int_{T_1}^{T_2} S_L(f, t) S_R(f, t+\tau) dt \right]^2}{\left[ \int_{T_1}^{T_2} S_I^2(f, t) dt \right] \int_{T_1}^{T_2} S_R^2(f, t) dt}, \tag{2}
\]

where \( f \) is the center frequency of the GTFB and \( \tau \) is the lag time ranging from \(-1.2 \) to \(1.2 \) ms. Although \( \pm1 \) ms is commonly used as the lag time limit for geometrical reasons, we extend this limit to \(1.2 \) ms to observe the behavior of the correlation over a wider time lag. The IACF is estimated using a Hamming window with a frame length of 100 ms at a sampling frequency of 44.1 kHz. A 50% overlap between successive frames was applied to the L and R time sequences. The maximum IACF value (IACC) and corresponding lag time (ITD) are obtained from the IACF defined by Eq. (2). The ITD and IACC are considered to correspond to the sound source location and sound width, respectively. The IACF is also expected to include information about the ambience, which is produced by the number and the coherence of independent sound sources.

Since stationary pink noise was used as the sound source, the IACF value of each frame was stable, and the sound image could be evaluated easily.

3. Evaluation of the IACF patterns

Typical auditory scenes experienced were subjectively classified into five types: F, FC, B, FB and FCB [8].
Corresponding spatial impressions of these five scenes are summarized in Table 2.

The relationship between the IACF pattern and the sound image, such as the position and width, and the extent of the ambience, were studied through simulation.

### 3.1. Sound localization

The same pink noise signal was supplied to the center and the right loudspeaker as a sound source. Using the tangent law, each loudspeaker gain was determined to obtain the virtual sound image. Figure 3 show the IACF patterns where the angle of the sound image localization is varied from the right (30°) to the midpoint (15°) to the front center (0°).

### 3.2. Sound width

Figure 4 show the IACF patterns where the cross-correlation coefficient between front left and right signals is changed from 0 (independent) to 1 (same). When the cross-correlation of the signals is low (Fig. 4(a)), two peaks appear around a lag time of ±0.3 ms, which correspond to the open angle of the loudspeakers. However, when the cross-correla-
tion of the signals is high (Fig. 4(d)), the peak appears around a lag time of 0 ms and extends through all frequency bands. When the cross-correlation is low, the sound image is broadly extended in the foreground. When it is high, a sharp and fixed sound image is obtained. Therefore, it is established that the peak value in the IACF patterns corresponds to the sound width.

### 3.3. Ambience

To investigate the correspondence between the extent of the ambience and the IACF pattern, the position and the number of independent sound sources were varied. Figures 5–9 show the IACF patterns corresponding to the typical five audio scenes, where front, rear, and center signals are independent of each other.

Since the path difference from the loudspeakers to each ear under the rear ambience condition is longer than that under the front ambience condition, the frequency bandwidth under the rear ambience condition, where the correlation is high, becomes narrower than that under the front ambience condition.

---

**Fig. 4** IACF patterns where the cross-correlation coefficient $\rho$ between front left and right signals is changed from 0 (independent) to 1 (same).

**Fig. 5** IACF pattern where the cross-correlation between front left and right signals is set to zero (Scene F).

**Fig. 6** IACF pattern where the front 3 signals are independent of each other (Scene FC).

**Fig. 7** IACF pattern where the cross-correlation between the rear left surround and rear right surround signals is set to zero (Scene B).

**Fig. 8** IACF pattern where the front and rear 4 signals are independent of each other (Scene FB).
condition. This is why the low frequency bandwidth in Fig. 5, where the value of IACF becomes high, is wider than that in Fig. 7. In addition, the positions of the high correlation peak, shown as eyelike patterns in the medium frequency bands (GTFB#10-#25) in Figs. 5 and 7, are $\pm 0.3$ ms and $\pm 0.7$ ms. These time lags should correspond to the loudspeaker positions. The total ambience condition (FB) in Fig. 8 shows the intermediate IACF pattern between Figs. 5(F) and 7(B). A comparison of Fig. 5 with Fig. 6 shows that the IACF value of low frequency in Fig. 6 becomes higher than that in Fig. 5. This is thought to be the effect of the addition of the center channel. Moreover, the positions of the high correlation peak in Fig. 5, shown as an eyelike pattern in the medium frequency bands (GTFB#20-#30), are more accentuated than in Fig. 6. These characteristics can also be observed in Figs. 8 and 9.

From the results of these series of simulations, it is concluded that the IACF pattern reflects the differences in the ambience impression well.

4. Conclusions

Spatial impressions were classified using the sound source localization, sound image width, and ambience. To obtain the physical measures corresponding to the spatial impression, the IACF pattern was adopted. The effectiveness of the IACF pattern was studied, focusing on the evaluation of a five-channel reproduction system through simulation. It is clear that the peak position, the peak value, i.e., IACC, and the distribution of the IACF pattern correspond to the sound source localization, source width, and ambience, respectively. In future work, we will apply the IACF pattern to the time-varying general sound signals and evaluate whether fidelity to the originally intended spatial characteristics can be realized using the reproduced multichannel sound.

References