FIFTH INTERNATIONAL CONGRESS ON SOUND AND VIBRATION
DECEMBER 15-18, 1997
ADELAIDE, SOUTH AUSTRALIA

RELATION BETWEEN TRICC AND SOUND IMAGE PERCEPTION IN A SOUND FIELD WITH A SINGLE ECHO

H. Yanagawa(1), *T. Tagaeto*(2), *and* *H. Komatsu*(2)

(1) Chiba Institute of Technology and (2) Kogakuin University

ABSTRACT

To estimate the quality of a binaural sound image and spatial impression, we used a burst white noise signal as an example of transient signals and evaluated the temporal change of the TRICC (maximum value of TRansient Interaural Cross Correlation function) in the sound field. The TRICC is unity when we can observe a direct sound without a clear echo. The magnitude of an echo, however, can be large in a real sound field. When the loudness of the echo becomes equal to that of the direct sound, a sound image created by the noise burst (direct sound) with the single echo broadens between the directions of the leading sound and the echo. We investigated the broadening of the sound image of a noise burst pair, and we analyzed the relationship between the sound image broadening by the single echo and the temporal change of TRICC. The precedence effect on the temporal change of sound image was clearly observed, when the time delay of an echo is within 2 ms. Accordingly TRICC must be still studied in order to reasonably predict the sound image broadening due to a transient signal.

1. INTRODUCTION

IACC is closely related to the preference of the sound quality of concert halls [1]. Moreover, IACC and Interaural Correlation Coefficients (ICCs) are inversely proportional to subjective diffuseness [2, 3]. The ICCs or IACC are derived from a pair of steady state responses to a random noise. The sound image, however, can be temporally changed particularly for transient signals. After the sound image is localized by the direct sound, the sound image becomes blurred by superposition of reflected sounds. And the decay process of the sound field makes subjective diffuseness after the sound source is put off. Therefore, a transient interaural cross correlation function has been introduced by Yanagawa et al. [4] to estimate such a temporal change of the spatial impression. The maximum value of a normalized transient interaural cross correlation function is called TRICC. Yanagawa et al. [4] showed that the temporal change of the TRICC measured using a white noise burst signal could be a efficient estimator for the temporal changes of the spatial impression in rooms.

In this article we study a sound with a single echo, since one of the most simple sound fields is composed of a direct sound and a single reflected sound. If the incident direction of the leading sound is different from that of the echo under the condition that both sound levels are equal, a sound image is localized in the direction of the leading sound. This is known as the precedence effect [5, 6]. As the sound level of the echo with a time delay increases, we can find a level difference L_{dif} between the leading sound and the echo so that the loudness of the leading sound is equal to that of the echo. Under this equal-loudness condition, the sound image can be distributed between the directions of the leading sound and the echo. This sound image broadening, however, might depend on the time delay of the single echo even under the equal-loudness condition. In this work we investigate the changes of the sound image broadening due to the delay time of the echo under the precedence effects. The transient interaural cross-correlation functions are analyzed in order to evaluate the temporal change of the sound image broadening.

II. TRANCIENT INTERAURAL CROSS CORRELATION FUNCTION IN A SOUND FIELD[7]

In order to obtain a transient interaural cross correlation function, it is convenient to study the radiation, build up, and termination of signals as proceeding in steps (Fig. 1). A cross correlation function is fundamentally defined by using an ensemble average, so that it can be applied to the response to transient signals. When a white noise $s(t)$ radiates at $t = 0$ and stops $t = T$, in a room, the temporal change of cross correlation function at both ears can be expressed by

$$\Phi_{lr}(t,\tau) = \left\langle \int_0^T s(\gamma)\, h_l(t-\gamma)\, d\gamma \int_0^T s(\gamma')\, h_r(t-\gamma')\, d\gamma' \right\rangle \tag{1}$$

where < > denotes taking an ensemble average. Both impulse response $h_l(t)$ and $h_r(t)$ are impulse responses from the sound source to both ears. Thus they are convolutions of the room impulse responses and the impulse response for each incident sound to both ears. Since the autocorrelation functions for the white noise become delta functions, Eq. 1 can be written as

$$\Phi_{lr}(t,\tau) = \int_{T_1}^{T_2} h_l(t)\, h_r(t+\tau)\, dt\,. \tag{2}$$

In a built-up process of the sound field, $T_1 = 0$ and $T_2 = t_u$ where t_u is an elapsed time after the sound source radiates. In a decay process of the sound field, $T_1 = t_d$ and $T_2 = \infty$ where τ_d is an elapsed time after the sound source stops. TRICC is the maximum value of the cross correlation function within a range of lag time $\tau = \pm 1$ ms.

Based on time and spatial characteristics of a sound field, we can surmise temporal changes of TRICC in transient states. In the case of a build-up process of the sound field, a direct sound arrives earliest at both ears. TRICC is almost unity. As time elapses, the reflected sound arrives from directions different to that of the direct sound and they are superposed on the direct sound. This causes a decrease in TRICC. As the sound field becomes steady state, TRICC converges into a value of the full IACC. After the sound source stops, the sound field is in the decay process. At first, the direct sound diminishes and then the reflected sounds diminish from the superposed sound. Thus, TRICC decreases to the value of a diffuse field (see Fig. 1).

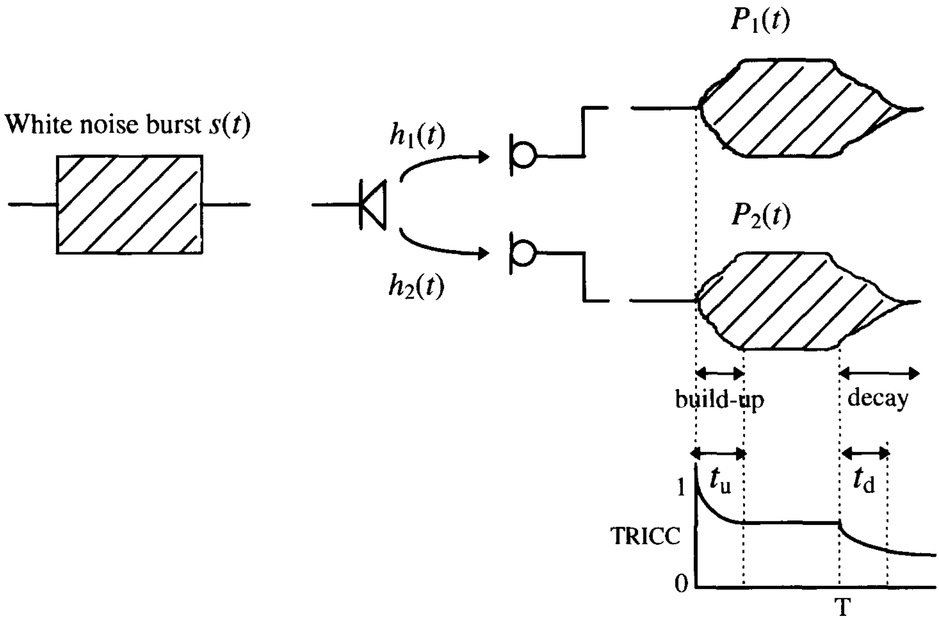

Fig. 1 TRICC in a reverberation sound field

This temporal change of TRICC corresponds to the temporal change of a spatial impression in a room. At first, the sound image is localized by the direct sound and sound image blur is produced by superposition of reflected sounds. Finally, the decay process of the sound field makes subjective diffuseness.

III. A SOUND IMAGE BROADENING BY A SINGLE ECHO

The relation between the size of a sound image and the delay time τ_s of the echo was investigated using a noise pair as a pointer signal whose ICC was changed. Figure 2 shows the arrangement for the experiments. The signal used for the leading sound and the echo is a weighted pink noise. The weighting curve is determined by following IEC publication 268-1, which simulates the average power spectrum of music sounds, since we focused on the quality of the sound image of music sounds. The experiments were performed in an anechoic room at Kogakuin University. The listening sound level was 80 dB. The time duration of the burst noise signal presented to the subjects was 4 s in order to avoid an onset effect which could have affected the subjective judgment for the loudness balance condition.

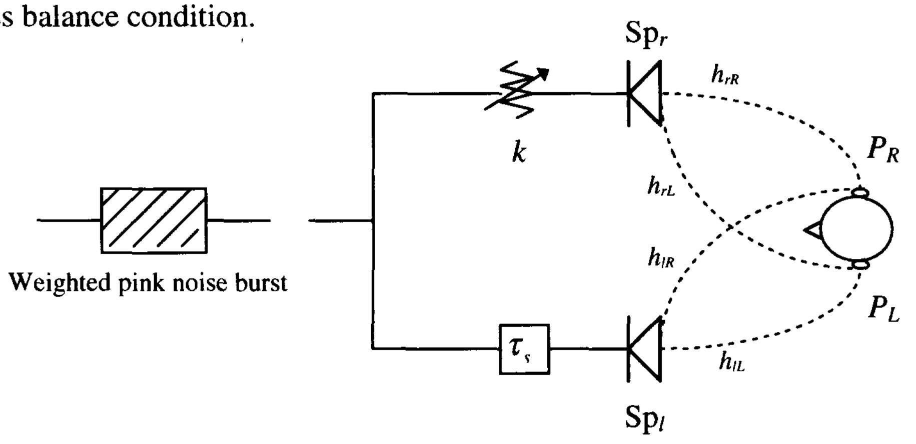

Fig. 2 The arrangement for the sound image experiments

Experiment 1. Time delay τ_s and loudness balance

The subjects were asked to change the sound level difference between the leading sound and the echo so that the loudness of the leading sound becomes equal to that of the echo. Figure 3 shows the level differences, under the equal-loudness condition, which were obtained by 3 subjects. The solid line shows the average, the vertical axis is the level difference L_{dif}, and the horizontal axis shows the time delay of the single echo. We can see that as τ_s increases (within 1 ms), L_{dif} becomes large. The maximum L_{dif} is given at 8 dB. After τ_s becomes larger than 2 ms, L_{dif}, however, decreases inversely proportional to the logarithm of τ_s.

Experiment 2. Time delay τ_s and size of a sound image

The same subjects who attended Experiment 1 attended Experiment 2. They were asked to change the correlation coefficients of the pointer noise so that the size of the sound image created by the pointer noise could be judged equal to that by the leading sound and the echo under the equal loudness condition. The pair of pointer noise consisted of two weighted pink noise signals whose weighting curves are the same as the leading sound and the echo. The pair of pointer noise signals with various correlation coefficients ρ_s were made by using two incoherent weighted-noise signals. The subject's interaural cross correlation function $\Phi_{LR}(\tau)$ is estimated from the correlation coefficients ρ_s of pointer noise signals by using HRTFs [8] as shown in Figs. 4. Figure 5 is the ICC which are defined by the interaural cross correlation functions $\Phi_{LR}(\tau)$ of $\tau = 0$. The time durations of the noise signals were 4 s.

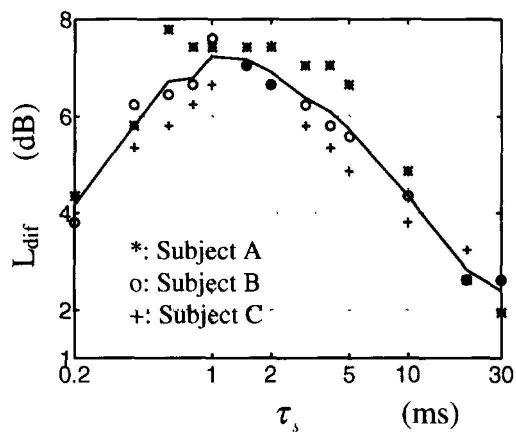

Fig. 3 Delay time of the echo τ_s and level difference L_{dif} between the echo and leading sound

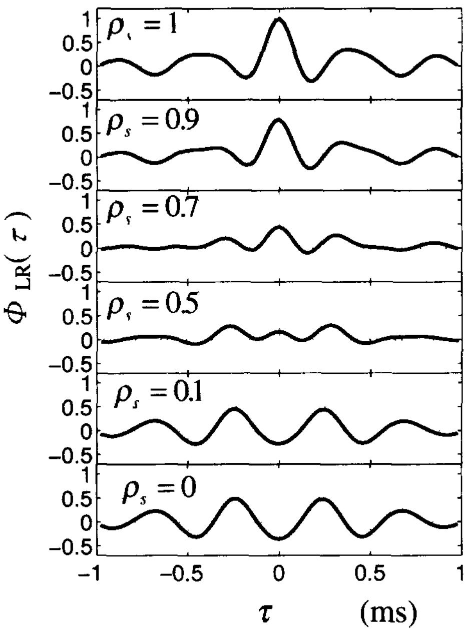

Fig. 4 Interaural Cross correlation function and ρ_s of the pointer noise

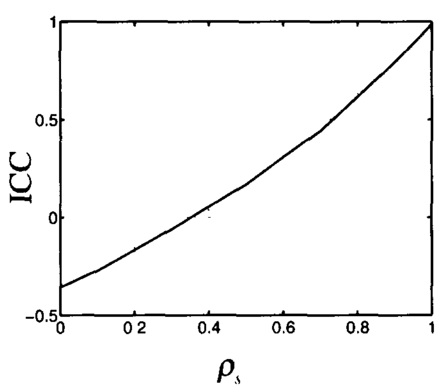

Fig. 5 ICC and ρ_s

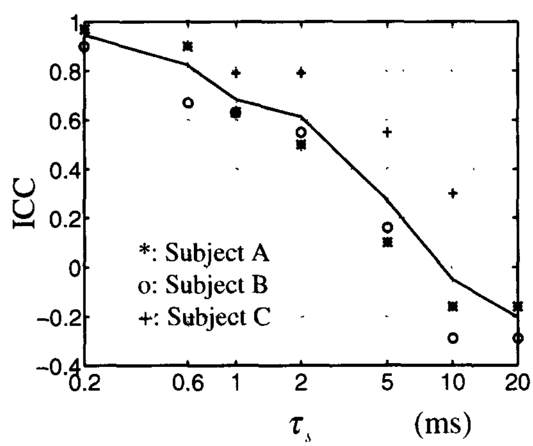

Fig. 6 Delay time of the echo τ_s and ICC of the pointer noise signal which sound image size is same as that of the leading sound and the echo

The experimental results are shown in Fig. 6. The solid line shows the average and the vertical axis is the ICC of the pointer noise signal. ICC is inversely proportional to the size of the sound image or subjective diffuseness [3]. The sound image broadens as τ_s increases. From the inner view report of the subjects, the timber of the sound image is perceived differently between the left and the right sides of the broadened sound image when τ_s reaches around 5 ms; however, a subject's judgment of the size could be made. When τ_s is 0, the size of the sound image is minimum. However, the sound image broadens as τ_s increases, and when τ_s become longer than 10 ms, the timber difference stated above is not significantly observed and the sound image broadening makes the subjective diffuseness similar to that obtained by using the pointer noise.

IV. TRANSIENT INTERAURAL CROSS CORRELATION FUNCTION OF A SOUND FIELD WITH A SINGLE ECHO

To observe how the interaural cross correlation function changes temporally by a single echo, transient interaural cross correlation functions were analyzed. Here we used the following set of equations.

$$P_R(t) = s(t) * \{k \cdot h_{rR}(t) + h_{lR}(t - \tau_s)\}$$

$$P_L(t) = s(t) * \{k \cdot h_{rL}(t) + h_{lL}(t - \tau_s)\}$$

$$\Phi_{RL}(t,\tau) = \frac{\langle P_R(t) \cdot P_L(t+\tau) \cdot W(t) \rangle}{\sqrt{\langle P_R^2(t) \cdot W(t) \rangle} \cdot \sqrt{\langle P_R^2(t) \cdot W(t) \rangle}} \qquad (3)$$

where $\mathrm{W(t)} = 1\text{:} \quad 0 < \mathrm{t} < t_u$

$\mathrm{W(t)} = 0\text{:} \quad \text{else where}$

$k = 10^{-\mathrm{L_{dif}}/20}$

Here, h_{rR}, h_{lR}, h_{rL}, and h_{lL} are convolutions of the impulse responses from the loudspeaker to each ear (see Fig. 2) and impulse responses of the IEC weighting filter and the equalization filter used for getting the pink noise from a white noise $s(t)$. The time t_u is an elaped time after the first arriving of the leading sound at the ear.

Figure 7 shows the sample of the temporal changes of the transient interaural cross-correlation functions. The transient interaural cross correlation functions are time averaged by a time constant of 5 ms [9]. We can see the first peak which shows the arrival of the leading sound. This peak decreases if the echo signal arrives in the time window used for calculating the transient interaural cross correlation function. Thus the 2nd peak which is smaller than unity builds up as the echo arrives. The delay time at the 2nd peak can be estimated from the echo delay time and its direction. That is, when the echo arrival time is shorter than 1 ms, the 2nd peak can be seen within ± 1 ms of lag time τ. In Figure 7 (b) we can see that both peaks are on the same side of τ as that for the leading sound; however, our subjective experiment results indicate that the sound image is not perceived clearly localized to the direction of the leading sound. As τ_s is lager than 20 ms, the noise burst and the echo become incoherent. thus the transient interaural cross correlation function converges into the interaural cross correlation function of approximately $\rho_s = 0.1$ (Fig. 4 and Fig. 7 (d)) and the sound image perceived as the subjective diffuseness.

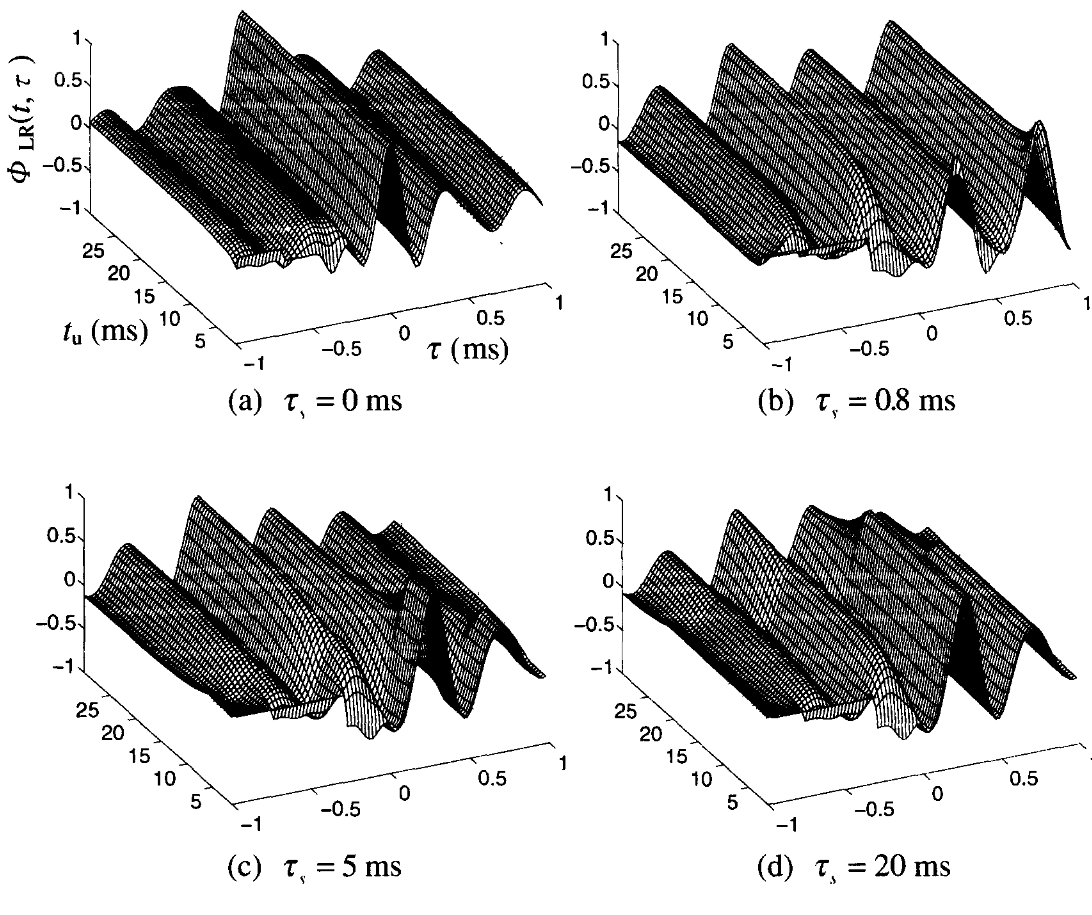

Fig. 7 Transient interaural cross correlation functions for the noise burst and the echo

VI. DISCUSSION

We have investigated the temporal change of a sound image created by the burst noise signals followed by a single echo. The level difference L_{dif} that makes loudness of the burst noise equal to that of the echo has been obtained. We could evaluate inhibition of the precedence effect according to L_{dif} [10]. The inhibition increases as the echo time delay reaches 1 ms. As the echo delay time becomes longer than 2 ms, the inhibition decreases inversely proportional to the logarithm of the delay time τ_s. The effective time range of the inhibition might be related to the time duration T_t of the waveform, which is given by the convolution of the signal's autocorrelation function and the impulse response between the sound source and the outer ear entrance. Since T_t, which is approximately 3 ms in these experiments, corresponds to the delay time τ_s at which L_{dif} begins to decrease (Fig. 3). The transient interaural cross correlation function describes the temporal changes of the interaural cross correlation function for a transient signal such as a noise burst and with a single echo. As the τ_s lager than 20ms, the transient interaural cross correlation function converges into the interaural cross correlation function of approximately $\rho_s = 0.1$ and the sound image perceived as the subjective diffuseness. Though ICC of the pointer noise is -0.2 which corresponds to the subjective diffuseness, however, TRICC and also IACC are 0.5. Therefore, TRICC or IACC, which is defined by the maximum value or magnitude of the interaural cross correlation function, could not estimate the size of the sound image for such a simple sound field consisted of the noise burst and the single echo.

VII. REFFERENCES

[1] Schroeder, M. R., D. Gottlob, and K. F. Siebrasse (1974) J. Acoust. Soc. Am. 56, pp.1195-1201

[2] Damaske, P. and Y. Ando (1972) Acustica 27, pp.231-238

[3] Yanagawa, H., T. Anazawa, and T. Itow (1990) Acustica, 71(3), pp.230-232

[4] Yanagawa, H., Y. Yamasaki, and T. Itow (1988) J. Acoust. Soc. Am., 84(5), pp.1728-1733

[5] Haas, H. (1951) Acustica 1, pp.49-58

[6] Wallach, H., E. B. Newman, and M. R. Rosenzweig (1949) Am. J. Psychol. 57, pp.315-336

[7] Yanagawa, H. (1996) Music and Concert Hall Acoustics, pp.251-259

[8] Gardner, B., and K. Martin, MIT Media Lab. (1994)

[9] Blauert, J., and W. Cobben (1987) Acustica 39, pp.96-103

[10] Aoki, S. and T. Houtgast (1992) Hearing Res. 59, pp.25-30