ICIC Express Letters
Part B: Applications
Volume **13**, Number **4**, April **2022**

ICIC International ©2022 ISSN 2185-2766
pp. 379–387

A STUDY ON A BINAURAL MODEL FOR ESTIMATING THE SOUND IMAGE WIDTH IN A CAR CABIN AUDIO LISTENING ENVIRONMENT

Koji Sakamoto

CI Business Group
Denso Ten Limited
1-2-28 Gosho-dori, Hyogo-ku, Kobe, Hyogo 652-8510, Japan
kohji.sakamoto_01@denso-ten.com

Received August 2021; accepted October 2021

Abstract. *Improving sound image localization and width is an important element for increasing the sense of presence in a car cabin. The construction of estimation model for sound image localization and width is required in order to evaluate objectively these effects. This study proposes a binaural model for estimating the sound image width in a car cabin audio listening environment. Based on a binaural model for the sound image azimuth estimation proposed in a previous study, a new feature quantity was extracted and was expanded to a model capable of estimating the sound image width. The feature quantity was extracted from the attack region of the gammachirp auditory filter bank outputs. Physical measurements and psychological experiments in the car cabin were then carried out, and the proposed method was compared with the conventional methods. As a result, it was confirmed that the proposed method could obtain a higher estimated accuracy than the conventional methods in the car cabin audio listening environment.*

Keywords: Sound image azimuthal binaural model, Gammachirp auditory filterbank, Attack region, Extraction of feature quantity

1. **Introduction.** Recently, three-dimensional (3D) sound technology in a car cabin has become recognized, and comfortable acoustic spaces can be offered to the occupant by increasing the sense of presence through the 3D sound. Improvements to sound image localization (SIL) to enhance the sense of presence, and construction of an SIL estimation model that objectively evaluates the effect are considered crucial elements in 3D sound technology [1, 2]. A previous study proposed a binaural model for the sound image azimuth (SIA) estimation as a component of SIL in a car cabin audio listening environment [2]. In addition, the reproduction capability of the surrounding sound is another element that increases the sense of presence, such that construction of a sound image width (SIW) estimation model is required. The balance of these estimates can be objectively evaluated if the SIA and SIW, which are elements of the sense of presence, can be estimated, as it contributes to the optimization of sensitivity design on such sense. Moreover, SIA and SIW estimation models will be expected to apply to signal alerts issued by advanced driver-assistance systems (ADAS) [3], and robot hearing [4], etc.

As physical measures for apparent source width (ASW), the interaural cross-correlation (IACC), a renowned factor of interaural cross-correlation function (IACF), had its effectiveness confirmed as an indicator in the real sound field of a concert hall [5]. The degree of interaural cross-correlation (DICC) without an ear simulator and A-weighting for IACC was then proposed and developed into a multiple regression model with DICC and binaural summation of sound pressure level (BSPL) as independent variables and the angle of ASW as a dependent variable [6]. It has been confirmed that these conventional methods

DOI: 10.24507/icicelb.13.04.379

are highly accurate. However, the estimated accuracy may be lowered when applied to a car cabin audio listening environment because it is different from the environment in which conventional methods are constructed. Although objective evaluation methods for various sounds in a car cabin have been researched [7-10], limited studies have reported an SIW estimation model used in a car cabin audio listening environment. Therefore, in this study, the binaural model applied for SIA estimation in a car cabin audio listening environment was expanded into a model for estimating the SIW and was compared with other conventional methods. The results indicate that the proposed method can obtain a higher accuracy at a statistically significant level.

Sections 2 and 3 of this paper describe the conventional and proposed methods, respectively. Section 4 details the physical measurements and psychological experiments conducted in a car cabin. Using the experimental data of a car cabin, the models applied by the conventional and proposed methods are verified in Section 5. Finally, Section 6 concludes the paper.

2. **Conventional Method.** The conventional methods for ASW are described below. The normalized IACF is expressed through Equation (1):

$$\Phi_{\mathrm{lr}}(\tau) = \lim_{T \to \infty} \frac{\dfrac{1}{2T}\displaystyle\int_{-T}^{+T} f_{\mathrm{l}}(t) f_{\mathrm{r}}(t+\tau)\,dt}{\dfrac{1}{2T}\sqrt{\displaystyle\int_{-T}^{+T} f_{\mathrm{l}}^2(t)\,dt \int_{-T}^{+T} f_{\mathrm{r}}^2(t)\,dt}} \tag{1}$$

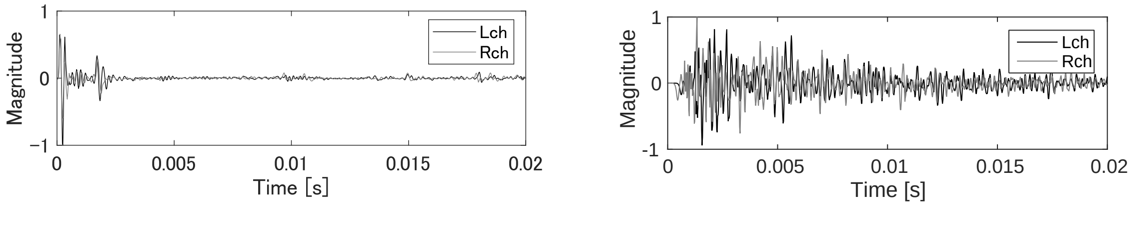
Here, $f_{\mathrm{l,r}}(t)$ represents the binaural signal filtered through the A-weighting filter, and τ represents the interaural time difference (ITD). IACC is expressed through Equation (2):

$$\mathrm{IACC} = |\Phi_{\mathrm{lr}}(\tau)|_{\max}, \ |\tau| \le 1 \ [\mathrm{ms}] \tag{2}$$

The multiple regression model with DICC and BSPL is expressed in Equation (3):

$$\mathrm{ASW} = -39.6X + 1.55Y - 31.9 \ [\mathrm{deg.}] \tag{3}$$

Here, X indicates DICC and Y represents BSPL [dB]. Figure 1 shows an example of impulse responses from a sound source to the left and right ears of the dummy head at a listening position in a concert hall (small hall) and in a car cabin. The conventional methods have been verified in a sound field in which the reflected sound arrives after the direct sound reaches the listening position, such as in a concert hall. IACC and DICC are lowered as there are many reflected sounds with high randomness for the direct sound, and the ASW increases. By contrast, because a car cabin is narrow and composed of complicated shapes and materials, and owing to a restriction of the loudspeaker mounting position and direction, a direct sound and reflected sound both coexist. Because IACC, DICC, and BSPL may change over time in such a car cabin audio listening environment, stable estimation results might not be obtainable.



(a) Concert hall (small hall)

(b) Car cabin

FIGURE 1. Example of impulse responses

3. **Proposed Method.** With the proposed method, a binaural model for SIA estimation in a car cabin audio listening environment was used and expanded into a model capable of estimating the SIW.

A block diagram of the proposed method is shown in Figure 2. The calculation method is explained below. First, the left and right ear input signals $l(n)$ and $r(n)$ are passed through the ear canal and middle ear filter then through a gammachirp auditory filterbank (GCFB) [11] to obtain the filterbank outputs $l(n, z)$ and $r(n, z)$. n is the time index, and z is the frequency channel of the auditory filter. GCFB enables the analysis by the time-frequency resolution of the multiple structures, which is effective for the analysis of the steep time change [7, 12-14], and was adopted in the model, because the auditory sense simulation accuracy is high. Then, for $l(n, z)$ and $r(n, z)$, energy was obtained for each time frame T, and the left and right averages were collected to obtain $E(n_T, z)$. The importance of the attack of sound is described as a cue regarding the SIA [2, 15]. The time-frequency region up to n_T, where the energy per z reaches the maximum value, is defined as the attack region. Thereafter, the interaural level difference (ILD) $\xi(n_T, z)$ and interaural phase difference (IPD) $\theta(n_T, z)$ were calculated within the attack region of $l(n, z)$ and $r(n, z)$, and the search object directions $D_\xi(n_T, z, \phi)$ and $D_\theta(n_T, z, \phi)$ were calculated through a comparison with a previously constructed head-related transfer function (HRTF) database. Here, ϕ is the azimuth. To obtain the search object direction $D(n_T, z, \phi)$, the weighted average is calculated using the weight $\beta(z)$ as in Equation (4).

$$D(n_T, z, \phi) = (1 - \beta(z))D_\xi(n_T, z, \phi) + \beta(z)D_\theta(n_T, z, \phi) \tag{4}$$

$D(n_T, z, \phi)$ was then multiplied by the weight function $E(z)$ proportional to the power of the GCFB output signal. The direction showing the mean of the marginal distribution $P(\phi)$ was made to be an SIA estimation $\hat{\phi}$.

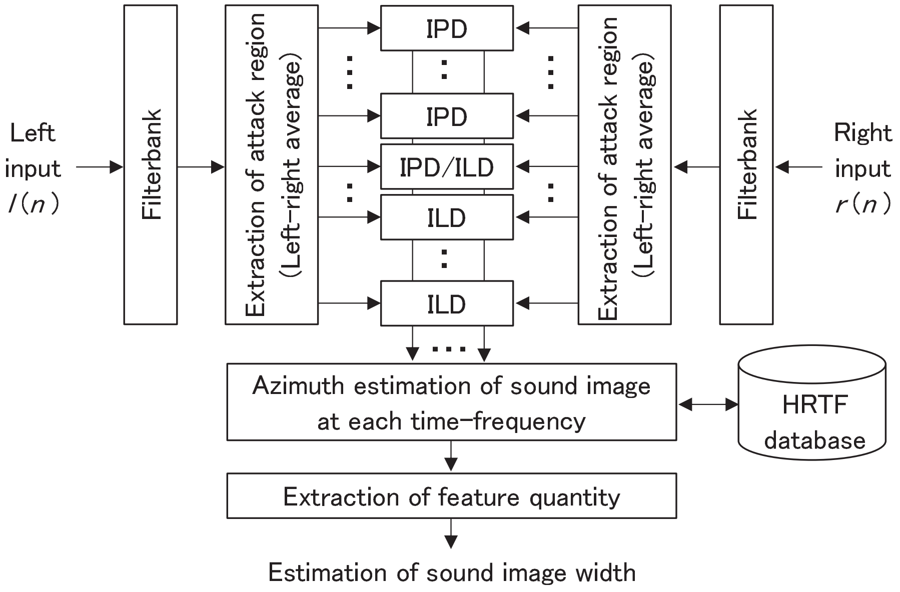

Left input $l(n)$
Filterbank
Extraction of attack region (Left-right average)
IPD
IPD
IPD/ILD
ILD
ILD
Extraction of attack region (Left-right average)
Filterbank
Right input $r(n)$
Azimuth estimation of sound image at each time-frequency
HRTF database
Extraction of feature quantity
Estimation of sound image width

Figure 2. Block diagram of the proposed method

An attempt to extract the physical features of the SIW estimation was made. Because the mean value of $P(\phi)$, $\hat{\phi}$, was confirmed to correspond to the azimuth of the sound image center in the SIA estimation model, the elements of the $P(\phi)$ distribution might correspond to the elements of the sound image shape, and the width of the $P(\phi)$ distribution might correspond to the SIW. Moreover, because direct and reflected sounds coexist and binaural input signals are random, and because the left and right loudspeakers are in asymmetry with the listening position, $P(\phi)$ might have a non-normal distribution.

Therefore, $P(\phi)$ was divided into left and right directions from the mean direction $\hat{\phi}$; in addition, the directions showing the mean for each divided distribution, $\hat{\phi}_{\mathrm{L}}$ and $\hat{\phi}_{\mathrm{R}}$, were obtained. Consequently, the difference between $\hat{\phi}_{\mathrm{L}}$ and $\hat{\phi}_{\mathrm{R}}$ was calculated as $\hat{\phi}_{\mathrm{LR}}$.

$$\hat{\phi}_{\mathrm{LR}} = \hat{\phi}_{\mathrm{R}} - \hat{\phi}_{\mathrm{L}} = \sum_{\phi=\hat{\phi}}^{+90} P(\phi)\phi \Bigg/ \sum_{\phi=\hat{\phi}}^{+90} P(\phi) - \sum_{\phi=-90}^{\hat{\phi}} P(\phi)\phi \Bigg/ \sum_{\phi=-90}^{\hat{\phi}} P(\phi) \tag{5}$$

Subsequently, the SIW was estimated using the nonlinear regression equation with the estimated value of $\hat{\phi}_{\mathrm{LR}}$ as an independent variable and the experimental value as a dependent variable.

4. Car Cabin Experiment.

The physical measurements and psychological experiments in the car cabin were carried out. The conditions were made to be at the time when the car was stationary.

4.1. Physical measurements.

The vehicle evaluated was a right-hand drive car (Camry, Toyota) (inside length of 2.08 m, inside width of 1.525 m, and inside height of 1.21 m). The head and torso simulator (HATS) (Type 5930, Brüel & Kjær) was set in the driver's seat, and the time-stretched pulse (TSP) was reproduced for both the left and right reproduction channels. Thus, the impulse response was measured using an audio sound analyzer (ASA-2, Etani Electronics). The sampling frequency was 48 kHz, and the synchronous addition was performed eight times using a genuine front door woofer and an instrument panel tweeter (Lch, Rch). In addition, the front door woofer is mounted at the listener's feet facing inwards, and the instrument panel tweeter is mounted upwards at the left and right ends of the vehicle, creating a sound field where direct sound and reflected sound coexist. Sound sources that controlled the inter-channel level difference (ICLD) between the left and right channels for the following evaluation sound sources are shown in Table 1.

1) White Noise, Band Noise (WN, BN: 2-12 kHz)
   Using a WN (AUDIO TEST CD-1, Japan Audio Society), a BN was made through an FIR-type bandpass filter with a filter degree of 512, a passband attenuation of −6 dB, and stopband attenuation of −60 dB. The signal length was 50 ms, and fade-in and fade-out processing were carried out using a raised-cosine ramp in the first 10 ms and last 10 ms of the 50 ms signal length, respectively.
2) Harmonic complex tone (HCT1: $f_0 = 250$ Hz, HCT2: $f_0 = 1$ kHz)
   Harmonic structure spectra such as music, speech and sign sound, which are objects in a car audio reproduction, were simulated. The second through the seventh harmonics were synthesized at a fundamental frequency of 250 Hz or 1 kHz because the seventh order HCT has already been used in previous study [16]. The harmonic level was −5 times the order +5 dB for the fundamental frequency level. The signal length and fade-in and fade-out processing are the same as those used in the WN.
3) Speech (S)
   Japanese Narration, Female (Impact 2, Japan Audio Society)'s "Haru" (0 min 35.8 s to 0 min 36.3 s) for 500 ms.
4) Music (M)
   Classical music was selected as a practical sound source [6] in the ASW evaluation. Mozart's Symphony No. 41 studio recordings (fourth movement, 94 bars to the 2 s section) were converted into a monophonic source.

Here, because the sound image may change with the signal length of the sound source, the signal length was made to be as short as possible but each sound source could still be recognized. The BSPL of the WN and HCT1 100 ms, S 500 ms, and M 2 s sections (ICLD = 0) measured by HATS was 73 dB. WN and HCT1 of short sound source signal

Table 1. Condition for car cabin experiment

Experiment No.	ICLD (Rch-Lch) [dB]
1	Lch only
2	−6
3	0
4	6
5	Rch only

lengths were made to be values of the interval including the reverberation in the car cabin. On the physical measurement result, the ICLD control sound source resampled at 48 kHz in sampling frequency and impulse responses (4800 taps) were convolved for each left and right reproduction channels; after the signal of every channel was synthesized, they were input to the model, and the estimate was obtained.

In the model calculation, because the microphone position of the HATS used is the ear canal entrance, ear canal [17] and middle ear characteristic approximation filters were designed as described in [18]. Here, z of the proposed method was made with an equivalent rectangular bandwidth (ERB) number of 4 to 33 (frequency is 123 to 7743 Hz) while the number of channels is 30 so that auditory filters with ERB may be arranged at equal intervals. T is the time width corresponding to one wavelength of the center frequency of each z. In Equation (4), $\beta(z) = 1$ below the center frequency of 750 Hz, $\beta(z) = 0$ above 1.5 kHz, and the frequencies between them were linearly changed from $\beta(z) = 1$ to 0. In the HRTF database, the HATS used in the car cabin experiment was used to correspond to the observation system: (i) the impulse response from the loudspeaker (TD510ZMK2, Denso Ten) to HATS was measured in an anechoic chamber, and (ii) the impulse response was measured by setting a microphone (Type 4191, Brüel & Kjær) at the center head position of HATS, instead of using HATS as is, and the HRTF was obtained based on the complex number division of the Fourier transformation results. The object was the range in which the front face of the HATS was set at 0° and was horizontally rotated by 5° from −90° in the left direction to +90° in the right direction. The TSP was used for measurement, and the sampling frequency was set to 48 kHz. The synchronous addition was performed eight times. Then, the database was constructed by passing the head-related impulse response (HRIR) through the GCFB.

4.2. **Psychological experiment.** In the psychological experiment, a sound source (ICLD = 0) was presented to the subject as a reference for each evaluation sound source. The ICLD control sound sources were presented in random order. The reference sound source was presented for each experiment, and both the reference and ICLD control sound sources were repeatedly presented until the subject signaled. Then, the perceived sound image was written in the evaluation paper of Figure 3, and the SIW angles of the elliptical shape of the sound image were totaled. The subjects were five males and two females, aged 20 to 50 years, and the ear position of all subjects was unified with the ear position of the HATS in physical measurement during listening.

The results of the psychological experiments are shown in Figure 4, depicting a box-and-whisker plot and showing the angle of the SIW for each experiment number in Table 1, for each evaluation sound source. The box-and-whisker plot, which expresses asymmetry and outliers of the distribution, was used. The box has a quartile range, and the lines in the box have a median. The maximum whisker length is 1.5 times that of the quartile range, and "+" indicates an outlier. It can be seen that the LR loudspeaker reproduction conditions of experiment numbers 2-4 tend to broaden the SIW, as shown in Figure 4. The sound sources BN and HCT2, which are composed of only high-frequency components, tend to have a narrower SIW than the other sound sources, and M has more reverberation

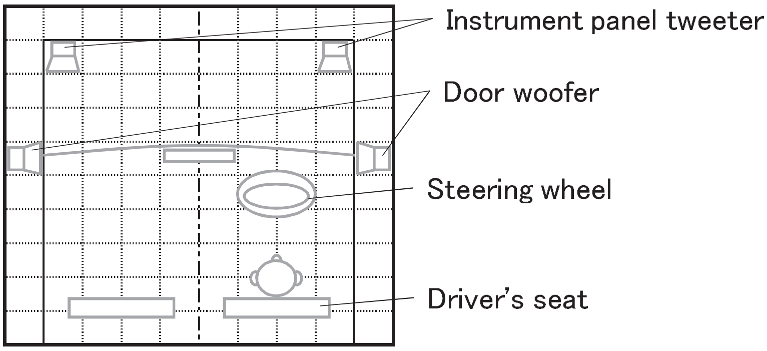
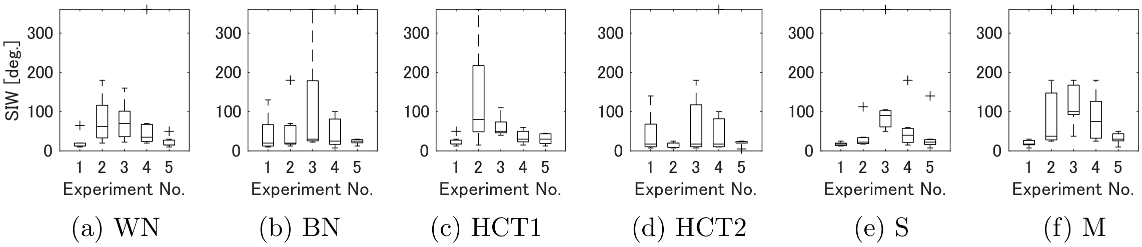


Figure 3. Evaluation paper (as viewed from the top of the car cabin)



(a) WN (b) BN (c) HCT1 (d) HCT2 (e) S (f) M

Figure 4. Car cabin psychological experiment results: box-and-whisker plot

components in the sound sources and tends to have a wider SIW than the other sound sources. In addition, the summed value may become an extremely large value 360° through the lateralization of a small number of persons under certain conditions. The median value, which is only slightly affected by the asymmetry and outliers of such a distribution, was used as a representative value of the experimental result.

5. **Model Validation.** The proposed method was compared with the conventional methods using the data obtained from the car cabin experiment. The conventional methods (CM) CM1 and CM2 are linear regression models with IACC and DICC as independent variables and experimental values as dependent variables, and CM3 is a multiple regression model, as shown in Equation (3). The proposed method (PM) is a nonlinear regression model in which $\hat{\phi}_{\mathrm{LR}}$ is an independent variable and the experimental value is a dependent variable. These data were normalized using the min-max normalization (MMN). As a result of the regression analysis for each model, the statistical significance was not obtained for CM1 through CM3. Because CM3 was constructed under specified classical music listening conditions, reconstruction may be necessary to apply it to the car cabin audio listening environment. A multiple regression analysis was then conducted using the experimental data with the same independent variables, and a statistical significance was not obtained. The results are discussed below. It was suggested that the lower the IACC, the lower the estimation accuracy of the ASW, and that the estimation value scatters significantly for IACC < 0.5 [19]. Moreover, the discrimination limit increases as the DICC decreases [6]. Because the direct sound and reflected sound coexist in the car cabin experiment, the interaural correlations were extremely low (IACC, mean $M = 0.41$, standard deviation $SD = 0.18$; DICC, $M = 0.47$, $SD = 0.23$), which is a factor that resulted in a low estimation accuracy. Furthermore, with the conventional methods, it was suggested that among the energy components of the reflected sound against the direct sound, the component that does not exceed the upper limit (%-split) of the separation of the sound image that satisfies the law of the first wavefront contributes to the ASW [20]. Compared to this, the car cabin experiment environment differs from the structure of the

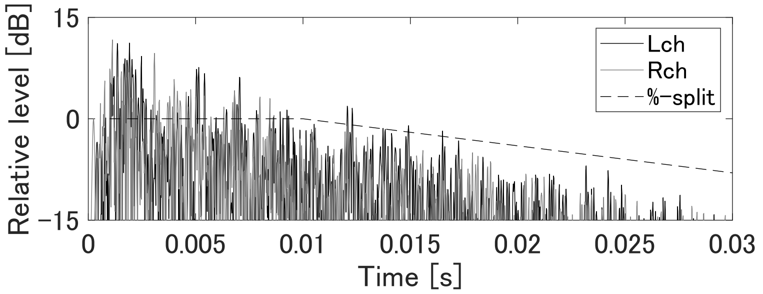
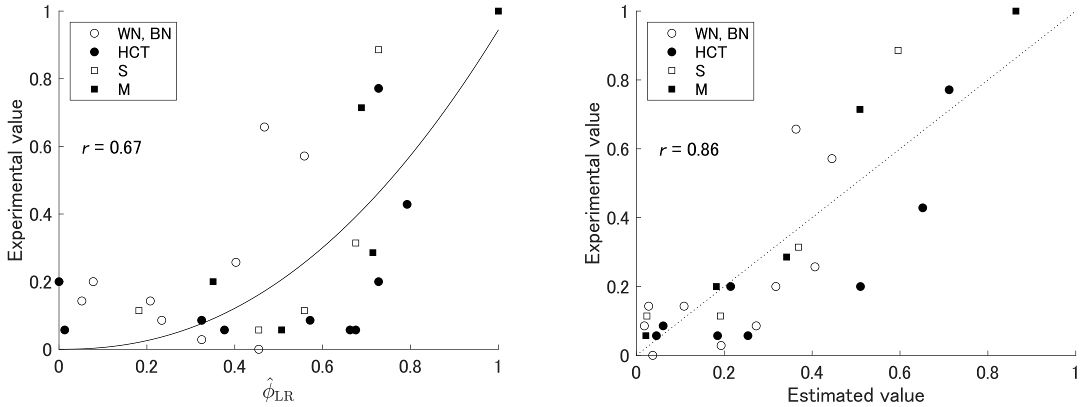


Figure 5. Impulse response and %-split



(a) PM

(b) Multiple regression with PM and SID

Figure 6. Relationship between estimated values and experimental values

direct and reflected sounds constructed using the conventional method. Figure 5 shows the impulse responses from a sound source to the HATS under the reproduced conditions during car cabin experiment No.2 and %-split for the speech applied. When the peak of the first wave of the impulse response is assumed to be the reference, i.e., 0 dB, the initial energy component of the mixture of both the direct and reflected sounds, which is dominant in the car cabin audio listening environment, exceeds the %-split. Moreover, the upper limit varies with the frequency and duration of the sound sources [21]. These factors are considered to be a reason for the decrease in the estimated accuracy of the conventional method applied under the car cabin audio listening conditions. Therefore, the construction of a new SIW estimation model in a car cabin audio listening environment is required. PM is expressed through Equation (6).

$$\mathrm{SIW} = a\hat{\phi}_{\mathrm{LR}}^{b} \tag{6}$$

Here, $a = 0.94$, $b = 2.24$. Statistical significance was obtained (a, $t(29) = 5.09$, $p < 0.001$; b, $t(29) = 4.11$, $p < 0.001$). The root mean squared error (RMSE) was 0.20. Figure 6(a) shows the relationship between $\hat{\phi}_{\mathrm{LR}}$ and the experimental values.

Thereafter, the estimated error is examined. In Figure 6(a), the experimental value tends to be larger than the estimated value under six conditions, and is 0.5 or greater. Because the experimental value of the SIL tends to be close to the listener under these conditions, a multiple regression analysis was carried out using the PM and experimental value of the sound image distance (SID) as the independent variable and the SIW as the dependent variable. Thus, the effect of the SID was verified. As a result, statistical significance was obtained (multiple regression coefficient, $F(2, 27) = 36.9$, $p < 0.001$; partial regression coefficient (PM), $t(27) = 4.03$, $p < 0.001$; partial regression coefficient (SID), $t(27) = -5.36$, $p < 0.001$; partial regression coefficient (constant term), $t(27) = 5.46$,

$p < 0.001$). The multiple regression equation has an adjusted coefficient of determination of $Rh = 0.71$. The RMSE was reduced to 0.14. Figure 6(b) shows the relationship between the calculated results using the multiple regression equation and the experimental values. The estimated error tended to be small, and the possibility that the SID affected the SIW estimation was indicated. In future studies, it is possible to improve the accuracy of the SIW estimation by verifying the model together with the SID. Thus, the PM seems to be more effective than the CMs in a car cabin audio listening environment.

6. **Conclusions.** The binaural model for the SIA estimation in the car cabin audio listening environment was expanded to a model that can estimate the SIW. As a result of comparative verification of the CM1 through CM3 and PM, the conventional methods were unable to achieve statistically significant results, whereas the proposed method achieved significant results. Thus, the significant results obtained in this study contribute to the SIL performance and improved reproduction capability of the surrounding sound in the on-board acoustic equipment development of the car audio by objectifying the sound image evaluation. This seems to be the basis of the sensitivity design in terms of sense of presence. In addition, if the acoustic design parameters are updated automatically, allowing the estimated value to reach closer to the desired characteristics, e.g., through artificial intelligence (AI), the sensitivity design can be automated. However, because in this study, the listening condition was limited to a car that was stationary, it is necessary to develop a model close to actual listening, such as by expanding the range of conditions of car models, sound sources, investigating when the car is operational, and expanding to a model in which SIL, including SID and sound image elevation (SIE), is possible.

REFERENCES

[1] K. Ozawa, S. Tsukahara, Y. Kinoshita and M. Morise, Development of an estimation model for instantaneous presence in audio-visual content, *IEICE Trans. Inf. & Syst.*, vol.E99-D, no.1, pp.120-127, 2016.

[2] K. Sakamoto, Azimuth estimation binaural model of sound image for audio listening environment in car cabin, *Trans. Jpn. Soc. Kansei Eng.*, vol.20, no.3, pp.285-289, 2021.

[3] V. K. Kukkala, J. Tunnell, S. Pasricha and T. Bradley, Advanced driver-assistance systems: A path toward autonomous vehicles, *IEEE Consum. Electron. Mag.*, vol.7, no.5, pp.18-25, 2018.

[4] K. Nakadai and H. Okuno, Robot audition and computational auditory scene analysis, *Advanced Intelligent Systems*, vol.2, no.9, pp.1-9, 2020.

[5] Y. Ando, S. Sato, T. Nakajima and M. Sakurai, Acoustic design of a concert hall applying the theory of subjective preference, and the acoustic measurement after construction, *Acta Acustica United with Acustica*, vol.83, no.4, pp.635-643, 1997.

[6] M. Morimoto and K. Iida, A practical evaluation method of auditory source width in concert halls, *J. Acoust. Soc. Jpn. (E)*, vol.16, no.2, pp.59-69, 1995.

[7] S. Ishimitsu, K. Sakamoto, T. Yoshimi, Y. Fujimoto and K. Kawasaki, Study on the visualization of the impression of button sounds, *International Journal of Innovative Computing, Information and Control*, vol.5, no.11(B), pp.4189-4203, 2009.

[8] Y. Soeta, S. Nakagawa, Y. Kamiya and M. Kamiya, Subjective preference for air-conditioner sounds inside a car in summer and winter, *Journal of Ergonomics*, vol.6, pp.1-7, 2016.

[9] S. Tatsukami, S. Ishimitsu, Y. Soeta and S. Nakagawa, Effects of active control of noise with music on subjective auditory impression and brain activity, *ICIC Express Letters, Part B: Applications*, vol.9, no.3, pp.195-201, 2018.

[10] A. Kanda, S. Ishimitsu, K. Wakamatsu, M. Nakashima and H. Yamanaka, Objective evaluation of sound quality for audio system in car, *ICIC Express Letters, Part B: Applications*, vol.10, no.4, pp.335-342, 2019.

[11] T. Irino and R. D. Patterson, A dynamic compressive gammachirp auditory filterbank, *IEEE Trans. Audio, Speech, and Lang. Process.*, vol.14, no.6, pp.2222-2232, 2006.

[12] G. Onishi, S. Ishimitsu, K. Sakamoto, T. Yoshimi, Y. Fujimoto and K. Kawasaki, Automatic evaluation of button sound impressions using a neural network, *ICIC Express Letters*, vol.4, no.3(A), pp.683-689, 2010.

[13] K. Sakamoto, S. Ishimitsu, K. Sugawara, T. Yoshimi and K. Sasaki, The study of audio equipment evaluations using the sound of music, *ICIC Express Letters, Part B: Applications*, vol.2, no.3, pp.597-602, 2011.

[14] K. Sakamoto, S. Ishimitsu, T. Arai, T. Yoshimi, Y. Fujimoto and K. Kawasaki, A study of evaluating the button sounds for car audio main units –First report, feature analysis using wavelet transform–, *Trans. Jpn. Soc. Kansei Eng.*, vol.10, no.3, pp.375-385, 2011.

[15] B. C. J. Moore, *An Introduction to the Psychology of Hearing*, 3rd Edition, Academic Press, 1989.

[16] J. R. Miller and E. C. Carterette, Perceptual space for musical structures, *J. Acoust. Soc. Am.*, vol.58, no.3, pp.711-720, 1975.

[17] K. Sugiyama, M. Nishimoto and M. Satoh, Transmission characteristics of ear canal of artificial head, *Acoust. Sci. & Tech.*, vol.26, no.1, pp.67-70, 2005.

[18] B. R. Glasberg and B. C. J. Moore, A model of loudness applicable to time-varying sounds, *J. Audio Eng. Soc.*, vol.50, no.5, pp.331-342, 2002.

[19] Y. Ando and Y. Kurihara, Nonlinear response in evaluating the subjective diffuseness of sound fields, *J. Acoust. Soc. Am.*, vol.80, no.3, pp.833-836, 1986.

[20] M. Morimoto, K. Nakagawa and K. Iida, The relation between spatial impression and the law of the first wavefront, *Applied Acoustics*, vol.69, no.2, pp.132-140, 2008.

[21] J. Blauert, *Spatial Hearing –The Psychophysics of Human Sound Localization–*, The MIT Press, 1997.