Proceedings of Meetings on Acoustics

Volume 19, 2013

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Architectural Acoustics
Session 2pAAa: Adapting, Enhancing, and Fictionalizing Room Acoustics II

2pAAa10. Physical and subjective factors of spatial envelopment impression of surround sound reproduction

Toru Kamekawa* and Atsushi Marui

*Corresponding author's address: Musical Creativities and the Environment, Tokyo University of the Arts, 1-25-1, Adachi-ku, 120-0034, Tokyo, Japan, Kamekawa@ms.geidai.ac.jp

How we feel envelopment? For the evaluation of spatial impression, `envelopment' is one of the key issues among several attributes of spatial impression. When the authors attempted to elicit common attributes from surround sound recordings by triadic elicitation procedure, three attributes, `brightness', `temporal separability', and `spatial homogeneity of envelopment' were elicited. In this paper, authors focused on `spatial homogeneity of envelopment' and considered related physical parameters such as IACC (Inter-aural Cross Correlation). From comparison between the result of subjective test and the data measured by a dummy-head facing toward different angles, the proposal method was defined as the IACC weighting of the 1/3 octave band level variation and its difference between frontal and rotated positions (0°, -30°, -15°, +15°, and +30°) of a dummy head.

Published by the Acoustical Society of America through the American Institute of Physics

©2013 Acoustical Society of America [DOI: 10.1121/1.4800238]
Received 25 Jan 2013; published 2 Jun 2013
INTRODUCTION

When we evaluate recording techniques for surround sound reproduction, many attributes are used such as brightness, powerfulness, and so on [1, 2, 3, 4, 5, 6]. Particularly, spatial impression is one of the most important factors when compared with conventional two-channel stereo reproduction. The typical attributes of spatial impression are the following (also described in Figure 1):

• Localization: the apparent location of the sound source,
• Depth: the apparent spatial distance between front-end and back-end of the sound source,
• Width: the width of frontal image similar to ASW (apparent source width),
• Envelopment: the enveloped feeling, surrounded laterally and from behind similar to LEV (listener envelopment) [7], and
• Presence: the impression of being at the actual performance.

However, it is difficult to evaluate the perceptual differences between actual music excerpts such as differences in microphone set up [5, 6]. The subjective evaluation test using these attributes often results in no significant differences, because while differences between the excerpts are small, the differences between impressions of these attributes among listeners are relatively large. In order to evaluate on more subtle differences between the sound stimuli while enabling listeners to share the common meanings for each attribute descriptors, following three attributes were elicited using triadic elicitation procedure [11]:

• brightness,
• temporal separability, and
• spatial homogeneity.

From the authors' previous report, it is difficult for naive listeners to describe spatial impression compared to timber or temporal impressions [8]. Although for professionals such as recording
engineers who are familiar to surround sound, spatial impression is first priority for making surround sound reproduction. Above all "spatial homogeneity" is one of the most important attributes for describing spatial impressions.

“Spatial homogeneity” is defined as following;

surrounded by sound which comes from all direction, and each sound is not identified and linked seamlessly and homogeneity.

Right panel of Figure 2 shows an image of sound environment having high "spatial homogeneity" while left panel having less.

**Figure 2:** The explanatory images of s"patial homogeneity". Right panel shows higher spatial homogeneity than the left panel.

**PHYSICAL FACTORS RELATED TO SPATIAL IMPRESSIONS**

**IACC**

ASW (apparent source width) and LEV (listener envelopment) are well known attributes to evaluate spatial impression, and IACC (Inter-aural cross correlation coefficient) is used to explain these impressions. IACC is calculated by the equation (1) [9].

\[
IACC = \frac{\int_{t_1}^{t_2} p_l(t)p_r(t-\tau)dt}{\sqrt{\int_{t_1}^{t_2} p_l^2(t)dt \int_{t_1}^{t_2} p_r^2(t-\tau)dt}}_{max}
\]

(for -1 ms < \(\tau\) < +1 ms)

The \(\tau\) is defined as ITD (Inter-aural time difference) between the left and the right sound signals, recorded with a dummy-head, when the correlation of the two signals is at maxima within ±1 ms time window. It is considered that when the arrival direction of a sound source is located specified position, the \(\tau\) value becomes constant.

The concept of "spatial homogeneity" seems to be comparable to LEV. IACC is calculated from the full band signal of IR (impulse response) measured by dummy-head in a spatial condition. Morimoto, et al. recommend to use IACC calculated on a signal divided into octave bands [7]. Meanwhile Anazawa [10] and Griesinger [11] proposed the methods to use 1/3-octave bands for calculating IACC for musical signals.

The calculation interval (\(T = t_2 - t_1\)) is set to 30 ms and the analysis interval is set to half of \(T\), i.e.15 ms, in this paper. It was referred that the boundary value of perception between pitch and time is 33 ms [13].
THE ANALYSIS USING SURROUND SOUND SOURCES

Recording method

Five music excerpts recorded in a medium-sized concert hall (Casals Hall, Tokyo) were used. They were recorded by the microphone array “Omni+8”, which consists of four omni-directional microphones and one figure-8 microphone [14] (figure 3). The position of each microphone was set by a professional recording engineer’s ears (called the ‘reference’). Additional samples were recorded by moving the microphone arrays 30 cm from the reference position to back, forward, up, and down. The automated player piano was employed to play “Fantasie Impromptu” by Chopin (bars 29 to 42), which resulted in a total of five samples that were used in the subsequent experiment.

![Microphone Setup](image)

**Figure 3:** The layout of sample recording using Omni+8, which is consisted from two omni-directional microphones for left and right and a figure-8 for center, two omni-directional microphones for rear. Right panel shows five kinds of microphone placement for the recording at Casals Hall. Unit of the values are in centimeters.

Comparison to subjective listening test

The listening test was conducted at the Sound Production Studio located at the Senju Campus of Tokyo University of the Arts (a.k.a. Tokyo Geidai). The studio was built under the listening room specifications in ITU-R BS 1116. The five stimuli mentioned above were presented via five active full bandwidth loudspeakers (Genelec 8050) arranged according to ITU-R BS 775-1, with a height of 1.2 m from the floor and with a radius of 2.6 m from the central listening position (no LFE signal was prepared). Calibration of loudspeaker level was performed using an audio analyzer (Klark Teknik DN6400 with NTI N2010 microphone) with A-weight and fast response. Each loudspeaker output was individually calibrated to 79 dBA using a $-10$ dBFS pink noise input signal, giving 85 dBA for the total of all five speakers. Average listening level of each music excerpt was adjusted ca. 85 dB (LAeq).

The experiment system was implemented via Scheffe’s Paired Comparison test using the “A-B Comparison mode” of the “STEP” program by Audio Research Labs [16]. A listener compared spatial homogeneity and made a two-comparison forced-choice test methodology using a 7-point response scale.

Twenty-one subjects, including 11 students and 10 professional recording engineers and researchers, evaluated all linear combinations of the five recorded excerpts ($5 \times (5 - 1)/2 = 10$ pairs). Durations of the samples were approximately 20 seconds. A subject could listen to each piece until he/she was satisfied with the evaluation of each pair.
Statistical analysis

From the results of the analysis of variance (Nakaya variation of Scheffe’s ANOVA [17]) by all subjects, significant differences between microphone placements and effects of subjects’ individual differences were observed for all attributes.

To check the error in each subject, the existence or non-existence of circular triads was checked [17, 18]. Circular triads are caused by internal inconstancy when the subject rated A higher than B, B higher than C, and C higher than A.

The subjects who had more than two circular triads were removed from the list. Finally, 17 out of 21 subjects remained. The analysis of variance (ANOVA) after removing the subjects with inconstancies still had the effect of subjects’ individual differences, however. So, remaining subjects were divided into two or three groups by cluster analysis until the effect of subjects’ individual differences had no significant difference. Computing of cluster analysis was done using the “kmeans” command in Matlab.

Final result is shown in figure 4. From the analysis of variance after clustering significant differences between microphone placements and effects of subjects’ individual differences were observed.

![Comparison of the microphone placements regarding 'Spatial Homogeneity' ratings obtained by the subjective experiment. Whiskers indicate 95% confidence interval. The vertical axis indicates the score of each sample based on 7-point response scale.](image)

**FIGURE 4:** Comparison of the microphone placements regarding ‘Spatial Homogeneity’ ratings obtained by the subjective experiment. Whiskers indicate 95% confidence interval. The vertical axis indicates the score of each sample based on 7-point response scale.

**PHYSICAL FACTORS OF SPATIAL HOMOGENEITY**

**Comparison of IACC**

From the result of subjective experiment, the physical difference between two sounds which was able to recognize the difference was considered.

The measurement was conducted at the same room as the listening test. A dummy-head (B&K HATS Type 4128) was used to measure IACC. To adapt various sound sources, impulse responses of each position were measured using swept sine, and the dummy-head recording sources were obtained by convolution technique.

Figure 5 shows the time variation of IACC of the impulse response of the microphone arrays located reference position (after this, abbreviated “Ref”) and forward position (after this, abbreviated “For”) that had the difference by subjectivity evaluation. A clear difference is not seen from these graphs.
**Consideration of the difference arising from the decay level of each frequency band**

Figure 6 shows the time variation of 1/3 octave band level (rms) of each the impulse response. The authors assumed that the effect of IACC is connected with the level of a corresponding frequency band.

In consideration of decreasing with time of the impulse response and difference of each 1/3 octave band level, alternative index performed weighting of the level variation was considered (2).

\[ f(t)_{IACC} = (1 - IACC_{1/3oct}) \cdot L_{1/3oct} \]

\( IACC_{1/3oct} \) and \( L_{1/3oct} \) mean the IACC and level of 1/3 octave band corresponding to the time \( t \).

**Consideration of the difference from the direction angle of dummy-head**

Usually IACC is calculated to the signals recorded with a dummy-head facing towards the sound source. If “spatial homogeneity” is the condition that arrival direction of sound is
nonconstant, it would be necessary to measure not only front direction but some rotated
direction of dummy-head as well. A dummy-head was set at five rotated positions (0°, −30°,
−15°, +15°, and +30°) indicated in figure 7.

**Figure 7:** Image of the measurement of plural point of IACC

Figure 8 shows time variation of $f(t)_{IACC}$ corresponding each microphone array.

**Figure 8:** The time variation of $f(t)_{IACC}$ of each angle of dummy-head of the microphone arrays of the “Ref” (left) and “For” (right) position

To compare “spatial homogeneity”, following function is defined from the differences of the
time variation of $f(t)_{IACC}$ corresponding each angle of dummy-head (3).

$$f_{SH} = \frac{\int_{T_1}^{T_2} \sum_{n=1}^{n_a} |f_{IACC_n}(t) - f_{IACC\_median}(t)| \cdot f_{IACC\_median}(t) dt}{n_a} \times 100 \% \quad (3)$$

$f_{IACC\_median}(t)$ means the median of the value of all the angles at time $t$. $T_1$ and $T_2$ mean
start time and end time of the calculation section of an impulse response ($T_1=20$ ms,
$T_2=750$ ms). $n_a$ means the number of the position of dummy-head. In this case, $n_a = 5$.

Figure 9 shows the time variation of proposed index comparing each microphone array. The
values with that integrated from 20 msec to 750 msec are following;

$$f_{SH}(Ref) = 7.6\%$$
$$f_{SH}(For) = 13.1\%$$

This value is considered that the lower one has high spatial homogeneity, and these values
 correspond with the result of the evaluation experiment.
CONCLUSION

For the evaluation of spatial impression, “spatial homogeneity” was discussed. To compare “spatial homogeneity” of music sources, the proposal method was defined as the IACC weighting of the 1/3 octave band level variation and its difference between frontal and rotated positions (0°, –30°, –15°, +15°, and +30°) of a dummy head. Pairwise comparison was implemented to evaluate “spatial homogeneity” using five different microphone placements for surround sound recording. From the results of ANOVA, significant differences between microphone placements and interaction among subjects were observed. To reduce significant differences in the interaction of subjects, subjects who had circular triads were removed, and the rest were divided into two or three groups by cluster analysis. Finally, one microphone array showed significant differences.

The proposed index value was calculated from the impulse response corresponding a pair of microphone array positions which is seen the significant difference. The value is considered that the lower one has high spatial homogeneity, and these values correspond with the result of the evaluation experiment.

This study is still in progress, and the final goal will be the establishment of training methods for evaluating spatial impression of surround sound recordings. As for the next step, confirmation of the appropriateness of the elicited attributes is planned. The authors will conduct further listening experiments using music excerpts controlling the time variation of the 1/3-octave band of IACC and/or other relevant physical parameters, and research the relationship between these physical parameters of “spatial homogeneity” and “diffuseness” of room acoustics in ecological condition.

REFERENCES


