AES Surround Study Group
Evaluation Tests Report Part2
Presented at the AES Surround Recording Experiments Project Report
2006-2007 AES Japan.

Correspondence Relationship between Physical Factors and Psychological Impressions of Microphone Arrays for Orchestra Recording

Toru Kamekawa 1, Atsushi Marui 2, and Hideo Irimajiri 3

1 Tokyo National University of Fine Arts and Music, Adachi-ku, Tokyo, 120-0034, Japan
kameawa@ms.geidai.ac.jp

2 Tokyo National University of Fine Arts and Music, Adachi-ku, Tokyo, 120-0034, Japan
marui@ms.geidai.ac.jp

3 Mainichi Broadcasting Corporation, Kita-ku, Osaka, 530-8304, Japan
iririn@mbs.co.jp

ABSTRACT

Microphone technique for surround sound recording of an orchestra is discussed. Eight types of well known microphone array recorded in a concert hall were compared in subjective listening test on seven attributes such as spaciousness, powerfulness and localization using a method inspired by MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor). The result of the experiment shows similarity and dissimilarity between each microphone array. It is estimated that directivity of a microphone and distance between each microphone are related to the character of microphone array and these similarities are changed by music character. The relations of the physical factors of each array were also compared, such as SC (Spectral Centroid), LFC (Lateral Fraction Coefficient), and IACC (Inter Aural Cross-correlation Coefficient) from the impulse response of each array or recordings by a dummy head. The correlation of these physical factors and the attribute scores show that the contribution of these physical factors depends on music and its temporal change

1. INTRODUCTION

Recently, 5.1 channel surround sound system, one of the multichannel stereophonic audio reproduction systems, has acquired stable popularity in movie theatres, DVDs, and digital broadcasting. There are a lot of reports introducing methods and ways to record surround sound [1–6]. However, it has been very difficult to compare between multiple microphone techniques under the same condition due to one or more of the following constraints; namely, the cost of collecting a large number of microphones in one hall on one day, time and space constraints of setting up all microphones on a stage, and mixing timbral/spatial differences created by different microphone techniques. In this paper, the simultaneous recordings done by the authors for orchestral sound in a concert hall using several microphone settings are explained, and comparison between characteristics of microphone techniques are discussed.

2. OUTLINE OF RECORDING

The recording session was held at the Symphony Hall in Osaka on September 25th to 27th, 2006, performed by Osaka Philharmonic Orchestra conducted by Shigeo Genda.

Figure 1 shows the layout of the main and ambience microphones. There were eight types of main microphone arrays, seven types of ambience microphone arrays, and 40 spot microphones, placing 98 microphones in the hall in total. These microphones were connected to microphone amplifier (Studer 962) and recorded on two sets of Digidesign ProTools HD in 96kHz sampling frequency and 24bits resolutions.

To provide the optimum sound of each microphone array, the microphones were placed at optimum locations sought through discussion among several recording experts. Regarding the playback system, two sets of 5.1 channel surround monitoring systems were built to quickly compare and check between microphone arrays. The layouts of loudspeakers were based on the recommendation of ITU-R BS775-1 [7].

Since there were a number of different opinions about what termed as the "orchestral sound," several types of orchestral pieces, explained in the next section, were performed to discuss their differences and similarities in sound qualities.

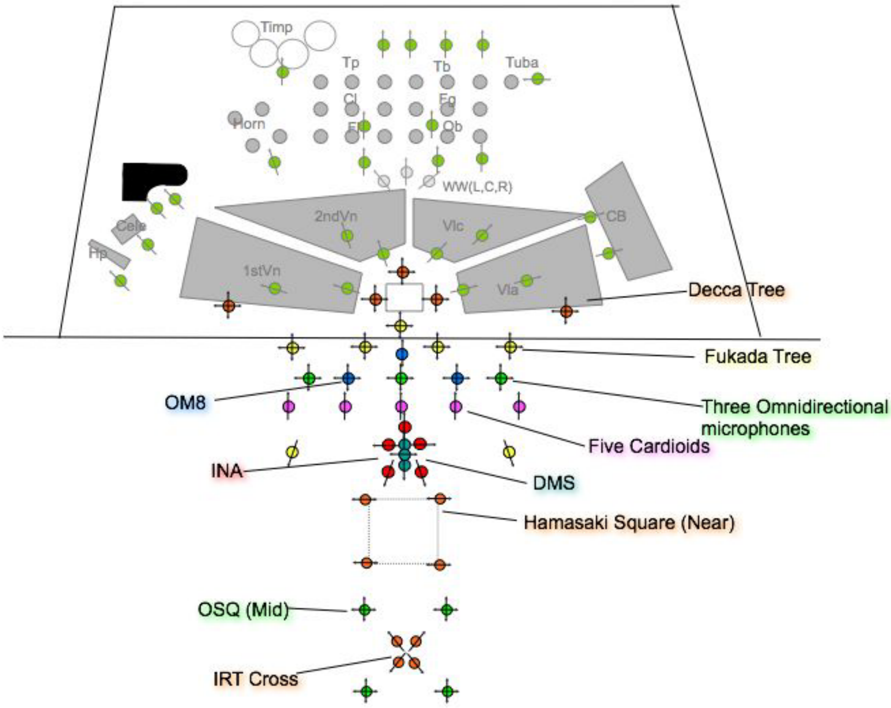


Figure 1 Layout of main and ambience microphones for the experiment

3. LISTENING EXPERIMENT

3.1. Microphone arrays for evaluation

For the subjective listening experiment, the authors chose following eight sets of main microphones.

1. Fukada Tree (Fukada)
2. INA5 (INA)
3. Double MS (DMS)
4. Omni+8 (OM8)
5. Decca Tree + Omni Square (DT+OSQ)
6. Decca Tree + Hamasaki Square (DT+HSQ)
7. Five Cardioids + Hamasaki Square (5C+HSQ)
8. Three Omnis + IRT Cross (3O+IRT)

The detail of each microphone array is as indicated in Appendix.The method used in the current subjective listening experiment was inspired by MUSHRA (MUltiple Stimuli with Hidden Reference and Anchor) [8], where each set of microphone array was assigned to a vertical slider enabling subjects to rate eight microphone sets simultaneously. The method was *inspired* in a sense that there were no specific anchor and reference.

3.2. Attributes for evaluation

Each microphone array was evaluated by following seven types of attributes.

1. *Spaciousness* (Spc): The width of frontal image.

2. *Envelopment* (Env): The enveloped feeling surrounded lateral and backward.

3. *Depth* (Dep): The apparent spatial distance of the sound source from the listener.

4. *Localization* (Loc): The apparent location of the sound source.

5. *Powerfulness* (Pow): Strong or heavy impression. Opposite meaning is "week" or "feeble."

6. *Softness* (Sof): "Mild" or "silky" impression. Opposite meaning is "hard" or "harsh".

7. *Preference* (Pref)

The ratings of eight microphone arrays were done for one attribute chosen in random order from the seven attributes. The exception was *preference* which was presented as the final attribute of the session for all subjects.

3.3. Playback environment

The experiment was conducted at Sound Production Studio located in Senju Campus of Tokyo National University of Fine Arts and Music (Tokyo Geidai). The studio was built under the listening room specification in ITU-R.BS1116. The 24 five-channel stimuli (eight microphone array by three pieces) were presented via five active full bandwidth loudspeakers (Genelec 8040) arranged according to ITU-R BS 775-1[7],with 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 Audio Analyzer (Klark Teknik DN6400 ,NTI N2010 microphone) with A-weight and fast response. Each loudspeaker output was individually calibrated to 79dB SPL using a -10dBFS pink noise input signal, giving 85dB SPL for the summed output of all five speakers. The Protools HD (Digidesign) with mixing console (D-command) was used for the surround sound reproduction system in 96kHz/24bits.

Twenty-two subjects included 13 students and 9 recording experts. Eight of faders on the mixing console (Figure 2) were used to indicate the order of each microphone array for each attribute. Using "Solo" switch on each fader, a subject could select each stimuli.

The following three pieces were chosen to have captured a wide variety of microphone characteristics for the evaluation.

- Overture from "The marriage of Figaro" *(Mozart)* : standard classical piece
- The beginning of "Pines of Rome" *(Respighi)* : large orchestra and less low pitch instruments
- Middle part of "Wellington's Victory" *(Beethoven)* : located small 'brass band' at the backward of audience seat

Lengths of each piece were approximately one minute. A subject could listen to each piece until he/she is satisfied in evaluating all eight microphone arrays.

Figure 2 Eight faders for evaluation of each microphone array

4. RESULT AND ANALYSIS OF THE EXPERIMENT

4.1. Comparison of the average score

Figure 3, 4 and 5 show the average score of each microphone array on each attribute.

Regarding "Figaro," DT+OSQ (Decca Tree with Omni Square) seemed to mark a high score at almost all attributes except *spaciousness* and *localization*. DMS (Double MS) has scored relatively low on all attributes. However other two pieces showed deferent results.

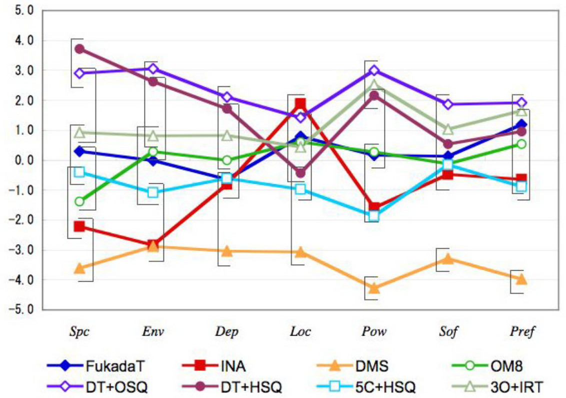


Figure 3 Average score of each microphone array of "The marriage of Figaro"

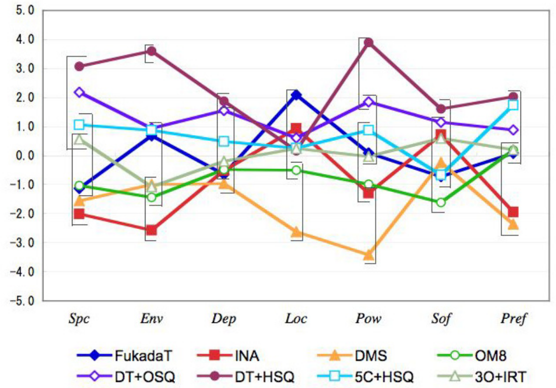


Figure 4 Average score of each microphone array of "Pines of Rome"

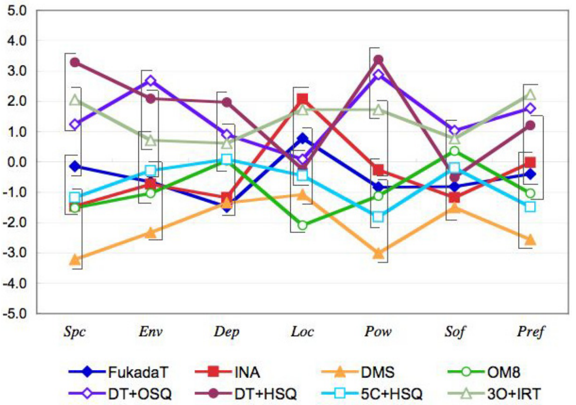


Figure 5 Average score of each microphone array of "Wellington's Victory"

The brackets in the figures show groupings of stimuli based on 95% confidence intervals. There were some groups of microphone arrays distinguished significantly within some of the attributes such as *spaciousness* and *powerfulness*, *softness*, and *preference* at "Figaro." But there were different results at other two pieces. In all three pieces, there were significant differences between each microphone array regarding *spaciousness* and *powerfulness*, but there was no difference on *depth* and *localization*. Also, the differences on *envelopment*, *softness*, and *preference* were dependent on the musical characters.

4.2. INDSCAL analysis

To compare similarity and difference of each microphone array, INDSCAL (individual difference scaling) analysis was conducted on correlation coefficients computed from the subjects' overall score of each microphone array. Typical use of INDSCAL analysis in psychoacoustics is to derive a map of stimuli from psychological distances obtained from global dissimilarity ratings. The assumption behind our use of correlation coefficients is that correlation values computed from *all* subjective attribute ratings are highly related to global dissimilarities in psychological domain.

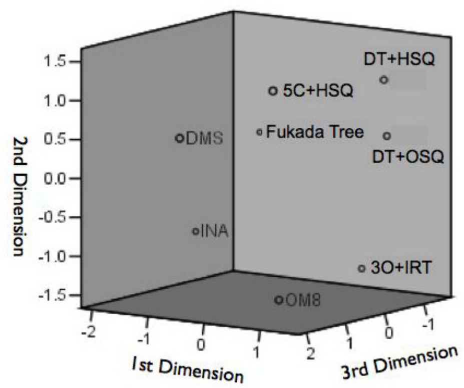


Figure 6 The spatial configuration of eight microphone array on three dimension regarding "The marriage of Figaro"

Figure 6 shows the spatial configuration (stimulus space) of eight microphone arrays of "Figaro" on three dimensions.

The interval of each microphone array means their similarity. Since it is hard to find these intervals from this 3-dimmension figure, it is converted two of 2-dimmensional layouts.

Figure 7 indicates two-dimensional layout of 1st against 2nd and 1st against 3rd dimension.

From the stepwise multiple regression analysis with each of the INDSCAL dimensions as dependent variable and seven attribute scores from all subjects as independent variables, the relation between each dimension and attributes were calculated as following;

1st dimension: *envelopment spaciousness*,
2nd dimension: *spaciousness*, *preference*,
3rd dimension: *powerfulness, spaciousness*,.

Left panel in the Figure 7 shows three groups from microphone arrays' similarity; Fukada-Tree and DMS, 3O+IRT and OM8, and DT+HSQ, DT+OSQ and 5C+HSQ. However, right panel in Figure 7 shows that these grouping do not hold on different axes. Especially, DT+HSQ and 5C+HSQ which use the same HSQ as ambience array step away from each other. Otherwise, DT+OSQ and DT+HSQ which using same frontal array keep each similarity.

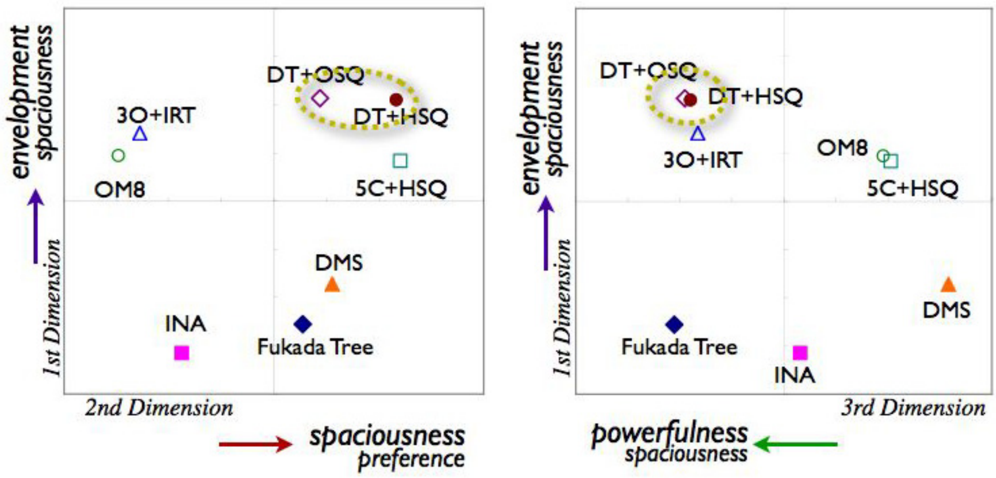


Figure 7 Two-dimension layout of 1st to 2nd and 1st to 3rd regarding "The marriage of Figaro"

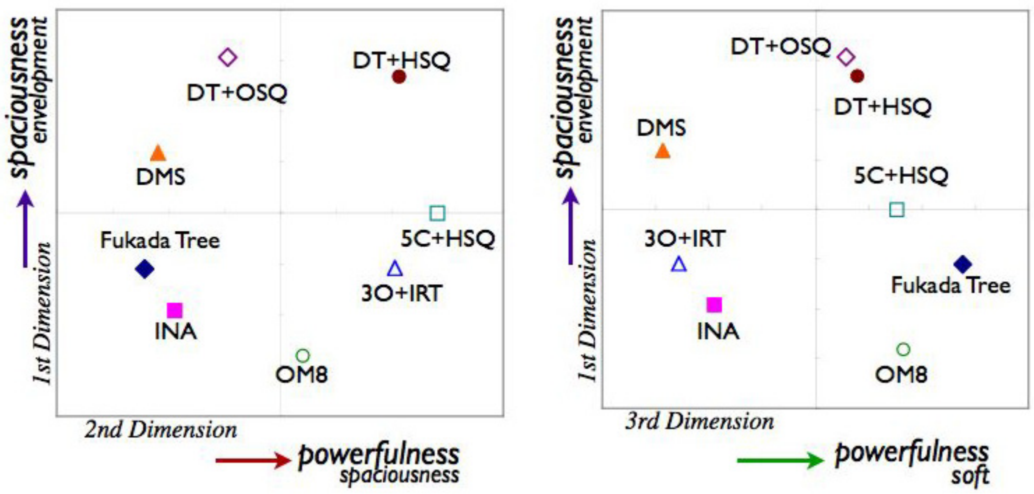


Figure 8 Two-dimension layouts of 1st to 2nd and 1st to 3rd regarding "Pines of Rome"

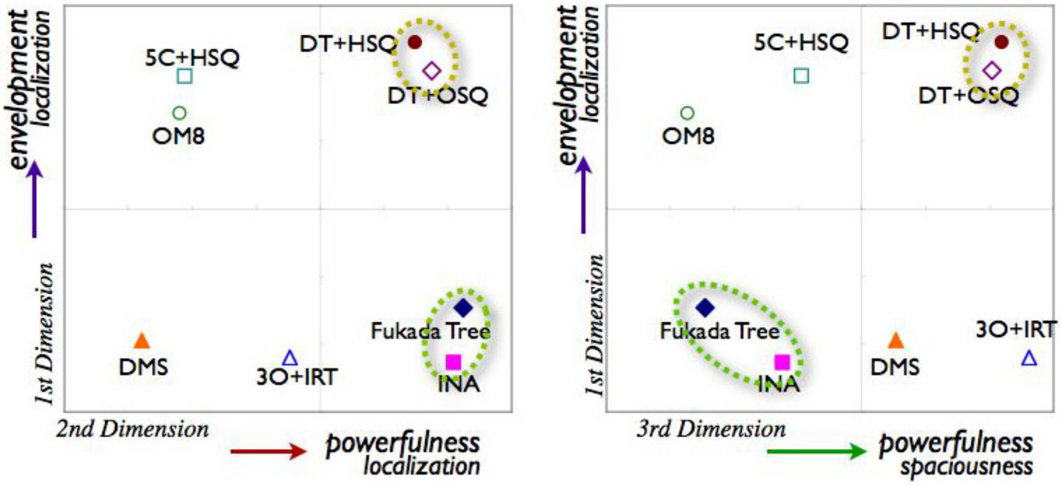


Figure 9 Two-dimension layouts of 1st to 2nd and 1st to 3rd regarding "Wellington's Victory"

Figure 8 indicates the same analysis regarding *"Pine of Rome"*. The result of the multiple regression analysis is following;

1st dimension: *spaciousness*, *envelopment*,
2nd dimension: *powerfulness, spaciousness*,
3rd dimension: *powerfulness*, *softness*.

The factors of 2nd and 3rd dimensions were *powerfulness* and they were different from the "*Figaro*" case. This is suppose the character of music affected the impression of them. Left panel in the Figure 8 show lower similarity of DT+OSQ and DT+HSQ compared to "Figaro".

Figure 9 indicates same analysis regarding *"Wellington's Victory"*. The result of the multiple regression analysis is following;
1st dimension: *envelopment*, *localization*,
2nd dimension: *powerfulness, localization*,
3rd dimension: *powerfulness*, *spaciousness*.

Since this music has Banda, a small band of brasses, woods, and percussion, at rear side of the listener, the factor of 1st-dimension is *envelopment* and *localization.*

From these results *envelopment* is most related to 1st dimension. Compared to "Figaro", this piece placed the importance on *powerfulness* rather than *spaciousness*.

The arrays that microphones for surround channels are relatively close to the frontal microphones such as Fukada Tree and INA, have higher similarity on both dimensions compared to other pieces.

4.3. Comparison of physical factor and psychological factor

4.3.1. SC and LFC

From the result of IR (Impulse Response) of each microphone array, SC (Spectral Centroid) and LFC (Lateral Fraction Coefficient) of each array were calculated from obtained IR at the microphones[9]. SC was expected to be related with the timbral attributes such as *powerfulness* and *softness*. While LFC was expected that it is related to the attribute of spatial impression such as *spaciousness*, *envelopment*, and *depth*.

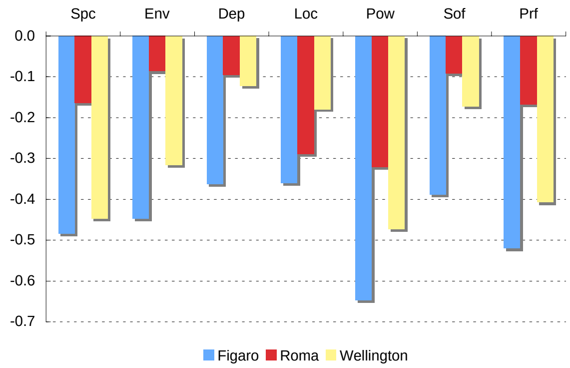


Figure 10 Correlation coefficient of SC calculated from IR and each attribute of each piece

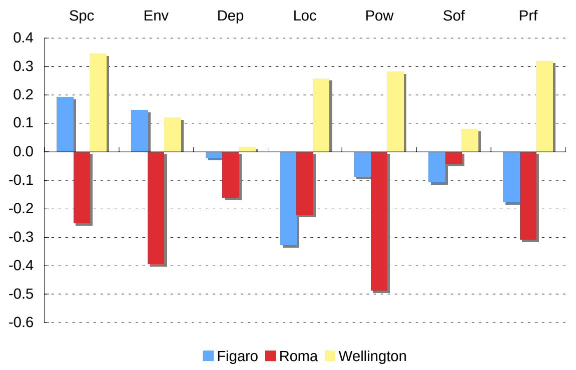


Figure 11 Correlation coefficient of LFC calculated from IR and each attribute of each piece

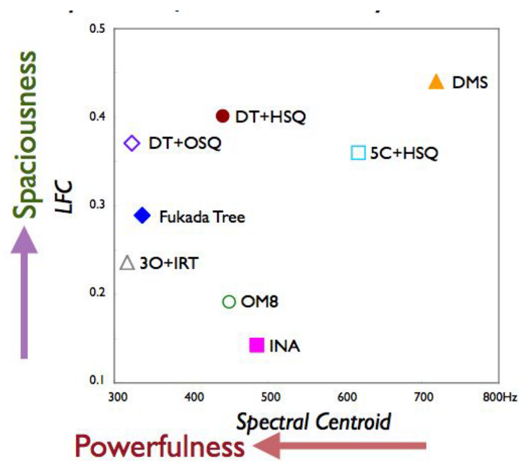


Figure 12 The relation between SC and LFC of each array calculated from IR

Figure 13 The recording sean using a dummy head microphone in the same environment

Figure 10 shows correlation of SC calculated from IR and the score of the arrays evaluated each attribute. SC correlates inversely with powerfulness. Figure 11 shows same as LFC, and it has weak correlation with spaciousness on 'Figaro' and 'Wellington'. However it shows inverse correlation regarding 'Pines of Rome'.

Figure 12 shows the relation between SC and LFC of each array. The arrays using omini-directional microphone such as DT+OSQ and DT+HSQ show similarity and would be expected spaciousness and powerfulness. Otherwise DMS seems to be relatively different from other arrays.

4.3.2. SC and IACC from the Dummy head recording

To compare the physical factor of each piece, IACC (Inter-Aural Cross-correlation Coefficient) and SC were calculated on the recordings done using the dummy head microphone (Head Acoustics) at the same playback condition as the listening experiment (Figure 13). The correlation coefficient between these physical factors and the score of each attribute by every subject was computed and are shown in Figure 14 and 15.

Since the values of IACC and SC are changing dynamically, the representative value of IACC and SC were calculated by the average through the pieces.SC shows strongish inverse correlation with powerfulness regarding 'Figaro' and 'Wellington'. IACC shows weak inverse correlation with spaciousness and envelopment except 'Wellington'.

The reason of these results are expected that these music change momentarily.

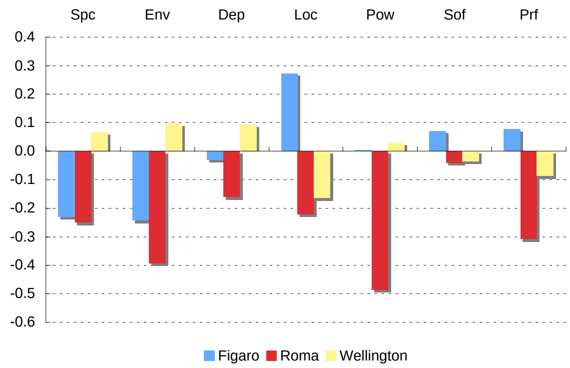


Figure 14 Correlation coefficient of IACC recorded by dummy head microphone and each attribute of each piece

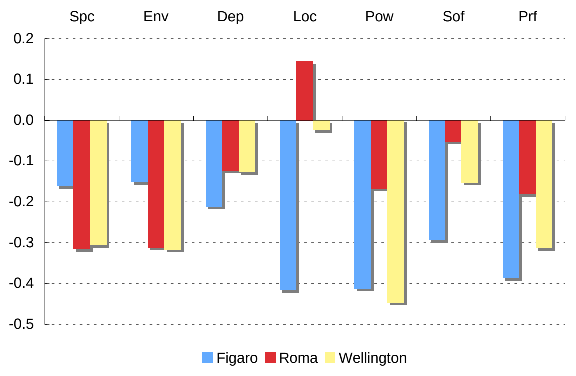


Figure 15 Correlation coefficient of SC IACC recorded by dummy head microphone and each attribute of each piece

4.3.3. Temporal change of SC and IACC of each pieces

The temporal change of SC of each array are shown in Figure 16. The purple waveforms on the figures indicate temporal transition change of correlation between SC value of each second and score of powerfulness. Since the SC has inverse correlation with powerfulness, this waveform plotted negative value to upside.

At first glance, powerfulness shows high correlation when the SC value moved to low frequency. It is estimated that powerfulness is evaluated when the music characters tend to low pitch.

Figure 17 ,18, and 19 show the temporal transitional change of IACC of each array. There is the tendency that the IACC value of the arrays using omni-directional microphones, are relatively lower regarding 'FIgaro'. However other two pieces show more complicated shapes.

The red waveforms on the figure 20 indicate temporal transition change of correlation between IACC value of each second and score of envelopment plotted negative value to upside. And the blue lines indicate temporal level change of each piece averaged each 0.5s.

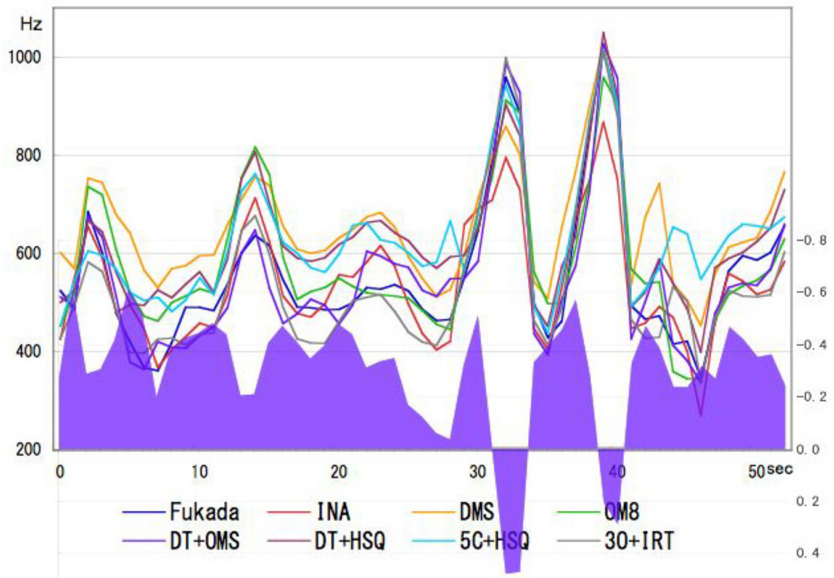


Figure 16 Temporal transitional change of SC of "The Marriage of Figaro" (colored lines) and the correlation between temporal SC values from dummy head recording and the scores of each array regarding *powerfulness* (purple waveform).

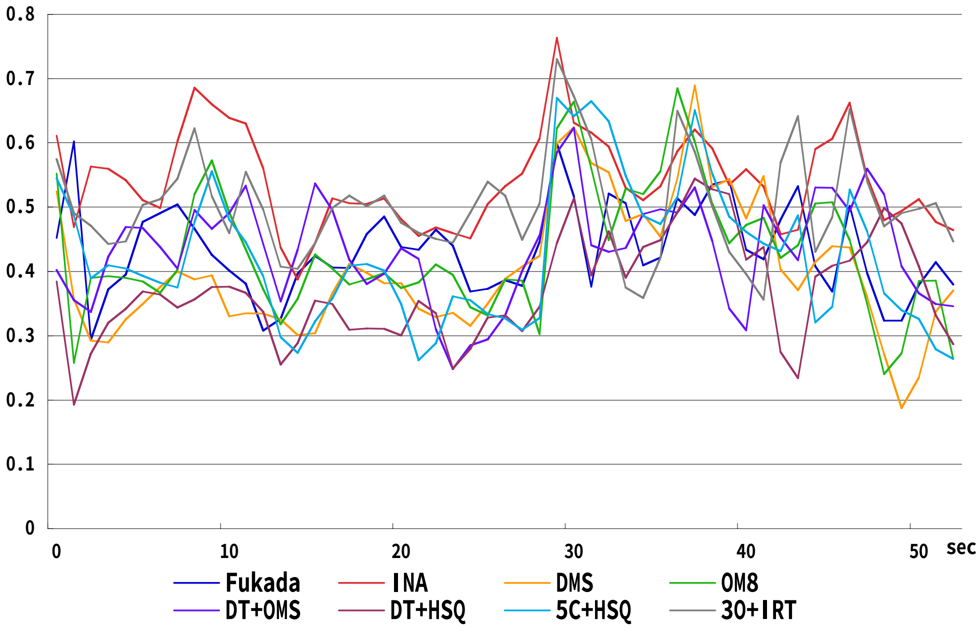


Figure 17 The temporal transitional change of IACC of each array found from dummy head regarding "The marriage of Figaro". The horizontal axis is time (seconds) and the vertical axis is IACC.

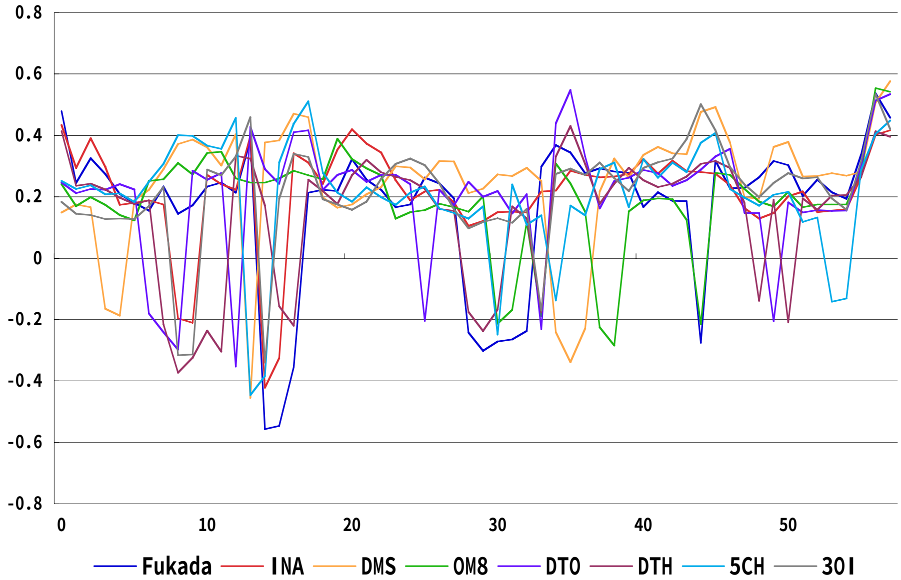


Figure 18 The temporal transitional change of IACC of each array found from dummy head regarding "Pine of Rome". The horizontal axis is time (seconds) and the vertical axis is IACC.

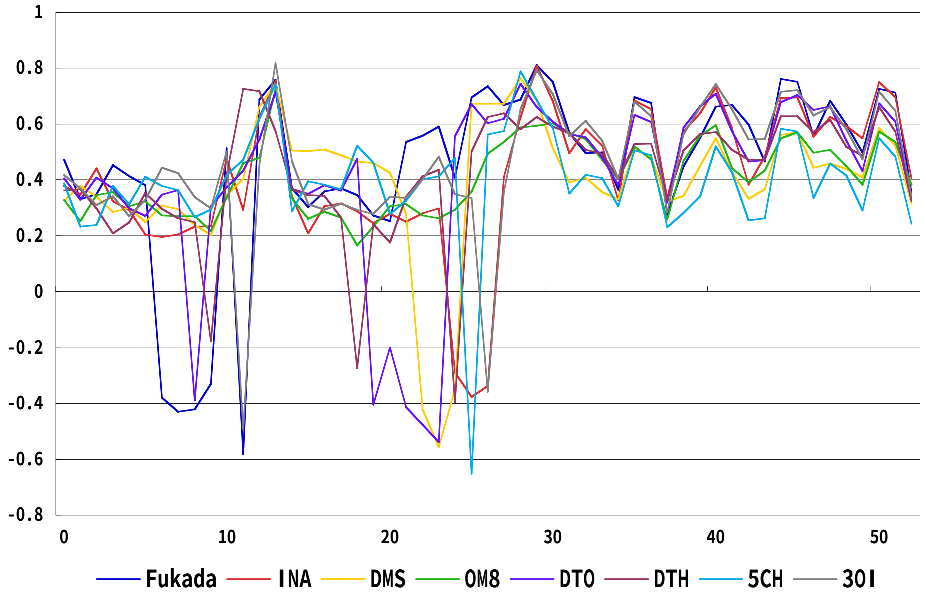


Figure 19 The temporal transitional change of IACC of each array found from dummy head regarding "Wellington's Victory". The horizontal axis is time (seconds) and the vertical axis is IACC.

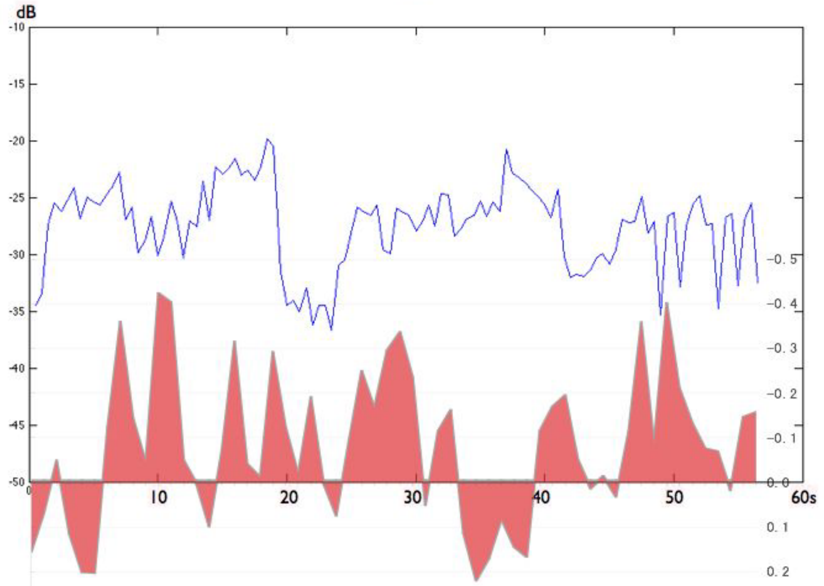


Figure 20 Temporal level change of "Pines of Rome" (blue line) and the correlation between temporal IACC values from dummy head recording and the scores of each array regarding *envelopment* (red waveform).

5. CONCLUSION

Several well-known microphone arrays for surround recoding of orchestra were discussed. The result of listening experiment shows that each microphone array has different character and it is related to the configuration of each microphone array. Furthermore these characters depend on impression of the character of music. Previous studies done by Martens and Kim et al also have been reported similar results [10-12].

To discuss character of music, SC and LFC were calculated from impulse response of each microphone array and SC and IACC from the data of all stimuli recorded by dummy head microphone were also found. From the correlation coefficient of these physical parameters and the scores of each attribute, the contribution of these physical factors to each attribute would change by music. Especially SC and IACC from dummy head were changing dynamically through the music. These changes are considered as a key factor in the variation of each subject.

From the temporal correlation between SC or IACC and corresponded attributes, following assumptions are drawn;

- ＊ Impression of powerfulness is evaluated at the music characters tend to low pitch.
- ＊ Impression of envelopment is evaluated at the part of 'afterglow' which the loudness of music changing from loud to soft.

The physical predictors such as LFC and IACC were originally developed to predict the spatial impression of a room acoustics, and thus more investigations on physical predictors for musical pieces may be necessary to improve the study.

6. ACKNOWLEDGEMENTS

This work was supported by Hoso-Bunka Foundation and many companies that agreed to cooperate on this project.

The authors make an address of gratitude to all participants in the experiment.

7. REFERENCES

[1] Francis Rumsey, "Spatial Audio" Chap.2 ,Focal press, 2001

[2] Michael Williams and Guillaume Le Dti, "The Quick Reference Guide to Multichannel Microphone Arrays Part 1: using Cardioid Microphones" AES 110th Convention Amsterdam Preprint 5336, 2001

[3] Michael Williams and Guillaume Le Dti , "MULTICHANNEL MICROPHONE ARRAY DESIGN" 108th AES Convention Paris Preprint 5157 , 2000

[4] Francis Rumsey, " Spatial Audio " Chap.7 (p192) , Focal press , 2001

[5] Kamekawa, "Impression Differences by Placement of Front and Rear Microphones for Multichannel Stereo Recording" AES 118 Convention Barcelona Preprint , 2005

[6] Kamekawa, "The Effect on Spatial Impression of the Configuration and Directivity if Three Frontal Microphones Used in Multi-channel Stereophonic" AES 28th International Conference Pitea, Sweden Proceedings, 2006

[7] Recommendation ITU-R BS.775: Multichannel stereophonic sound system with or without accompanying picture. International Telecommunication Union , 1993

[8] Recommendation ITU-R BS.1534-1: Method for the subjective assessment of intermediate quality level of cording systems , 2001-2003

[9] Hanyṳ, "Room acoustical parameters" , Journal of Acoustic Society of Japan, Vol. 60-2 pp.72-77 , 2002

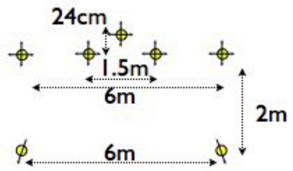
[10] Martens and Kim , "Verbal Elicitation and Scale Construction for Evaluating Perceptual Differences between Four Multichannel Microphone Techniques" AES 122nd Convention Vienna Preprint 2007

[11] William Martens et al, “Investigating Contextual Dependency in a Pairwise Preference Choice Task” AES 28th International Conference Pitea, Sweden ,Proceedings, 2006

[12] Sungyoung Kim et al, “An Examination of the Influence of Musical Selection on Listener Preferences for Multichannel Microphone Technique” AES 28th International Conference Pitea, Sweden ,Proceedings, 2006

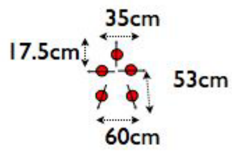
8. APPENDIX:

The detail of eight types microphone array using this experiment are following;



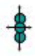
1. Fukada Tree

Five omni-directional microphones for front and two uni-directional (cardioid) microphones for rear.



2. INA (INA5)

Five cardioid microphones configured as star shape.

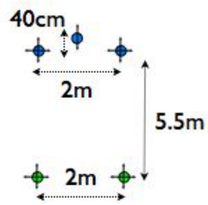


3. DMS (Double MS)

Two cardioid microphones located at front(Mf) and rear(Mr) and a bi-directional microphone (figure of 8) placed at lateral direction(S), producing 4 channels by following equations ;

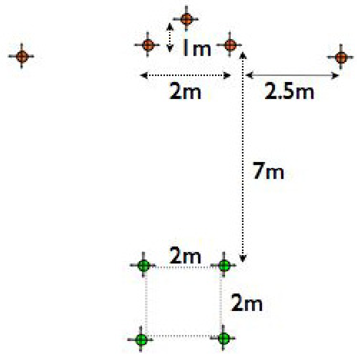
L=Mf+S, R=Mf-S

Ls=Mr+S, Rs=Mr-S

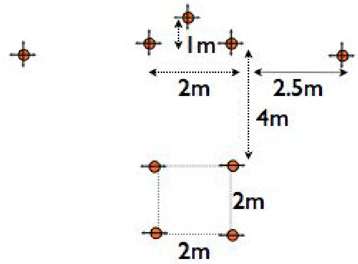


4. OM8 (Omni+Figure of 8)

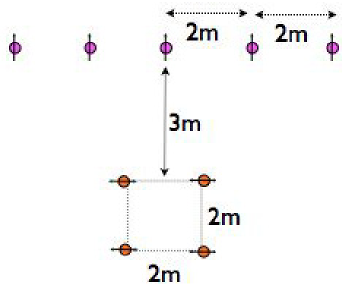
Two omni-directional microphones for left and right and a figure of 8 for center, two omni-directional microphones for rear.



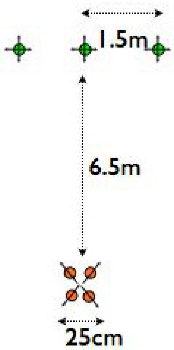
5. DT + OSQ
Five omni-directional microphones so called 'Decca Tree' as frontal array and four omni-directional microphones located as square shape apart from frontal array



6. DT + HSQ
Five omni-directional microphones (so called 'Decca Tree') as frontal array and four figure of 8s configured in square shape (so called 'Hamasaki-Square') apart from frontal array



7. 5C+ HSQ
Five cardioids as frontal array and four figure of 8s configured in square shape (Hamasaki Square) apart from frontal array



8. 3O + IRT-cross
Three Omni-directional microphones as frontal array and four cardioids located as 25cm square shape apart from frontal array