



AES Surround Study Group

Evaluation Tests Report Part 3

Presented at the AES Surround Recording Experiments Project Report
2006–2007 AES Japan Section

各種サラウンド・マイクアレイにおける主観評価テスト パート3

各種サラウンドマイクアレイ印象の試聴者による差異に関する主観評価実験

Atsushi Marui¹, Toru Kamekawa¹, and Hideo Irimajiri²

¹ Department of Musical Creativity and the Environment, Faculty of Music,
Tokyo National University of Fine Arts and Music, Tokyo, Japan
marui@ms.geidai.ac.jp
kamekawa@ms.geidai.ac.jp

² Mainichi Broadcasting System, Inc. Osaka, Japan
iririn@mbs.co.jp

ABSTRACT

In order to investigate how recorded sounds of different surround sound microphone arrays are perceived, subjective evaluation of two classical music pieces recorded using seven microphone arrays were conducted in Fukuoka, Osaka, Tokyo, and Vienna. The listening tests were done using Scheffé's pairwise rating method on *powerfulness*, *width*, *softness*, and *preference*. Although, in general, the ratings were highly dependent on a choice of music piece, high and low ratings on *powerfulness* were seen for Decca Tree with Omni Square array and Double M/S array, respectively for both of the pieces used in the experiments. Also, prediction equation of *preference* was formulated from perceptual attributes, and the equation showed that higher *powerfulness*, higher *width*, and lower *softness* yielded higher *preference*.

1. マイクアレイの印象に関する主観評価実験 <日本編>

各サラウンド・マイクアレイによって録音された音源が、試聴者にどのような心理的印象を想起させるかを調べる為、2007年3月に、福岡、大阪、東京の3カ所にてアンケート調査および主観評価実験を行なった。本報告においては、主観評価実験の手法および分析結果を報告する。

本実験の主な目的は、サラウンド・マイクアレイの違いが、音楽の試聴者にどのような心理的印象を与えるかを調査する事であった。また、試聴者の職種や背景によって評価に偏りがある

のか、聴取位置によって印象が異なるのかなどを調べる目的もあった。単印象（迫力、拡がり、柔らかさ）から全体印象（好み）の推定も試みた。

1.1. 音刺激

実験に用いた音刺激は、2種類の音楽を7種類のサラウンド・マイクアレイを用いて収録したもので、合計14個であった。2種類の音楽は「ローマの松」および「ウェリントンの勝利」であった。選曲に関しては、パート1を参照して頂きたい。

実験に用いた音刺激は、7種類のマイクアレイによって録音された音源である。7種類のマ

イクアレイは、それぞれの楽曲の特徴をよく表現できているもの、音質に特徴が強く表れているもの、すでに音質に関する一定の評価を得ているもの、これまで評価される事が少なかったもの、等を総合的に考慮して筆者らが主観的に選んだ以下のものである。

- Fukada Tree
- Decca Tree & Omni Square Mid
- 3 Omni & IRT Cross
- 5 Cardioid & Hamasaki Square Near
- Omni 8
- INA 5
- Double MS

なお、マイク・アレイに関しても、楽曲と同様、パート1において詳細に説明されている為、ここでは省略する。

1.2. 手法

実験には「シェッフェ (H. Scheffé) の一対比較法」を用いた。一対比較法は複数個の試料(音刺激)から2個ずつ取り出して比較し、最終的に全ての試料の相対的評価を行なう方法である[1]。

シェッフェの方法に限った事ではないが、一対比較法の弱点として、試料の組み合わせの数が多くなってしまふ事が挙げられる。たとえば k 個の試料がある時、一対の音刺激を(A-BとB-Aの様)に順番を変えて2回比較がされるとすると、全ての組み合わせは $k(k-1)$ 対になり、試料数の増加に伴う組み合わせの数は多項式的に増える事が判かる。今回の実験においてはマイクアレイを7種類選んだので、 $7 \times (7-1) = 42$ 回の比較を行なう必要があった。しかし、今回は、被験者の拘束時間を長くできない反面、被験者数が比較的多く得られる事が予想されていた為、一対比較における全組み合わせを複数被験者に分散し、各被験者の負担を減らす事にした。具体的には、各被験者はA-B比較とB-A比較のどちらか一方(一対比較行列の上半分もしくは下半分のみ)について評価する事、評価する形容詞を被験者ごとに替えること事、の2点である。

一回の実験には最大で14名がスイート・スポットを中心に着席し、同時に実験が行なわれた。この為、スイート・スポットから若干外れた場所にて試験を行なった被験者もいた。

1.3. 被験者

被験者は福岡で54人、大阪で87人、東京で111人の計252人であったが、うち11人のデータに明確な記入ミスや無回答があった為、241名のデータのみを分析した。年齢は19歳～76歳(平均39.3歳/標準偏差11.1歳)であった。男女比は、87対13であった。

被験者本人の自己申告による職種は、学生や専業主婦/主夫から放送技術スタッフや録音機材の開発者まで、多岐にわたった。本報告においては、放送関連のオーディオ専門職、非放送関連のオーディオ専門職、非オーディオ専門職の3つに分類し、分析に用いた。

1.4. 教示

「シェッフェの一対比較法」を使った実験では、音刺激に対する形容詞評価を行なう為、形容詞の選定も実験計画の重要な要因となる。今回の実験においては「迫力」「拡がり」「柔らかさ」「好み」の4通りを選んだ。「迫力」「柔らかさ」「好み」の3つはそれぞれ、音色表現因子として様々な研究者により確認されている迫力因子、金属性因子、美的因子に対応したものである。そこに空間的な印象を表す代表的な単語として「拡がり」を加えた。

ヒトが一对の刺激の比較を行なう時には、ある特定の感覚属性において比較する場合と、全体的な印象において比較する場合がある。特定の感覚属性において比較する、とは、例えば2個の試料を「明るさ」や「大きさ」等の単独の感覚属性のみに着目して比較するものである。全体的な印象の比較においては、それら単独の属性すべてを考慮した、「好み」や「ふさわしさ」などの統合的な判断がなされる[2]。例えば、今回の実験に用いた4語のうち、「好み」は全体の印象、それ以外の3つは単独の感覚属性を問うものである。

シェッフェの方法は、特定属性における比較にも全体的な印象比較にも用いる事が可能である為、今回の実験においては、単独の感覚属性および全体印象の両方に関して質問をした。ただし、印象の偏りを避ける為、全体印象に関する質問は必ず単独の感覚属性に関する質問より後に行なわれた。

被験者には次のような教示を与えた。

この聴き比べでは1問につき2つの音を聴いて頂き、1つ目に聴いた音に対して2つ目に聴いた音がどうであったかをお答え頂きます。各問いに対し1~5の当てはまるものをお答え下さい。

- 最初に聴いた音に比べ後の音の方が、迫力が(1:ない、2:ややない、3:どちらともいえない、4:ややある、5:ある)
- 最初に聴いた音に比べ後の音の方が、迫力が(1:ない、2:ややない、3:どちらともいえない、4:ややある、5:ある)

以上の教示には「迫力」について比較する部分のみを記載したが、実際に比較評価に用いた形容詞は、実験では「迫力」「拡がり」「柔らかさ」「好み」の4通りであった。

1.5. 分析結果

「シェッフェの対比較法」による分散分析を行った結果、2曲ともにすべての形容詞において、1%の危険率でマイクアレイの違いに有意差が現れた。それぞれのマイク・アレイを使った録音物の試聴印象をFigure 1(「ローマの松」)、及びFigure 2(「ウェリントンの戦い」)に示す。棒グラフはそれぞれ評価点の平均値と95%信頼区間を表している。

信頼区間は、実験に用いた尺度値と親和性がよい区間推定値を与えてくれ、帰無仮説有意検定で用いられる p 値とのつながりも存在する為、アメリカ心理学会の「スタイル・マニュアル」においても報告が推奨されている[4, p.18]。「95%信頼区間」とは「母数を含む確率が95%である区間」という意味であるが、今回の報告においては、仮説検定による平均値の差の検定ではなく、同等の検定力を持ちつつも信頼区間を利用した直感的な方法を紹介し、対比較法の結果を見ていく。

CummingとFinchによって提案された方法[3]は、信頼区間を目視し、そこから検定とほぼ同等の結果を推論できる方法である。彼らの論文には7つの法則が挙げられているが、その中から最も簡単に平均値の有意差を確認できる法則を紹介する。なお、これ以降「統計的に有意」

という時は、特に断らない限り $p<.05$ レベルである。

【第4法則】第1群の95%信頼区間の中に第2群の平均値が含まれていない時、両群の平均値には $p<.05$ で統計的有意差がある。また、両群の信頼区間が重ならない時、両群の平均値には $p<.01$ で統計的有意差がある。(両群の標本数が10以上かつ両群が独立の時。)

この法則によると、「ローマの松」(Figure 1)においては、Decca Tree & Omni Square Mid、及び5 Cardioid & Hamasaki Square Nearの「迫力」があり、反対にDouble MSは「迫力」がない、と言える。「拡がり」に関しては、5 Cardioid & Hamasaki Square NearとDouble MSの対を除く全てのマイク・アレイにおいて統計的に有意な差は見られなかった。「柔らかさ」については、Double MSに対して有意差が認められるのはFukada Tree、5Cardioid & Hamasaki Square、Omni 8、INA 5の4アレイである。「好み」に関してはDouble MSに対してFukada Treeを除く5つのマイク・アレイが統計的に有意であった。

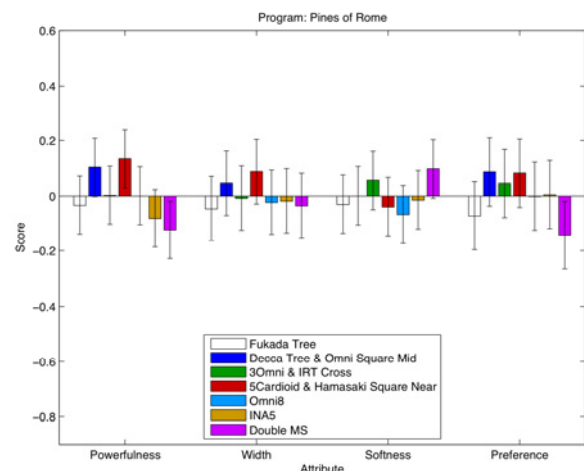


Figure 1: Scheffé's pairwise comparison analysis results from "Pines of Rome" for Japanese participants. Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

「ウェリントンの戦い」(Figure 2)においては、「ローマの松」よりもはっきりとした評価がなされている。まずDecca Tree & Omni Square Midに「迫力」があり、続く3 Omni & IRT CrossとOmni 8もそれぞれ統計的に有意に「迫力」の差異が認められる。5 Cardioids & Hamasaki Square NearとDouble MSは互いに有意差はないが、他

のマイク・アレイとは有意な「迫力」の違いがあった。

同様に、マイク・アレイ同士の平均値と信頼区間とを比較する事により、「拡がり」においてもDecca Tree & Omni Square Midと3 Omni & IRT Crossは5 Cardioids & Hamasaki Square Near及びDouble MSと統計的有意差がある事が判かる。「柔らかさ」に関しては3 Omni & IRT CrossとDouble MSとの間で小さいながらも有意差が見られるが、その他のアレイ対では有意差はなかった。

5 Cardioids & Hamasaki Square NearとDouble MSは互いに有意差はないが、他のマイク・アレイとは有意な「好み」の違いがあった。この点に関しては「迫力」や「拡がり」と似た差異の出方になっている事から、「迫力」、及び「拡がり」と「好み」の間に何らかの関係があるとも考えられる。

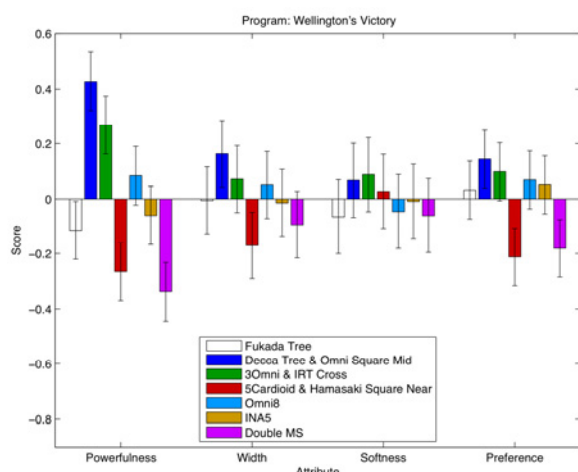


Figure 2: Scheffé's pairwise comparison analysis results from "Wellington's Victory" for Japanese participants. Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

1.6. 視聴者の条件による違い

被験者を職種（放送関連のオーディオ専門職、非放送関連のオーディオ専門職、非オーディオ専門職）に分類し、それぞれの分析結果を調べたが、職種による平均値の違いは明らかではなかった。なお、非オーディオ専門職の信頼区間が広がったが、回答にばらつきが大きかった事が主たる要因だと考えられる。また、

他のグループ（約35名）と比べて被験者数が少なかった（約20名）事も一因として挙げられる。

次に、被験者を試聴ごとに分類し、同様の分析を行なった（Figure 3 および Figure 4）。「ローマの松」において試聴位置が後方の場合（Figure 3 下図）は、全てのマイクアレイについて有意差が確認できなかった。後方に着座した被験者数が少ない事その原因であろう。試聴位置が比較的前方にある場合（Figure 3 上図）では、まず各マイクアレイの「迫力」の感じ方に有意差が見られた。Decca Tree & Omni Square Mid、3 Omni & IRT Cross、5 Cardioids & Hamasaki Square Near はどれも、Fukada Tree、Omni 8、Double MS と有意差があった。「ローマの松」は、前方からは楽器音が、後方からは残響音が主として聞こえてくる楽曲である為に、前方の座席では楽器音の迫力に敏感になっていると考えられる。次に「拡がり」においては、Decca Tree & Omni Square Mid と 5 Cardioids & Hamasaki Square Near が高く、Fukada Tree、Omni 8、INA 5、Double MS の 4 組が低くなっている。「柔らかさ」に関しては 5 Cardioids & Hamasaki Square Near、Omni 8、INA 5 と比べて Double MS が突出している。「好み」の傾向には座席位置による大きな違いはない様である。

また「ウェリントンの勝利」（Figure 4）では、「迫力」「好み」の傾向は前後位置でそれほど大きな変化はない様に見える。しかし「ローマの松」と同様に、試聴位置が後方の場合（Figure 4下図）は被験者数が少ない為に信頼区間が広がっている。

「ローマの松」「ウェリントンの勝利」とともに、着座位置が中央から横に移動した場合、右と左での大きな差は見られなかった。ただし、両方に共通していた事として、「拡がり」の判断における信頼区間がより広がった点が挙げられ、左右への拡がりを判断する事が比較的困難になったと示唆される。

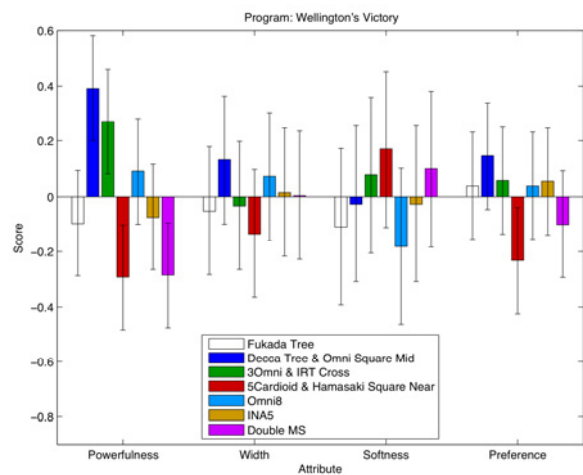
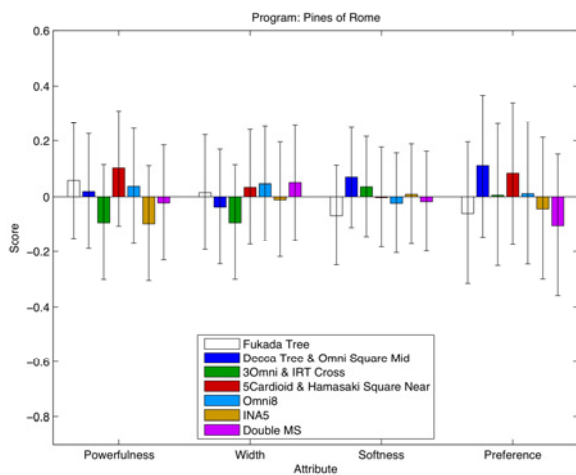
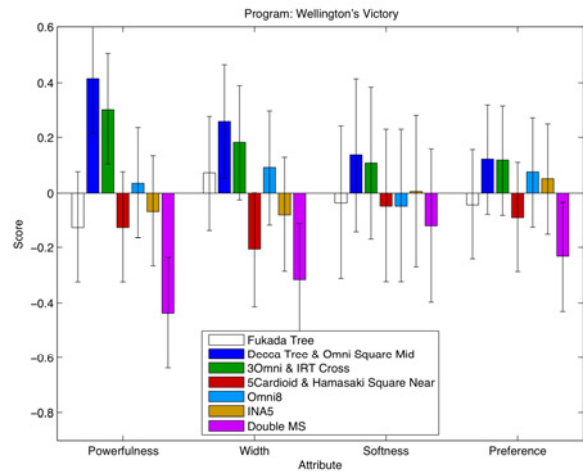
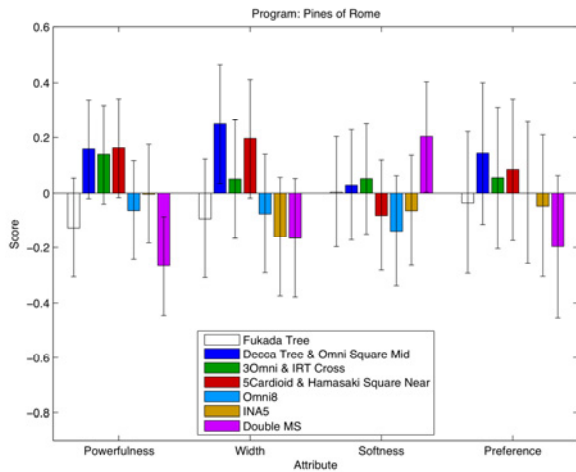


Figure 3: Results from participants seated in the front area (upper panel), and the rear area (lower panel), for “Pines of Rome.” Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

Figure 4: Results from participants seated in the front area (upper panel), and the rear area (lower panel), for “Wellington’s Victory.” Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

2. マイク・アレイの印象に関する主観評価実験<ウィーン編>

2.1. 手法・被験者

日本で行なわれたものと同様の実験を、2007年5月にウィーンにて開催された122nd AES Convention会場にて行ない、様々な国籍の65人の被験者を得た。ただし、再生環境や被験者の制約により、日本国内の3カ所とウィーンでは、用いた言語やアンケート内容が異なっている為、直接の比較は難しい。

試聴実験に使った楽曲は日本で行なった実験と同じく「ローマの松」および「ウェリントンの勝利」であった。質問項目は「powerfulness（迫力）」「width（拡がり）」「softness（柔らかさ）」の3項目とし、「好み」については設問から外した。また、提示したマイクアレイは以下の4種類だった。

- Fukada Tree
- Decca Tree & Omni Square Mid
- INA 5
- Double MS

コンベンションで使用された共通語が英語であった為、英語を母国語としない被験者に対しても、英語で教示が与えられた。

2.2. 分析結果

実験および分析は、日本で行なわれた実験と同様に「シェッフェの対比較法」にて行なわれた。分散分析の結果、2曲ともすべての形容詞において、1%の危険率でマイクアレイの違いに有意差が現れた。結果をFigure 5（「ローマの松」）およびFigure 6（「ウェリントンの戦い」）に示す。前節と同様、棒グラフはそれぞれ評価点の平均値と95%信頼区間を表している。

「ローマの松」では、「迫力」に関してDouble MSが、その他の3つのマイクアレイ（Fukada Tree、Decca Tree & Omni Square Mid、INA 5）との違いが有意だった。「拡がり」においてはFukada Treeが他の3つのマイクアレイと比べ高かった。「柔らかさ」においては、Fukada TreeとDouble MSが両端に評価され、残りの2つは同程度に評価された。

「ウェリントンの戦い」では、日本の結果と同様にDecca Tree & Omni Square Midは「迫力」

があり、Double MSは「迫力」がないとされた。「拡がり」については、個々のマイクアレイを比べた時には有意差の出た対もあるものの、全体的にそれほど差は大きくなかった。Decca Tree & Omni Square Midは「柔らかさ」に関しては他の3つのマイクアレイと比べ、有意に低い評価であった。

また、ウィーンでの実験結果においても職業や着座位置ごとの分析を行なってはみたものの、各グループの被験者が少なかった為、はっきりとした結果は得られなかった。これはCumming & Finchによる第4法則の但し書きである「各群の標本数が10以上」という点によるものであろう。

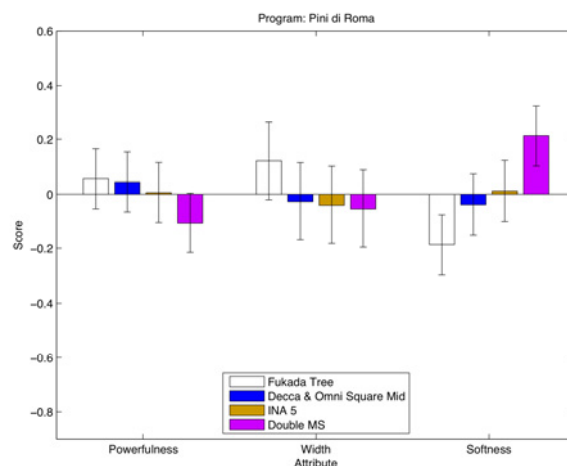


Figure 5: Scheffé's pairwise comparison analysis results from "Pines of Rome" for participants in Vienna. Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

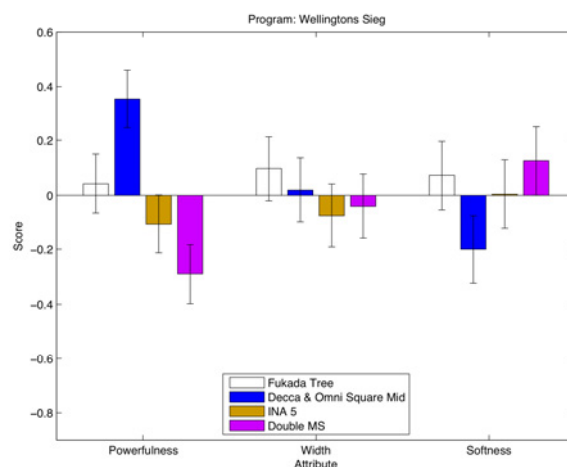


Figure 6: Scheffé's pairwise comparison analysis results from "Wellington's Victory" for participants in Vienna. Colored bars show relative mean ratings and whiskers show 95% confidence intervals around the mean.

3. 各評価属性と好みの関係

1.4節において「迫力」「拡がり」「柔らかさ」の3つは単独の印象属性であり、「好み」は全体的な印象であると述べた。文献[2]で説明されるフィルター・モデルでは、全体的な印象は単独の印象属性によって予測する事ができる、としている。この予測式は各単独印象属性に対する被験者の応答を独立変数に、全体印象属性に対する被験者の応答を従属変数とする重回帰分析を通して作る事ができる。

重回帰分析を日本で行なった実験結果に適用したところ、「ローマの松」では、

$$y_{\text{pref}} = .56x_{\text{power}} + .37x_{\text{width}} - .12x_{\text{softness}}$$

となり ($R^2=.74$)、「ウェリントンの戦い」の結果からは

$$y_{\text{pref}} = .37x_{\text{power}} + .46x_{\text{width}} - .69x_{\text{softness}}$$

が得られた ($R^2=.89$)。ここで x は独立変数(添え字はそれぞれの属性の応答)、 y_{pref} は従属変数とする。

「好み」の予測式中の各属性の重みは楽曲によって異なるものの、「迫力」と「拡がり」が高く、「柔らかさ」の低い録音物に対して、より高い「好み」が得られるものと考えられる。

ただし、これは今回の限られた楽曲とマイクアレイの組み合わせから得られた予測式である為、これ以外の組み合わせの場合にも正しく試聴者の「好み」を予測できる、とは保証できない。また、モデルの「当てはまりの良さ」を示す R^2 値に高い数値が得られてはいるものの、これは従属変数の標本数に対して独立変数の個数が多い為でもあるので、注意が必要である。

なお、ウィーンでの実験では「好み」等の全体的な印象を聞いていない為、同様の分析を行なう事はできなかった。

4. サラウンドとステレオの比較

日本およびウィーンにて、評価実験の一部として同じ楽曲をサラウンド再生とステレオ再生をしたときの印象の違いを尋ねた。サラウンド再生とステレオ再生では、それぞれに調整を行なったものを用いており、サラウンドからのダウンミックスではない。日本とウィーンでは実

験に使える時間が異なっており、形容詞数も異なっている。日本では「迫力」「奥行き」「左右への拡がり」「上下への拡がり」「包まれ感の自然さ」「余韻の自然さ」「好み」の7項目、ウィーンでは「powerfulness」「depth」「width」「envelopment」「natural reverberation」「preference」の6項目について、サラウンド音源とステレオ音源の一对比較を行なった。

結果をFigure 7に示す。×印は平均値を表しており、縦線は95%信頼区間を表している。破線はステレオとサラウンドの中間点、すなわち両者に明確な違いがないという点を表している。全体的にサラウンドの方が高い評定点を獲得しているが、日本における「迫力」ではサラウンドとステレオの違いは有意ではない。また、「上下への拡がり」に関しても明確に違うとは言いきれない。ウィーンでの結果は日本の結果と比べて信頼区間が広いが、これは被験者数の違いによるものである。Natural reverberation(残響の自然さ)においてサラウンドとステレオの違いがやや弱まるものの、その他の評価項目に関してはサラウンドが統計的に有意に高い評定点を獲得している。

5. まとめ

各サラウンド・マイクアレイによって録音された音源が、試聴者にどのような心理的印象を想起させるかを調べる為、日本国内3カ所とウィーンにて「シェッフエの一对比較法」による主観評価実験を行なった。

分析の結果、マイクアレイによって音楽の試聴者に異なる心的印象を与える事が確認された。「迫力」や「拡がり」のあるマイクアレイとしてはDecca Tree & Omni Square Midが2曲の楽曲に共通して挙げられていた。また、「迫力」がなく「好み」も低いマイクアレイとしてはDouble MSが挙げられていた。その他のマイクアレイの印象評価は楽曲による変化が大きく、楽曲、演奏、演奏される空間等に大きく依存する様に思われた。

試聴者の職種による大きな違いは確認できなかった。着座位置によって有意差のあるマイクアレイもあったが、これも楽曲による差が大きい。

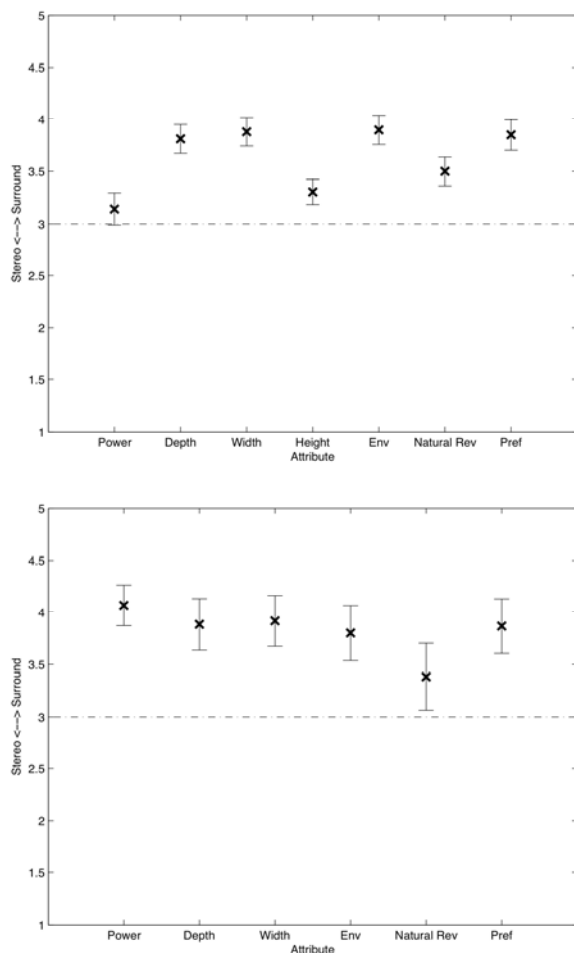


Figure 7: Comparison of subjective impression regarding 5.1 channel surround playback and 2 channel stereo playback. Crosses show mean ratings and whiskers show 95% confidence intervals around the mean. The results were collected in Japan (upper panel) and Vienna (lower panel).

全体的な印象を、単独の感覚属性に関する応答から推定する事を重回帰分析によって試みた。迫力と拡がりがあり、柔らかくない録音物が、より好まれるのではないかと結果が得られた。

また、2チャンネル・ステレオ再生と5.1チャンネル・サラウンド再生の試聴比較実験も行った。ほとんどの評価項目においてサラウンドが統計的に有意に高い評定点を得ており、サラウンド再生が比較的大きな「左右への拡がり」や「包まれ感の自然さ」を感じるようになった。

6. 参考文献

- [1] 佐藤信 『統計的官能検査法』 日科技連出版社 (1985)
- [2] Søren Bech and Nick Zacharov. *Perceptual Audio Evaluation — Theory, Method and Application*. John Wiley & Sons. 2006.
- [3] Geoff Cumming and Sue Finch. Inference by Eye: Confidence Intervals and How to Read Pictures of Data. In *American Psychologist*, vol.60, no.2, pp.170–180. 2006.
- [4] アメリカ心理学会著 江藤ら訳 『APA論文作成マニュアル』 医学書院 (2004)