



## ブラインドテストレポート

スピーカ再生技術研究会

鈴木 茂

2012年10月7日

### 1. はじめに

製品の比較は雑誌記事の定番になっている。一般的には、雑誌の記事を読み、また、自らの体験に基き、製品には差があるということが半ば常識となっている。

しかし、スピーカーシステムを除いては、製品の差は殆どないという意見もある。自分は、従来、スピーカーシステム以外も、製品には差があると思っていた。しかし、差がないという意見を目にし、実際に試してみたことがある。もちろん、ブラインドテストではなく、自分で切り替える方法だったので答を知っていた。しかしながら、実際にテストしてみると、構成も年代も価格も違うアンプの差を感じることができなかつた。

このような勝手なレポートをブログに書いたところ、実際にやってみようということになり、今回のテストを実施することになった。

今回のテストは、初回ということもあり、すべてが試行錯誤で、なおかつ、一部機器のトラブルがあつたため、試験は途中までしかできなかつたが、それなりの結果が得られたと考える。統計処理できるだけのデータ数がなく、数学的には問題のある方法ではあるが、興味深い結果であった。今回の結果だけでは解釈が難しいが、スピーカー再生技術研究会のレポートとして公開することにした。

### 2. 試験の対象と方法

試験の方法は、予め作成しておいたドラフトを元に、セッティングの間、話し合いながら、詳細を決定した。このため、比較するアンプの選択、ソースの選択は、参加者全員で、全てが見える状態で実施した。準備している間、比較するアンプの音色や、聴き分けのポイントなどは、各自、検討する時間は十分にあつたと考えられる。このように、見える状態で準備した理由は、比較する夫々のアンプの状態に問題がないかを確認しておくことと、後のブラインド比較の結果に疑義を残さないためである。主催者が単独で、準備を実施してしまうと、何かタネがあるのではないか、と勘織られても仕方がない。しかし、その準備段階を全部公開しておけば、セッティングでの解答誘導操作はできなくなる。

#### 2.1 比較の対象

比較試験の対象は、アンプとした。その理由は、スピーカー再生に対してもっとも影響が大きく、また、ブラインドで試験するのが比較的容易だからである。CDプレーヤー等は、ブラインドで試験することがほぼ不可能である。何故なら、CDを入れ替える動作が推定されてしまうし、挿入してから音が鳴るまでの時間も違い、メカニカルな動作音も異なる。このためCDプレーヤなどは、部屋を分けない限りテストが非常に難しいと考える。その他、ケーブルなどもブラインドでの比較が困難なものと考る。

ABXテストでは、2種類の比較をするが、被験者にとって苦痛でしかないため、クイズに参加する要素を含め、3種類のアンプを使用した。できれば、デジタルアンプや真空管アンプなど、特徴あるものを含めたかったが、残念ながら、準備した中国製の真空管アンプキットの性能に問題があり、除外した。このアンプは、再生する周波数を変えると出力電圧が大きく違ってしまうため、出力を揃えることが困難と判断したためである。この結果、表1の通りローエンドクラスのプリメインアンプを2



機種、高級クラスのセパレートアンプを1機種とした。

表1 比較に使用したアンプ

記号	製造者	モデル名	定価(消費税抜き)	クラス	出力[W](8Ω)	発売年
A	LUXMAN	M-600A C-600f	500,000円 500,000円	A	30W+30W	2007
B	DENON	PMA-390IV	39,800円	AB	50W+50W	2001
C	YAMAHA	A-S300	39,000円	AB	60W+60W	2010

Aは現行機種の高級セパレート型である。更に上のクラスの製品は、重量が大きすぎ、移動に適さなかったため、比較試験においては、この機種が現実的な最高クラスと云える。B、Cはどちらもアナログプリメインアンプのローエンド機種であるが、発売年が9年違い、アンプの経年劣化も含め差が出ても不思議ではない。

## 2.2 比較の方法

試験は、アンプを適当な順番で鳴らし、前回と差を感じたかどうかを記録するという方法で実施した。この他にオプションとして、モデル名が分かれば記載、点数付けも実施したがこれらは、あくまでも参考のためというメモに留めた。従って、前回との差があるかないかを評価の対象とした。

また、ブラインドテストの実施直前に、それぞれのアンプの名称を開示して、テストと同じ音量で試聴ソースを一度ずつ再生した。

## 2.3 試聴のためのシステム

比較のための絶対条件は、アンプ以外を全て共通の構成とすること、及び、音量を揃えることである。アンプ以外は、下記の機種を使用し、増幅機能のない切替器を使用した。

表2 比較のためのシステム

コンポーネント	機種	備考
プレーヤ		
スピーカーシステム	Feastrex Nf5-Ex を使用した自作システム	励磁型高級フルレンジ

プレーヤは、参加者O氏のお勧めのもので、販売価格は十数万程度、スピーカーシステムは、鈴木所有のもので、ユニットの価格が50万円弱のものである。音出しの時に、これらのシステムでアンプの差を十分出せるであろうことを確認してから実施した。

## 2.4 試聴ソース

試聴ソースは10名の参加者が持ち寄ったものをそれぞれの順番に聴いてゆき、参加者が、『聞き分けが可能』と感じたものを抽出した。今回は、機器トラブルのため、下記1曲の比較試聴を終えたところで中断となった。

試聴ソース イーグルス ホテル カリフォルニアのトラック1の最初から2分20秒

## 2.5 ウォーミングアップ

各アンプは電源を入れたままの状態で保持し、ウォーミングアップは、試聴ソース選択時に切換えて鳴らした。この時間は、全部で4時間程度であったが、全てのアンプで同一条件とはならなかった。アンプAは、A級動作なので、準備の間常にウォーミングアップをしていたことになるが、BとCとは、鳴らしている間だけが負荷であり、かつ、出力は平均で1Wもない状態だったので、ウォーミングアップは不十分だったかもしれない。

## 2.6 試聴室と試験日時

試験は、9月17日(月)準備を含め、9:00から16:00までの間、三鷹市芸術文化センターの第三音楽練習室( $25m^2$ )にて実施した。試聴室は防音構造で、デッドな部屋であり、必ずしもオーディオの試聴に最適という訳ではないが、条件は同一である。

この部屋に、図1のように機器を配置した。座席は、前列をA1 - A5、後列をB1 - B5とし、オペレータを図の右側に配置した。オペレータは、鈴木が担当した。

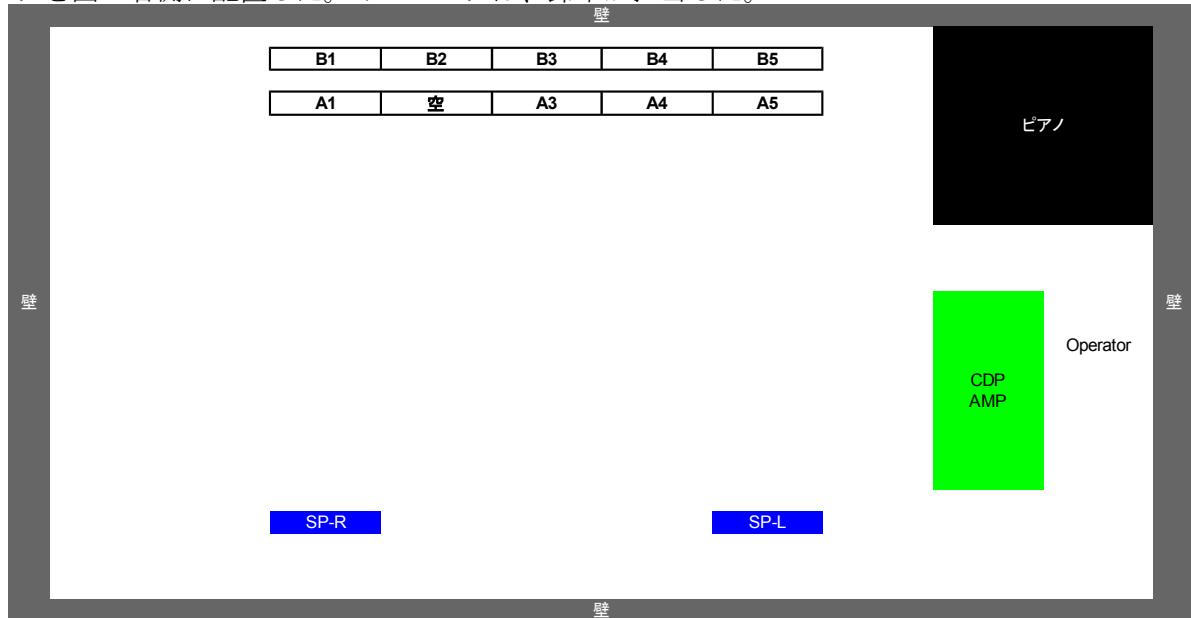


図1 試聴室の配置

## 2.7 音量合わせ

ソース選択の際に、試聴者の意見を聞きながら、騒音計の値を参考にし、およそその音量(75dB前後)を決める。その後、テストCDの正弦波を再生して、スピーカー出力の端子間電圧をデジタルテスター測定し、他のアンプ同じ電圧となるようモリウムを調整した。実際には、表3に示す通り、全く同じ値に調整することは出来なかった。

表3 テスト信号再生時の各アンプの出力電圧

アンプ	出力端子間電圧[V]	出力計算値[W]	最小出力値からの誤差[%]	備考
A	3.48	0.757	1.2%	インピーダンス 16Ω
B	3.52	0.774	3.5%	インピーダンス 16Ω
C	3.46	0.748	基準	インピーダンス 16Ω

## 2.8 試聴の順番と回数

試聴は、各アンプにつき3回、事前に決めておいた順に再生した。ランダムではなく、同じアンプが続く場合も含め、また、各アンプとも同じ回数登場するよう、鈴木が予め6通りの順番を決めておいたものから、その場で適当に選択した。この方針は、被験者に事前に説明したが、当然のことながら、順番は、事前には開示していない。3台のアンプを3回聴くため、ソースひとつにつき、全部で9回の試聴となる。

## 3. 試験の結果

試験の結果を表4に示す。既に記した通り、この試験は、モデル名を当てる試験ではなく、前回と今回との差を感じたか否かを回答するものである。

表4 試験結果

試行(切替)	0	1	2	3	4	5	6	7	8	正答率
↓座席/アンプ→	C	B	B	A	C	C	A	B	A	
A1	-	1	0	0	1	0	1	1	0	50.0%
A3	-	1	0	1	1	1	1	1	1	87.5%
A4	-	0	1	0	1	1	0	0	0	37.5%
A5	-	1	0	0	1	0	1	1	1	62.5%
B1	-	1	1	1	1	0	1	0	1	75.0%
B2	-	1	1	1	1	0	1	1	1	87.5%
B3	-	1	1	1	1	0	0	1	1	75.0%
B4	-	1	0	0	1	0	1	0	1	50.0%
B5	-	1	1	1	0	0	1	1	1	75.0%
正答率	-	88.9%	55.6%	55.6%	88.9%	22.2%	77.8%	66.7%	77.8%	66.7%
正答率累積平均	-	88.9%	72.2%	66.7%	72.2%	62.2%	64.8%	65.1%	66.7%	
出力変化	-	0.026	0.000	-0.018	-0.009	0.000	0.009	0.018	-0.018	

表4において、アンプのA,B,Cは、表3に対応し、左から順に鳴らした。正答の場合を1、不正答の場合を0として表記し、個人ごとの結果を横軸に、試聴ごとの結果を縦軸に表示した。個人は、図1に示す位置に着座し、試験の間は移動していない。

累積平均正答率は、初回の試行から、その回の試行までの平均を計算したものである。累積平均正答率の上昇や下降などの傾向は、学習や疲労などの評価のために使用する。

一応参考のため、試験結果の統計値を表5に示す。被験者数が少く、有為な結果であるとは云えないでの、あくまでも参考として記した。

表5 試験結果の統計値 (参考)

項目	値
被験者数	9
正答率平均	66.67%
正答率最大	87.50%
正答率最小	37.50%
正答率標準偏差	17.68%

## 4. 考察

### 4.1 被験者ごとの評価

被験者個人ごとに見ると、正答率は、37.5%から87.5%までばらついており、平均は66.7%である。8回の切替えのうち、変化があったのが6回、変化がなかったのが2回であり、無作為に回答した場合の正答率は、単純に50%とはならないはずである。例えば、全部変化があったと回答すれば、正答率6/8=75%、全部変化がなかったと回答すれば、正答率は25%となる。このため、完全にランダムな場合と比較して66.7%が期待値より大きかったと単純に断定することは出来ないことに注意すべきである。

### 4.2 試行ごとの評価

4.1により、個人ごとに比較するよりも、試行ごとに比較するほうが、合理的であると考えられる。そこで、正答率88.9%と最も高かった、試行の1回目と4回目を見ると、いずれも、変化があった場合である。また、正答率が22.2%と最も低かった5回目は、変化がなかった場合である。5回目については、気圧や電源電圧のゆらぎ等、別な要因が働き、結果として差を感じた可能性もあるが、思い込みの影響も否定出来ない。

上記が何を意味しているかを考察すると、結局殆どの被験者は、差を感じていなかったのではないかという仮説が可能になる。正答率が最も高かった2名の正答率87.5%も、全部『変化があった』と回答した場合より正答が1問多いだけである。これは、ばらつきの範囲と考えることができるのではないだろうか。

また、別に考慮しなければならぬのは、僅かながら出力(音量)に差があったことである。出力に差があったことを考慮し、出力変化の絶対値を横軸、正答率を縦軸として散布図を書いてみた(図2)。

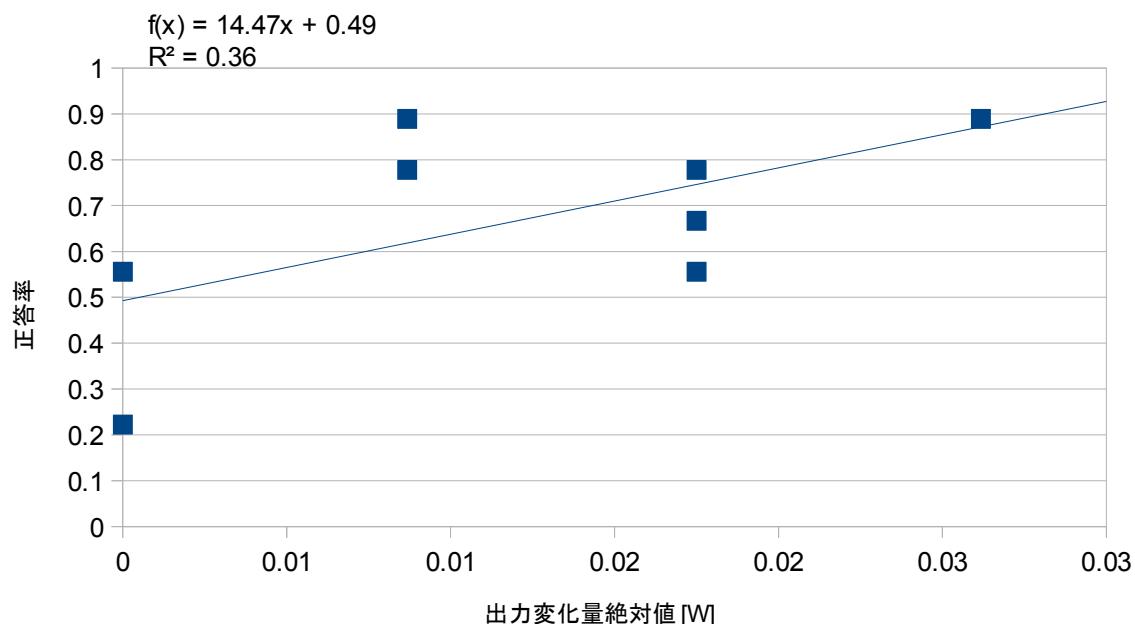


図2 出力の変化に対する正答率

図2を見ると、出力に変化があるほうが正答率が高いという傾向を見ることができる。被験者は音量の差を検知していたという仮説を立てることも可能であろう。しかし、点数が少いため、残念ながら、この散布図だけを以って、論じることは出来ない。

#### 4.3 試験による疲労及び学習の評価

初回から最後までの累積平均正答率を図3に示す。個々の試行での正答率を棒、累積平均正答率を折れ線で表示した。累積平均では、僅かに右下がりの傾向が見られるが、個々の試行に着目すると、8回の試行までの間には、必ずしも正答率が低下している訳ではない。従って、8回という変化の聴き分け回数は、疲労による影響が出る限度内であり、適切な範囲内と考えられる。

同様に、累積平均正答率に上昇傾向は見られないため、この試行の範囲内では、学習の影響も無かったものと推察される。

しかしながら、都度正解を示しながら比較を続けた場合には、学習効果が現れるかもしれない。学習効果があるのは、それなりの差がある場合に限られるので、学習効果を示すことができれば、アンプに聴き分けられる差のあることが証明できる。逆に、学習効果を示せない場合には、聴き分けられる差がないことが証明できる。

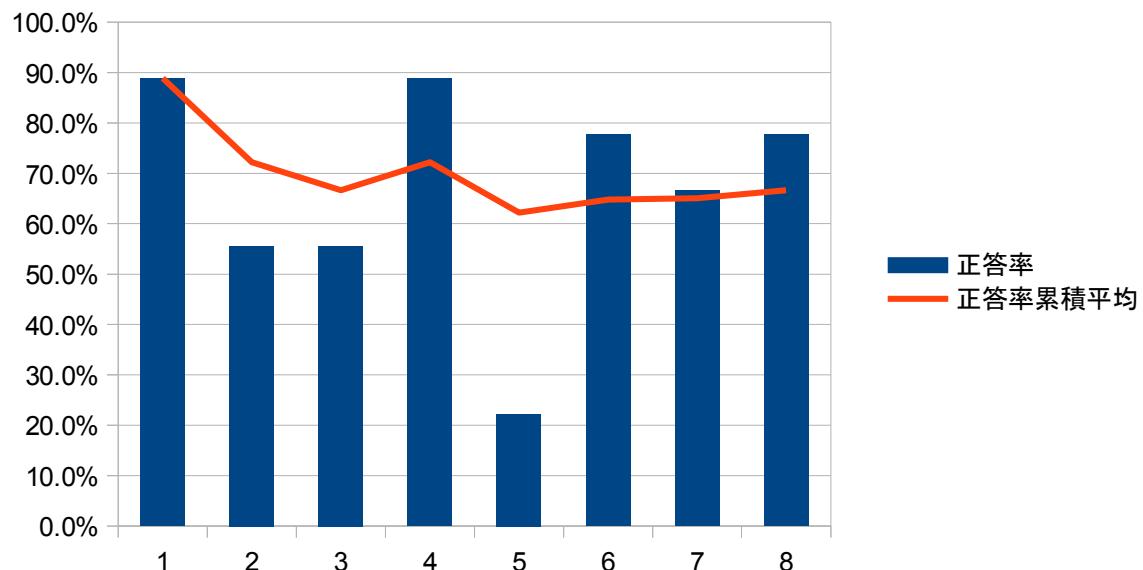


図3 累積平均正答率



## 5. 結論

今回の試験結果を以って、これらのアンプの差を聞き分けることが可能、または、不可能であった、という結論に至ることは出来なかった。厳密に評価するためには、試行回数も不足しており、また、ABX 法のように確立された試験法に準ずることが必要かもしれない。今回は、3つの試験体を用いるという遊びの要素を加えたにも拘らず、参加者には疲労を訴える意見が多かったため、2つの試験体による聴き分けという試行で回数を増やすことは、実現がほぼ不可能なのかもしれない。

今回の試験の結果を考慮すると、雑誌の比較記事のような方法では、差は判別出来ないと考えられる。恐らく、雑誌等の比較テストでは、音量その他の条件が揃えられておらず、また、思い込みの影響が強く出ているのではないだろうか。今回の比較は、ほぼ完全なブラインド状態であり、被験者は事前に答を知ることが出来なかった。そして、大きな価格差のあるシステムを比較したにも拘らず、完全には判別することが出来なかった。このことは、ソースを限定すれば、アンプを変えても聴き分け可能な差がない可能性を示唆している。しかし、音量や試聴ソースを変えれば判別が可能になる可能性も否定できず、追試験は十分に意義があると考える。今後オフ会などで、意見を募り、方法を工夫していきたい。

以上