Hearing is Believing vs. Believing is Hearing: Blind vs. Sighted Listening Tests, and Other
Interesting Things

3894 (H-6)

Floyd E. Toole and Sean E. Olive
Harman International Industries, Inc.
Northridge, CA 91329, USA

Presented at
the 97th Convention
1994 November 10–13
San Francisco

AUDIO
AES

AN AUDIO ENGINEERING SOCIETY PREPRINT

Hearing is Believing vs. Believing is Hearing:
Blind vs. Sighted Listening Tests, and Other Interesting Things

by Floyd E. Toole and Sean E. Olive
Harman International Industries, Inc.
Northridge, CA 91329 U.S.A.

Although it is taken for granted that many factors influence listeners as they form opinions of sound quality, it is interesting to actually put them to the test, and to assess the strength of the factors. These experiments had several objectives. On the subjective side, they were to determine the extent to which listeners' opinions about loudspeaker sound quality are affected by not seeing (blind tests) and seeing (sighted tests) the loudspeakers being evaluated, to examine the performance of listeners with and without experience in critical listening, and to examine the influence of the sex of the listener. On the product side, the objectives were to evaluate the differences among three high-quality expensive loudspeakers and a high-performance, small, inexpensive system which would serve as an "honesty" check for listeners in the sighted tests. The results contain some reassurances and some surprises.

0 INTRODUCTION

Many years of experience with listening tests, conducted under blind and double-blind circumstances have proven their worth and reliability in quantifying subjective responses to several measureable parameters. The results of these psychoacoustic investigations have been repeatable and, with retrospect, the relationships between the subjective and objective domains have been logical [1,2,3]. In properly conducted controlled tests, listeners have been shown to be extremely sensitive to small changes in sound quality.

Yet, there continue to be animated arguments about the validity of blind tests. The majority of these apply to the "great debate" issues, related to the question of whether certain differences are audible or not. This paper does *not* address this debate. All of the differences associated with the subjective evaluations in these experiments were *very* audible, even to untrained listeners. The question here was: given that certain differences among products are clearly audible, to what extent are listeners' opinions altered when they are aware of the products being listened to?

It is probably safe to say that everyone in audio has been involved in "sighted" subjective evaluations at one time or other. It is probably a safe

generalization to say that most people in audio think that they can ignore the effects of prior knowledge when they focus on the sounds of the products under examination. Others would argue that it is difficult to impossible *not* to be biased in some way by expectations. But . . . have you ever put it to the test? Probably not. Neither had we.

Experience is one of those variables among listeners that is very difficult to quantify. For example, musicians are experienced listeners but, is experience in focusing on musical attributes equivalent to that of focusing on timbral and spatial attributes? Some evidence suggests that it is not. Gabrielsson found that musicians who were not also audiophiles, were not especially good judges of sound quality[4]. The famous pianist Glenn Gould came to appreciate the insights of non musicians[5]. Our own tests have confirmed this. So, listeners with different backgrounds could be expected to have differing abilities or preferences in subjective evaluations. This is an enormously broad topic, but we thought that it would be interesting to take a first step towards understanding the importance of this variable.

Everyone knows that females have different preferences to males. Right? If so, where is the proof? Some earlier tests included both male and female participants [6]. In this case, they were all professional sound engineers and producers. In the final analysis, their opinions were indistinguishable from those of their male colleagues. There was one difference, however: a lower percentage of them had hearing loss so that, as a population, they were more reliable listeners. At the consumer level, there remains the question about sexual bias in listener preferences. We now have some data.

The issues addressed in these evaluations are important. They are also not unidimensional Definitive answers must await more data but, in the meantime, it is interesting to have some light shed on the issues.

2 OBJECTIVES

These tests had several objectives. On the subjective side they were:
- to determine the extent to which listeners' opinions about loudspeaker sound quality are affected by not seeing (blind tests) and seeing (sighted tests) the loudspeakers being evaluated,
- to examine the influence of having experience in critical-listening, and
- to examine the influence of the sex of the listener.
    On the product side, the objectives were:
- to evaluate the differences between two different variations of the same basic loudspeaker system. They employed the same drivers in the same enclosure, but the crossover networks were designed by two different engineers having slightly different opinions about the optimum spectral smoothness and balance. These are high-priced, high-end products.
- to compare these performances with that of a current audiophile favorite of a comparable price and size.
- to evaluate a compact, inexpensive, subwoofer/satellite system, and to use it as an "honesty" check for listeners in the sighted tests. Previous tests had shown that, within its power-handling capabilities, it performed in a manner that belied its low price and small size.

3 METHOD

So that the results would carry some weight, forty (40) listeners participated in the blind and sighted tests. They were all employees of Harman International companies. This means that, in the sighted tests, the listeners had one bias in common: brand loyalty.

The effects were tested using male experienced listeners and both male and female inexperienced listeners. In these tests, listeners were considered to be inexperienced if they had no previous experience in controlled listening tests. Other definitions are possible, which might include persons with no critical listening experience whatsoever. The participants were categorized under the following headings.

LISTENERS

Sex	BLIND TESTS		SIGHTED TESTS	
	Experienced	Inexperienced	Experienced	Inexperienced
Male	10	10	8	4
Female	0	5	0	3
Total	10	15	8	7

Unfortunately, it was not possible to balance all levels within each category. In Experiment 1 different listeners participated in the blind and sighted versions. As a further test, Experiment 2 was conducted, in which the same four experienced listeners participated in both versions.

The listening room was typical of a domestic listening situation. In the blind tests, the identities of the loudspeakers were hidden from the listeners with a visually opaque screen made of loudspeaker grille cloth. The grilles were removed from the loudspeakers so that, in effect, the grille cloth hid the entire loudspeaker, not just the drivers. In the sighted tests, the screen was removed and listeners were told the brand name, model number and retail price of each speaker prior to the start of the test.

The tests were conducted over a period of 1.5 weeks using a multiple (4 loudspeakers at a time) presentation method. The monophonic tests were conducted with the loudspeakers adjusted for equal loudness within 0.5 dB using B -weighted pink noise. Playback levels, which were constant throughout the tests, were set for typical "good listening". They were not intended to explore the power-handling capabilities of the systems.

Listeners completed two rounds, in each round giving ratings for four different loudspeakers for each of the 4 different programs. Between rounds, the speaker locations were changed. The order of rounds was randomized among the ten different listening groups. Listeners remained in the same seat locations throughout all tests.

Listeners rated the loudspeakers using a 10 point Preference Scale, where higher ratings indicate greater preferences. This is not the same as the Fidelity Scale used in earlier tests by Toole[1]. This scale is designed to accentuate perceived differences between the loudspeakers[7]. The duration of the entire test was about 20-30 minutes.

Excerpts from the 4 programs were digitally copied from CD on to a hard-disc, edited into 30 second repeating loops, and then transferred to R-DAT for the test playbacks. The programs used were:

ABBREVIATION	PROGRAM
TC	Tracy Chapman/Fast Car
LF	Little Feat/ Let the Good Times Roll
PS	Paul Simon/Graceland
SS	Full Orchestra /Stars & Stripes

4 RESULTS

4.1 EXPERIMENT ONE

The ratings from both experiments were analyzed using a repeated-measures analysis of variance (ANOVA). This is an analysis which evaluates the contribution of the individual variables, and the interactions between them, to the variations in listeners' numerical judgments.

APPENDIX 1 shows the ANOVA table for each source of variance and their interactions. If the H-F value is <= 0.05 then the source had an effect on the loudspeaker ratings with a probability of 95% that the listener responses did not occur due to chance. "Method" is the variable "blind" vs. "sighted".

It is reassuring that the influential factors and interactions are those that one would logically think should modify listener opinions. It is important to note that, in a test of sound quality, not all of the important variables were related to sound. Visual cues had several statistically significant influences.

- Speaker (This is a very strong influence, as it should be.)
- Speaker * Method (Seeing the products had a strong influence on ratings of the loudspeakers)
- Speaker * Seat (Where the listener sat in the room had a strong effect on the ratings. This is an acoustical effect related to the peculiarities of the listening room, and the directional properties of the loudspeakers.)
- Speaker * Program (The choice of music affected the ratings. Listeners' tastes in music occasionally are involved here but, much more important, are the facts that (a) musical selections are not all equally good at revealing problems in loudspeakers, and (b) that different recordings have different spectral balances ("voices) which interact with the different "voicings" of the loudspeakers.)
- Speaker * Program * Experience (The choice of music affected opinions of experienced listeners differently than those of inexperienced listeners. Some listeners, mainly inexperienced ones, tend to stay with a first impression, and not change it through quite wide variations in program material.)
- Speaker * Program * Experience * Method (As above, but seeing the loudspeakers made a difference)
- Speaker * Round (The locations of the loudspeakers in the room made a difference to the sound. This is a well known phenomenon. It has been scientifically demonstrated that, in tests involving good and closely rated loudspeakers, the locations of the loudspeakers in the room can be the dominant factor in determining the ratings [8,9].)
- Speaker * Round * Method (As above, but seeing the loudspeakers made the differences matter less.)

The specifics of these factors are discussed in the following sections.

4

4.1.1 Loudspeaker Preferences

Combining the results of all listeners in all rounds, indicates that models "D" and "G" were slightly to moderately preferred over "S" and "T", Figure 1. Be careful not to be misled by the scaling of this graph. The total vertical height is less than a third of the 10-point preference scale. In the normal context of listening tests, this is a small range of ratings, indicating a fairly close contest. There were no truly <u>bad</u> loudspeakers here.

Interaction Bar Chart
Effect: Speaker
Dependent: PreferenceRatings
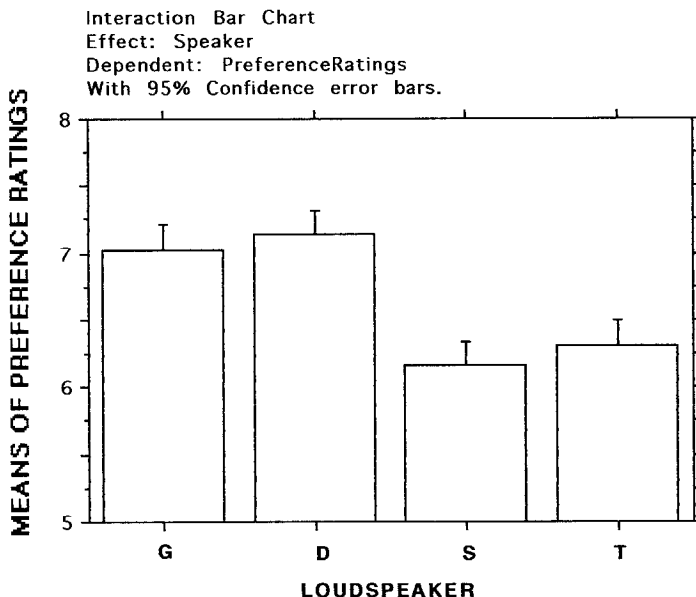With 95% Confidence error bars.

Figure 1. Combined results of all tests, blind and sighted, showing the cell means of preference ratings for all listeners. Note that the vertical scale is only a portion of the 0 - 10 preference scale.

The results show that the loudspeakers fell into two closely rated pairs, "G" and "D", and "S" and "T". The listeners clearly preferred the first pair to the second pair but, within the pairs, the error bars indicate that they had no preference that was statistically important. Remember that these results contain a mixture of all tests, sighted and blind. There is more to this story.

4.1.2 Effect of Test Method (Blind Versus Sighted) and Loudspeaker Position

There is now abundant evidence that the listening room and the positions of loudspeakers within the room are significant influences on

listener opinions of loudspeakers. In the blind tests, in which listeners had only the sound to rely on, preference ratings were strongly dependent on the locations of the loudspeakers, Figure 2.

Interaction Bar Chart
Effect: Speaker * Round * Method
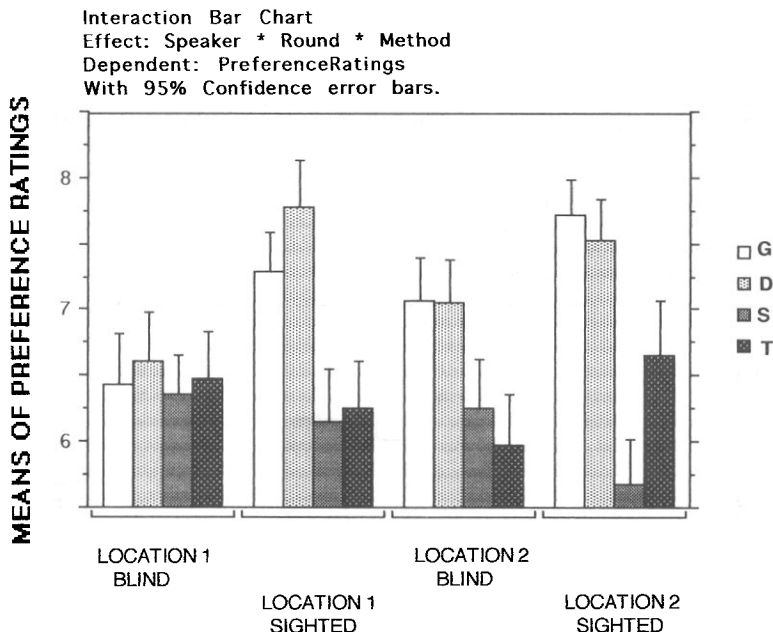Dependent: PreferenceRatings
With 95% Confidence error bars.

Figure 2 A comparison of results for the blind and sighted tests, for each of the loudspeaker locations.

In fact, in the blind tests, where opinions were based solely on sound, the location of the loudspeakers in the room had more of an effect in the preferences of these loudspeakers, than the loudspeakers themselves. In location 1, the listeners exhibited no clear preference for any of the four loudspeakers. In location 2, however, there were strong preferences. Obviously, important difference in sound quality were introduced by the position factor.

However, in the sighted tests, the ratings were very strongly differentiated, and they did not change appreciably between the two sets of room locations. In other words, in this test, when listeners knew what they were listening to, the opinions were dictated more by the product identity than by the sound.

If we isolate the visual and political factors we have the following possible scenario. It is easy to believe that loudspeakers "G" and "D" would be viewed favorably because they were the most expensive, the largest, quite attractive, and they were products of the company that employed the listeners. Loudspeaker "T" was slightly smaller, slightly less expensive, a prestige

product, but made by a competitor. Loudspeaker "S" was absolutely tiny, relatively inexpensive, and plastic. It was a product of the host company, but could anything that small and cheap be any good? Many listeners in the sighted tests admitted afterwards that before the music even started they believed that loudspeaker "S" would sound inferior, although they admitted its strong performance surprised them.

Combining the data from tests in both of the loudspeaker locations, Figure 3 isolates the effect of seeing the products that are being evaluated. The separation of the products into two groups is clear in both results, but the preference of loudspeaker "T" over loudspeaker "S" is evident only in the "sighted" results.

Interaction Bar Chart
Effect: Speaker * Method
Dependent: PreferenceRatings
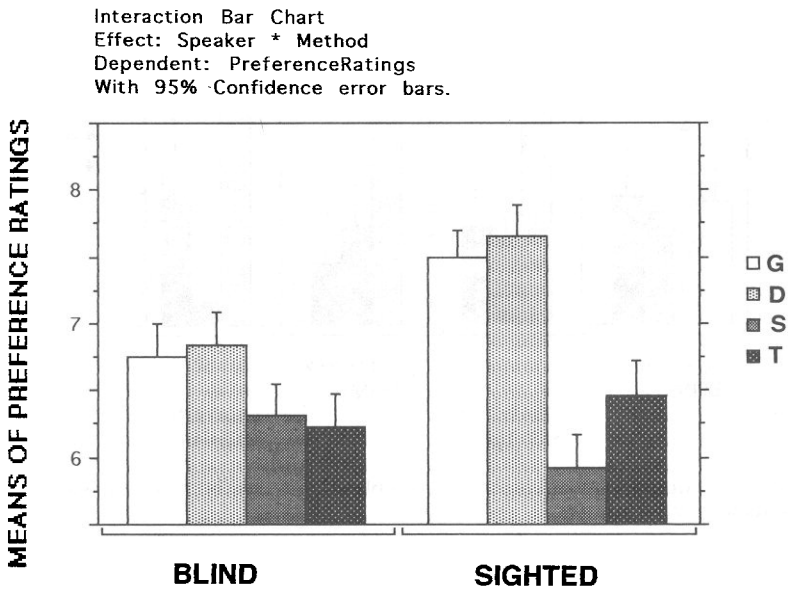With 95% Confidence error bars.

Figure 3 A comparison of listener preferences in blind and sighted tests, combining the results of tests performed in two loudspeaker locations.

Some of us would like to think that we can ignore visual factors when arriving at opinions of sound quality . . . but can we?

An overall effect that is interesting, but not of any real consequence here, is that listeners in the "sighted" tests (both experienced and inexperienced) used higher ratings compared to listeners in the blind tests, Figure 4. We can speculate that listeners may use higher ratings when they can see the products because they are more confident in their opinions, or they are less concerned about revealing inconsistencies in their judgments, or both.
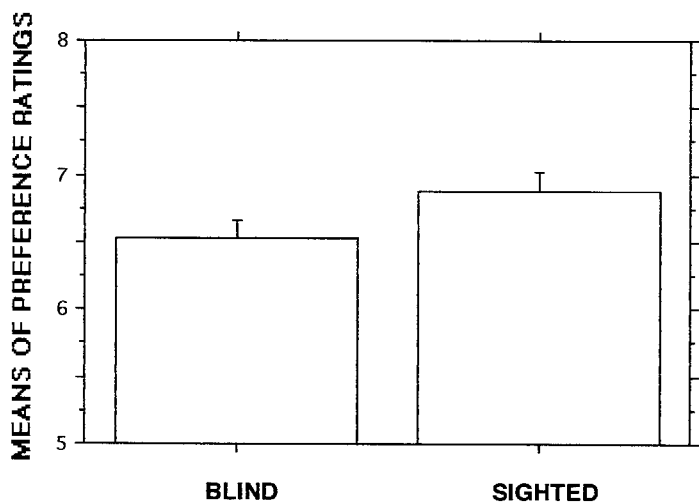
7

Figure 4 Averaged across all experiments, a comparison of the ratings used by listeners in blind and sighted tests.
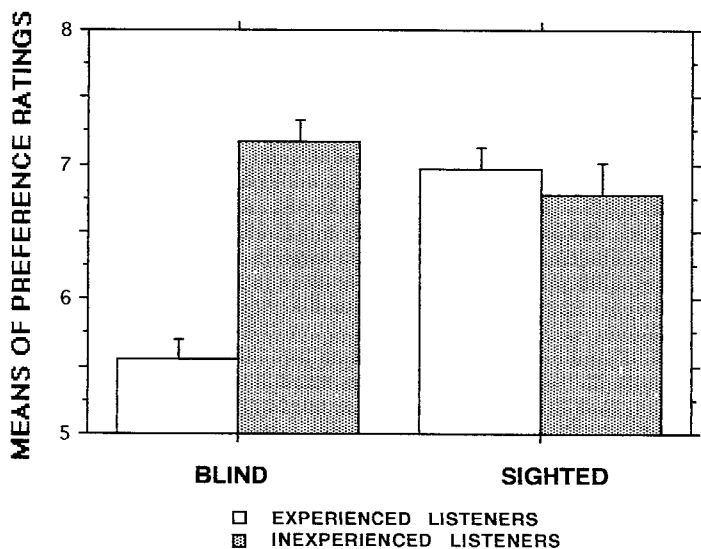

Figure 5 A comparison of the ratings used by experienced and inexperienced listeners in blind and sighted tests

4.1.3 The Effect of Listener Experience

Separating the listeners by experience, Figure 5, it becomes clear that it is the experienced listeners in the blind tests that caused the strongest differentiation. Experienced listeners used lower ratings than inexperienced listeners in the blind tests but in the sighted tests the difference disappeared.

While it is interesting to speculate about why this occurs, the absolute ratings used by listeners are of no consequence to the important result, which is the relative ratings of the products under evaluation. In this it can be clearly stated that the inexperienced male listeners had the same loudspeaker preferences as the experienced male listeners, Figure 6.
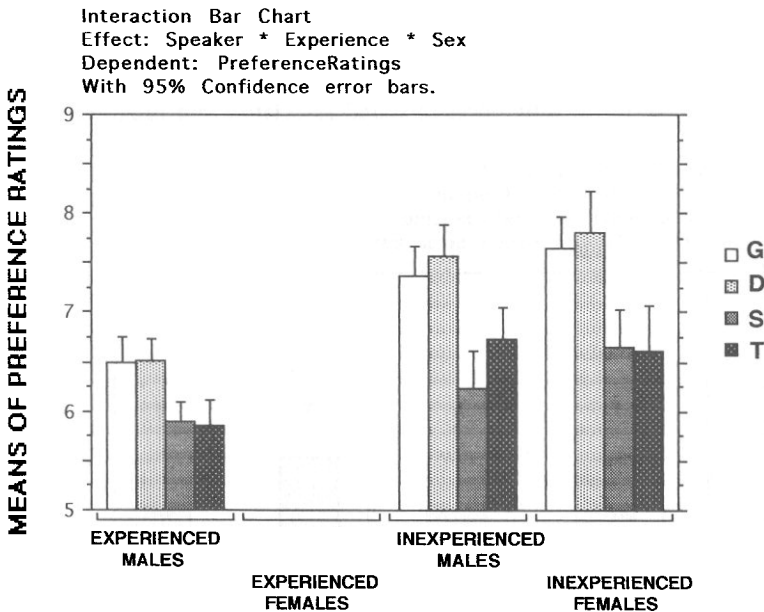
Interaction Bar Chart
Effect: Speaker * Experience * Sex
Dependent: PreferenceRatings
With 95% Confidence error bars.

Figure 6   Loudspeaker preferences classified by sex and listening experience.

4.1.4 The Effect of Sex

There is a popular belief that females have different preferences in loudspeakers than males. The folklore is rich with tales of irreconcilable differences between the sexes. Some evidence suggests that other factors may have been involved, such as price, size, style, loudness, purchasing priorities, etc. Here, though, we ignore everything but the sound, and ask the question. Figure 6 shows that the opinions of inexperienced male and inexperienced female listeners (see the two right-hand histograms) are remarkably similar. *Viva la similarité!*

9

4.1.5 The Effect of Listener Location

The interactions among loudspeaker, loudspeaker position and listener position are strong and complex. When comparing loudspeakers that are comparably good in terms of timbral accuracy, as these loudspeakers are, one must be aware of these effects if the results are to be trusted. In these experiments the physical differences in loudspeaker locations between rounds were much smaller than differences between listener locations. Listener location is therefore the stronger variable.

Figure 7 shows that front row listeners in seats 1 and 2 had similar loudspeaker preferences, although seat 2 listeners slightly preferred model "D". Back row listeners (seats 3, 4 and 5) showed differing preferences.
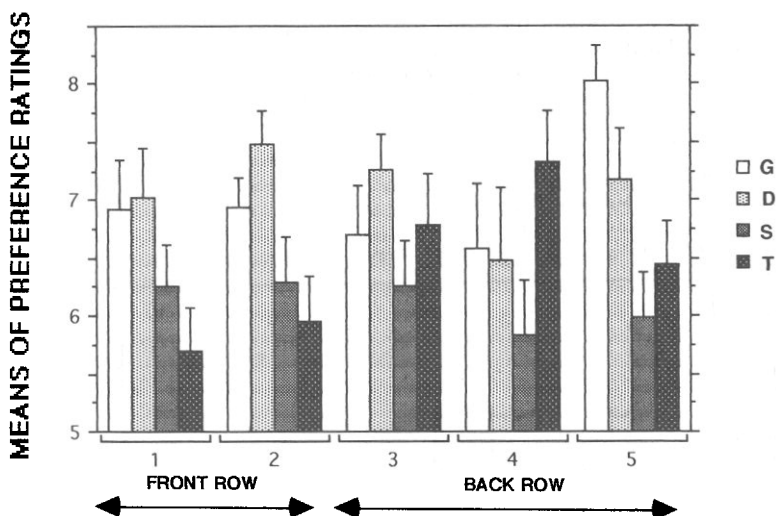

Figure 8 Loudspeaker preference ratings as a function of listener location.

It is interesting to look for trends in these data. Loudspeaker "G" stays within about 0.3 of a preference rating, except for seat 5. Loudspeaker "D" stays within about 0.5 of a preference rating, except for seat 4. Loudspeaker "S" stays within 0.8 of a preference rating for all of the seats, or 0.5 with the exception of seat 4. Loudspeaker "T" in contrast, spans 2.3 points on the scale.

To put this in perspective, listeners are instructed to separate ratings by at least 0.5 if they have a "slight preference", by about 1.5 if they have a "moderate preference" and by more then 2.0 if they have a "strong preference". The wide variations in the rating of loudspeaker "T" as a function of listener position is an indicator of at least two important things: (1) listener location is not to be ignored and, (2) loudspeaker "T" does not have a reliable relationship with the listening situation. Two possibilities come to mind: (1) there are large variations in low-frequency coupling as a function of listener location, (2) loudspeaker "T" has inconsistencies in directivity that are revealed differently in different loudspeaker/listener orientations. The low-frequency variations would apply to all of the loudspeakers, suggesting

10

that the second possibility may have some validity. This will be illustrated in section 6, when measurements are discussed.

These data underline the very great importance of having a good listening room and loudspeaker/listener arrangement, and knowing the biases that can be introduced by loudspeaker or listener position within the room. Thorough randomizing of these factors can help, but it prolongs the test enormously. It is better to avoid strong positional biases by working in an acoustical environment that is a known factor, something that is rarely possible, as we all know. It is also essential to track listener responses as a function of seat, since something of importance may be revealed.

5 RESULTS - EXPERIMENT TWO

In this experiment the same four experienced listeners, seated in the same seats, did the experiment in blind and sighted methods, in that order.

The ANOVA table for this experiment (see Appendix 2) shows significant interactions between *Method * Speaker* and *Speaker * Round* , both with H-F values near 0.03.

Interaction Bar Chart
Effect: Method * Speaker
Dependent: PreferenceRating
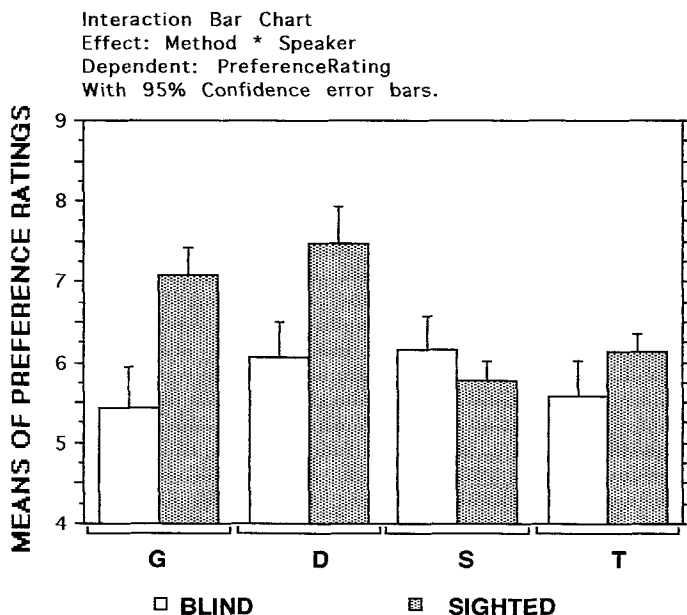With 95% Confidence error bars.

□ BLIND   ▦ SIGHTED

Figure 9   Blind vs. sighted ratings for the four loudspeakers, using the same four experienced listeners in both tests.

Figure 9 shows that, in the blind tests, it was a very close contest, with no strong preferences being evident. The group means suggest a slight preference for "D" and "S" over "T" and "G".

When the screen was removed and the test repeated, the results were very different. Even with the same experienced listeners, in the same seats, performing both tests, seeing the loudspeakers added the same sequence of strong biases that was seen in the results of Experiment One, which used different listeners in blind and sighted tests (see Figure 1). The biases: the ratings of loudspeakers "G" and "D" are increased by amounts suggestive of moderate preference, loudspeaker "S" drops by an amount suggesting slight (decline in) preference, and "T" increases by an amount suggesting a slight preference.

5.1 The Effect of Program

Figure 10 shows that, in the blind tests, the ratings varied with program, something that is to be expected, and which is commonly seen. In the sighted tests, this effect almost completely disappeared. Obviously, listeners' opinions were more attached to the products that they could see, than they were to the differences in sound associated with program.
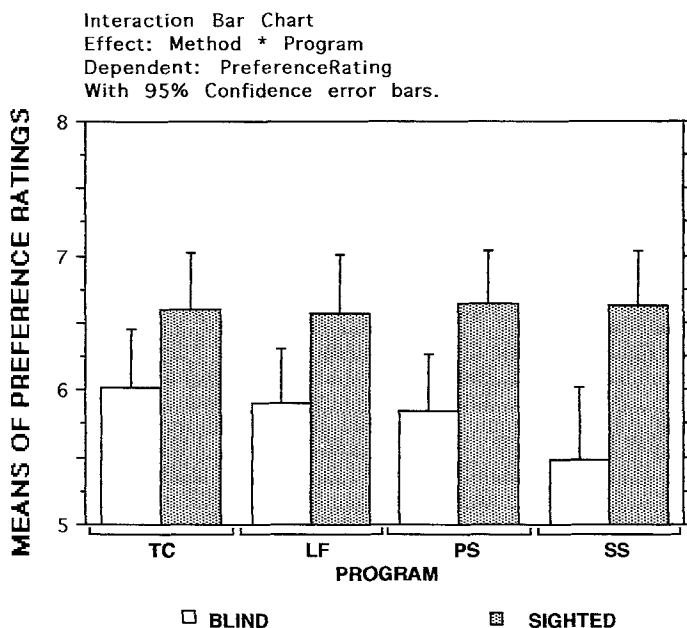

Figure 10 The effect of program on preference ratings for both blind and sighted tests.

12

5.2 The Effect of Loudspeaker Position

Figure 11 shows that the locations of the loudspeakers had strong effects on the ratings in the blind tests (open bars), while in the sighted tests (dark bars), the speaker placements had little effect on the ratings. Just as in Experiment 1, the fact that the listeners knew what was being listened to caused them to be much less responsive to real differences in sound quality.

Interaction Bar Chart
Effect: Method * Speaker * Round
Dependent: PreferenceRating
With 95% Confidence error bars.
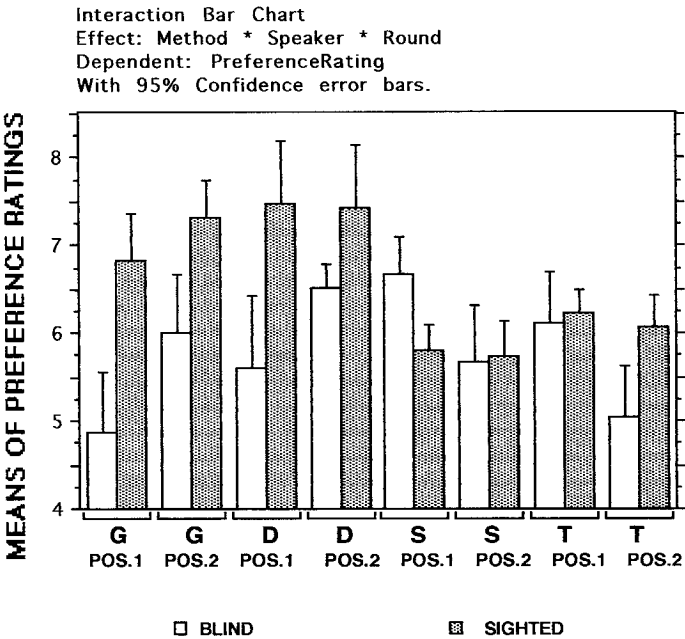
□ BLIND   ▨ SIGHTED

Figure 11. Preference ratings as a function of loudspeaker position, for both blind and sighted tests.

6 MEASUREMENTS

This project did not set out to be a test of opinion vs. measurements, but the temptation is strong to have a look at some limited objective data.

Anechoic measurements at 2 meters were performed on the four loudspeakers on axis, and 30 and 60 degrees horizontally off axis. These are shown in Appendices 3(a - d). Note that the measurements below 100 Hz are not accurate and should be ignored. Measurements of this kind constitute the absolute minimum useful data for loudspeaker assessment. Nevertheless, they can provide important insights into why the products might have performed the way they did.

Loudspeakers "G" and "D" (appendices 3(a) and (b) respectively) reveal their common origin, but there are significant differences. Overall, they are

well behaved, exhibiting relatively smooth, relatively flat axial curves, wide dispersion, and good directional uniformity. Viewed overall, these are creditable performances. Both systems are relatively free from resonant colorations, but loudspeaker "D" is the less refined of the pair. There are also distinctive spectral balances, with loudspeaker "G" being the brighter, more treble-biased, and loudspeaker "D" exhibiting a more temperate top end. Still these are not exaggerated cases, and this is supported by the high ratings that both received in the listening tests. The lack of a clear preference for either of these loudspeakers is evidence of listeners being divided in their acceptance of their different attributes.

Loudspeaker "S" (Appendix 3(c)) is also well behaved on and off axis. Directivity is relatively constant, failing noticeably only when the tweeter turns on just above 4 kHz. Overall, the level above 400 Hz is slightly elevated which might make instruments whose fundamentals fall below this frequency sound thin or forward. Of course, this will be dependent on how the separate subwoofer sums with the satellite in the room. This is a more than respectable performance, especially in this class of product. In the end, though, it is a small loudspeaker, with inexpensive components. As a result low bass output is limited, and at very high sound levels there are limitations. Its creditable performance in the blind listening tests is evidence of a design that does many things well, most of the time.

Loudspeaker "T" (Appendix 3(d) presents a more complex situation. On axis, it has a slightly bright balance and a small interference dip at 1 kHz. The dip is likely to be sensitive to vertical angle, but the measurement was made on the intended listening axis. By itself, it is not a large problem, but coupled with the obvious directional inconsistencies revealed in the off-axis curves, it becomes an issue. The inconsistent directivity is right in the middle of the very important voice-frequency range, and causes audible coloration in this range. The variable directivity also means that the sound quality will change with both loudspeaker and listener position. Still, it has other virtues, such as extended low bass performance, and the ability to play moderately loud without distress.

In Section 4.1.5 it was speculated that loudspeaker "T" might have inconsistencies in directivity that could account for the strong position-dependent variations in preference rating. That speculation appears to have been correct.

To sum up, even these simple measurements are sufficient to reassure us that the results of these subjective tests have a basis in physical reality. A more exhaustive inquiry would be interesting.

7 CONCLUSIONS

"It is important to note that, in a test of sound quality, not all of the important variables were related to sound. Visual cues had several statistically significant influences." (Section 4.1)

"In the normal context of listening tests, this is a small range of ratings, indicating a fairly close contest. There were no truly bad loudspeakers here." (Section 4.1.1) Some of the following conclusions may have been different if differences between the products had been greater.

14

". . . when listeners knew what they were listening to, the opinions were dictated more by the product identity than by the sound". (Section 4.1.1) The strength of the biases would be different in a test with products having greater performance differences. Nevertheless, the visual biases would still be present as unwanted influences.

"It can be clearly stated that the inexperienced male listeners had the same loudspeaker preferences as the experienced male listeners." (Section 4.1.3) In a race this close, it is clear that we had some very canny inexperienced listeners.

"the opinions of inexperienced male and inexperienced female listeners are remarkably similar. Viva la similarité! " (Section 4.1.3)

"These data underline the very great importance of having a good listening room and loudspeaker/listener arrangement, and knowing the biases that can be introduced by loudspeaker or listener position within the room. Thorough randomizing of these factors can help, but it prolongs the test enormously. It is better to avoid strong positional biases by working in an acoustical environment that is a known factor, something that is rarely possible, as we all know. It is also essential to track listener responses as a function of seat, since something of importance may be revealed." (Section 4.1.5)

"Even with the same experienced listeners performing both tests, seeing the loudspeakers added the same sequence of strong biases that was seen in the results of Experiment One, with different listeners in the blind and sighted tests." (Section 5) No one, it seems, is totally immune to the effect of visual biases.

"Obviously, listeners' opinions were more attached to the products that they could see, than they were to the differences in sound associated with program." (Section 5.1)

". . . the fact that the listeners knew what was being listened to caused them to be much less responsive to real differences in sound quality. [caused by changes in loudspeaker position in the room]" (page 11) If your opinion of Brand X were already on record, would you change it if you thought the same loudspeaker sounded different in another test? It could also be a special case of selective perception.

In summary, in listening tests where the audible differences between products were not difficult to hear, knowledge of product identity while listening had profound effects on listener opinions. In some instances, altered listener preferences resulted from listeners being less responsive to audible differences in the sighted tests than they were in the blind tests. For example: (a) they were less responsive to differences caused by loudspeaker location in the room, and (b) they were less responsive to differences associated with program material.

Overall, though, it was clear that the psychological factor of simply revealing the identities of the products altered the preference ratings by amounts that were comparable with any physical factor examined in these tests, including the differences between the products themselves. That an effect of this kind should be observed is not remarkable, nor is it unexpected.

What <u>is</u> surprising is that the effect is so strong, and that it applies about equally to experienced and inexperienced listeners.

Since all of this is independent of the sounds arriving at the listeners' ears, we are led to conclude that, under some circumstances, believing *is* hearing!

The bottom line: if you want to know how a loudspeaker truly <u>sounds</u>, you would be well advised do the listening tests "blind".

8 REFERENCES

1.   Toole, F.E. "Loudspeaker Measurements and Their Relationship to Listener Preferences", J. Audio Eng, Soc., vol. 34, pt.1 pp.227-235 (1986 April), pt. 2, pp. 323-348 (1986 May).

2.   Toole, F.E. and S.E. Olive, "The Modification of Timbre by Resonances: Perception and Measurement", J. Audio Eng, Soc., vol. 36, pp. 122-142 (1988 March).

3.   Toole, F.E., "Loudspeakers and Rooms for Stereophonic Sound Reproduction", Proceedings of the 8th International Conference, Audio Eng, Soc. (1990 May).

4.   Gabrielsson, A. and Sjogren, H., "Perceived Sound Quality of Sound Reproducing Systems", J. Acoust. Soc. Am., vol. 65, 1019 (1979).

5.   Gould, Glenn, "An Experiment in Listening - Who are the Most Perceptive Listeners", High Fidelity Magazine, vol.25, pp. 54-59 (August 1975).

6.   Toole, F.E., "Subjective Measurements of Loudspeaker Sound   Quality and Listener Performance", J. Audio Eng. Soc., vol.33, 2 (1985).

7.   Toole, F.E., "Subjective Evaluation", in "Loudspeaker and Headphone Handbook", Second Edition, edited by John Borwick, Butterworths, London (in press).

8.   P. L. Schuck, S. Olive, J. Ryan, F. E. Toole, S. Sally, M. Bonneville, E. Verreault, K. Momtahan, "Perception of Reproduced Sound in Rooms: Some Results from the Athena Project", pp.49-73, Proceedings of the 12th International Conference, Audio Eng. Soc. (1993 June).

9.   S.E. Olive, P. Schuck, S. Sally, M. Bonneville, "The Effects of Loudspeaker Placement on Listeners' Preference Ratings", 93rd Convention, Audio Eng. Soc., preprint no. 3350 (1992 Oct.).

Type III Sums of Squares

APPENDIX 1

Source	df	Sum of Sq...	Mean Sq...	F-Value	P-Value	G-G	H-F
Experience	1	81.988	81.988	3.777	.0662		
Method	1	74.625	74.625	3.438	.0785		
Seat	4	10.277	2.569	.118	.9744		
Experience * Method	1	153.770	153.770	7.084	.0150		
Experience * Seat	4	108.401	27.100	1.248	.3227		
Method * Seat	4	163.355	40.839	1.881	.1531		
Experience * Method * Seat	4	92.852	23.213	1.069	.3976		
Subject(Group)	20	434.130	21.707				
Speaker	3	206.662	68.887	9.346	.0001	.0001	.0001
Speaker * Experience	3	25.876	8.625	1.170	.3287	.3284	.3287
Speaker * Method	3	87.940	29.313	3.977	.0119	.0124	.0119
Speaker * Seat	12	166.372	13.864	1.881	.0552	.0566	.0552
Speaker * Experience * Method	3	39.672	13.224	1.794	.1580	.1591	.1580
Speaker * Experience * Seat	12	91.860	7.655	1.039	.4270	.4269	.4270
Speaker * Method * Seat	12	70.744	5.895	.800	.6490	.6471	.6490
Speaker * Experience * Method * Seat	12	99.301	8.275	1.123	.3596	.3602	.3596
Speaker * Subject(Group)	60	442.267	7.371				
Program	3	3.365	1.122	2.062	.1148	.1313	.1148
Program * Experience	3	1.200	.400	.736	.5349	.5047	.5349
Program * Method	3	.989	.330	.606	.6137	.5748	.6137
Program * Seat	12	7.704	.642	1.180	.3178	.3290	.3178
Program * Experience * Method	3	.436	.145	.267	.8488	.8000	.8488
Program * Experience * Seat	12	4.030	.336	.617	.8191	.7818	.8191
Program * Method * Seat	12	5.743	.479	.880	.5712	.5529	.5712
Program * Experience * Method * Seat	12	3.708	.309	.568	.8589	.8218	.8589
Program * Subject(Group)	60	32.637	.544				
Round	1	.721	.721	.684	.4180	.4180	.4180
Round * Experience	1	2.618	2.618	2.482	.1308	.1308	.1308
Round * Method	1	.295	.295	.280	.6026	.6026	.6026
Round * Seat	4	.411	.103	.097	.9821	.9821	.9821
Round * Experience * Method	1	.087	.087	.083	.7764	.7764	.7764
Round * Experience * Seat	4	3.581	.895	.849	.5109	.5109	.5109
Round * Method * Seat	4	1.011	.253	.240	.9126	.9126	.9126
Round * Experience * Method * Seat	4	3.837	.959	.910	.4774	.4774	.4774
Round * Subject(Group)	20	21.094	1.055				
Speaker * Program	9	23.334	2.593	3.173	.0014	.0136	.0014
Speaker * Program * Experience	9	15.741	1.749	2.140	.0283	.0742	.0283
Speaker * Program * Method	9	6.245	.694	.849	.5719	.5092	.5719
Speaker * Program * Seat	36	19.478	.541	.662	.9281	.8395	.9281
Speaker * Program * Experience * Method	9	16.100	1.789	2.189	.0247	.0686	.0247
Speaker * Program * Experience * Seat	36	24.804	.689	.843	.7219	.6463	.7219
Speaker * Program * Method * Seat	36	26.950	.749	.916	.6092	.5617	.6092
Speaker * Program * Experience * Method ...	36	26.963	.749	.917	.6086	.5612	.6086
Speaker * Program * Subject(Group)	180	147.085	.817				
Speaker * Round	3	34.095	11.365	3.451	.0220	.0370	.0220
Speaker * Round * Experience	3	17.365	5.788	1.758	.1649	.1819	.1649
Speaker * Round * Method	3	30.266	10.089	3.063	.0348	.0528	.0348
Speaker * Round * Seat	12	40.982	3.415	1.037	.4283	.4262	.4283
Speaker * Round * Experience * Method	3	5.932	1.977	.600	.6173	.5667	.6173
Speaker * Round * Experience * Seat	12	27.389	2.282	.693	.7516	.7068	.7516
Speaker * Round * Method * Seat	12	15.472	1.289	.392	.9615	.9291	.9615

Type III Sums of Squares

APPENDIX 2

Source	df	Sum of Squares	Mean Square	F-Value	P-Value	G-G	H-F
Subject	3	40.199	13.400				
Method	1	41.120	41.120	5.090	.1093	.1093	.1093
Method * Subject	3	24.235	8.078				
Speaker	3	31.226	10.409	1.361	.3156	.3289	.3258
Speaker * Subject	9	68.831	7.648				
Program	3	2.264	.755	.885	.4850	.4364	.4636
Program * Subject	9	7.678	.853				
Round	1	.056	.056	.099	.7739	.7739	.7739
Round * Subject	3	1.713	.571				
Method * Speaker	3	41.131	13.710	4.624	.0320	.0751	.0320
Method * Speaker * Sub...	9	26.685	2.965				
Method * Program	3	3.016	1.005	3.130	.0802	.1191	.0802
Method * Program * Su...	9	2.890	.321				
Speaker * Program	9	5.203	.578	1.566	.1758	.2778	.1758
Speaker * Program * S...	27	9.964	.369				
Method * Round	1	.063	.063	.045	.8461	.8461	.8461
Method * Round * Subject	3	4.193	1.398				
Speaker * Round	3	23.801	7.934	4.482	.0347	.0616	.0347
Speaker * Round * Subj...	9	15.931	1.770				
Program * Round	3	2.573	.858	4.531	.0337	.1050	.0789
Program * Round * Subj...	9	1.704	.189				
Method * Speaker * Pro...	9	.630	.070	.202	.9918	.8493	.9918
Method * Speaker * Pro...	27	9 348	.346				
Method * Speaker * Rou...	3	12.163	4.054	3.970	.0468	.1115	.0698
Method * Speaker * Rou...	9	9.193	1.021				
Method * Program * Ro...	3	.547	.182	1.452	.2916	.2998	.2916
Method * Program * Ro...	9	1.129	.125				
Speaker * Program * R...	9	2.106	.234	.779	.6369	.5188	.6369
Speaker * Program * R...	27	8.110	.300				
Method * Speaker * Pro...	9	2.905	.323	1.685	.1417	.2573	.1439
Method * Speaker * Pro...	27	5.173	.192				

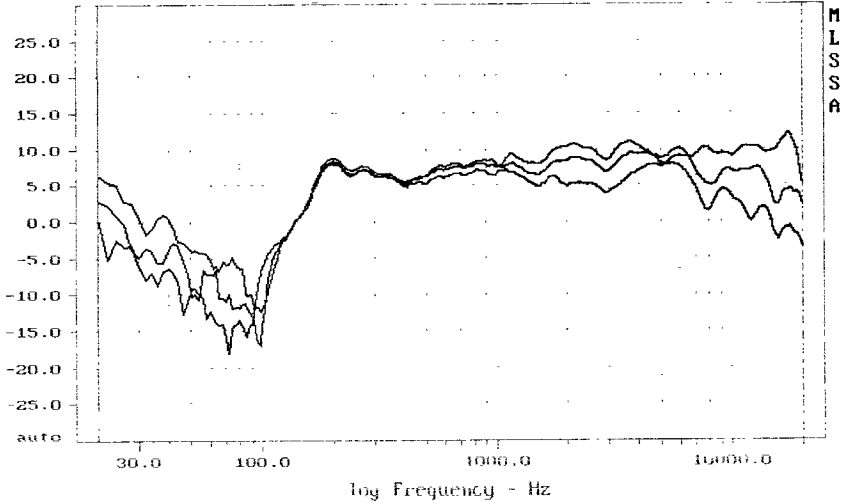Dependent: PreferenceRating

APPENDIX 3
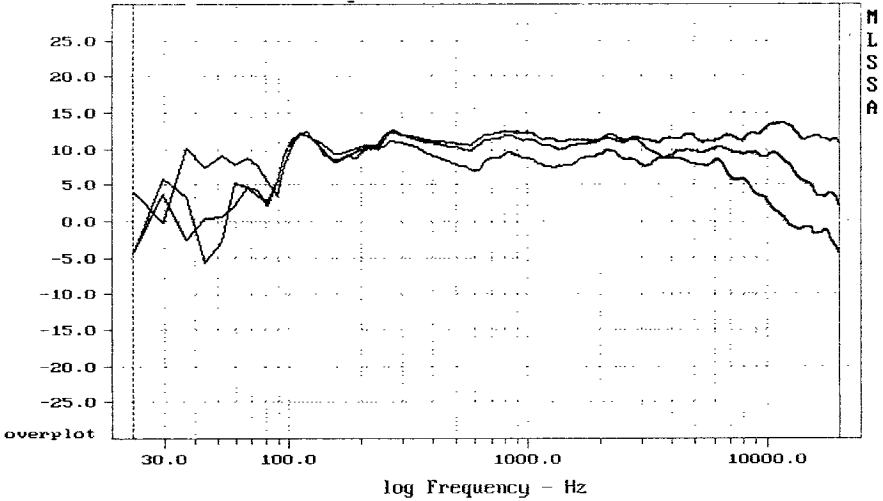
Appendix 3 (a)   Free-field frequency response measurements of loudspeaker "G" showing (top to bottom) 0° (on the listening axis), 30° off axis horizontally and 60° off axis horizontally. The data were 0.17-octave smoothed. Data below about 100 Hz are not accurate.
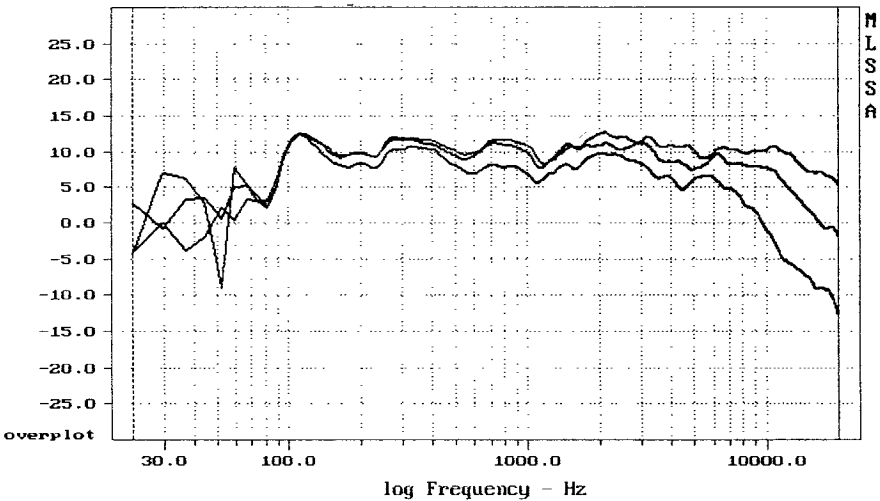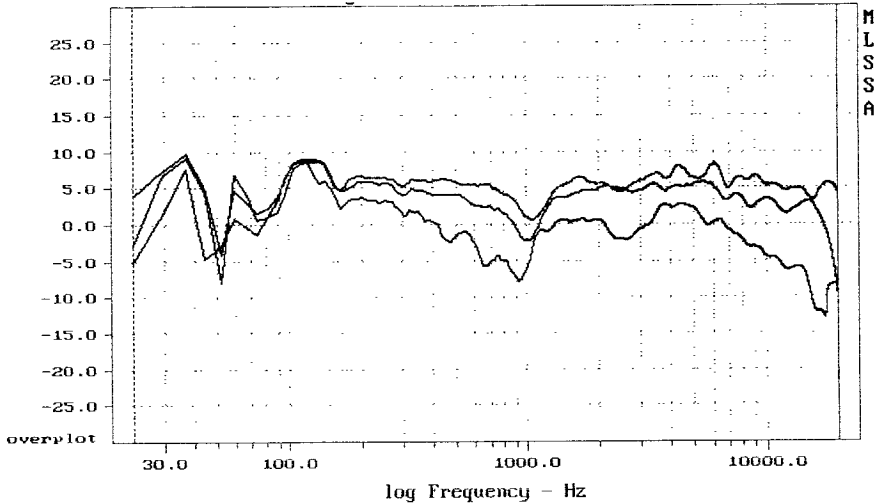

Appendix 3 (b)   Free-field frequency response measurements of loudspeaker "D" showing (top to bottom) 0° (on the listening axis), 30° off axis horizontally and 60° off axis horizontally. The data were 0.17-octave smoothed. Data below about 100 Hz are not accurate.

APPENDIX 3

Appendix 3 (c)   Free-field frequency response measurements of loudspeaker "S" (*satellite only, no subwoofer*) showing (top to bottom) 0° (on the listening axis), 30° off axis horizontally and 60° off axis horizontally. The data were 0.17-octave smoothed. Data below about 100 Hz are not accurate.

Appendix 3 (d)   Free-field frequency response measurements of loudspeaker "T" showing (top to bottom) 0° (on the listening axis), 30° off axis horizontally and 60° off axis horizontally. The data were 0.17-octave smoothed. Data below about 100 Hz are not accurate.