Stereophonic Localization in the Presence of Boundary Reflections, Comparing Specular and Diffuse Acoustic Radiators 4684 (P14-5)

Neil Harris
NXT Huntingdon UK

Sheila Flanagan
NXT Huntingdon UK

Malcolm O. J. Hawksford
University of Essex Essex UK

Presented at the 104th Convention 1998 May 16-19 Amsterdam

AN AUDIO ENGINEERING SOCIETY PREPRINT

Stereophonic localization in the presence of boundary reflections, comparing specular and diffuse acoustic radiators.

Neil Harris*
Sheila Flanagan**
Malcolm O. J. Hawksford***

Abstract

Experiments on sound localization using specular radiators are often performed under near anechoic conditions which do not correspond closely to normal listening environments. A hypothesis is proposed that localization precision as a function of room acoustics is minimized by the use of diffuse acoustic radiators such as DML panels. This proposition is investigated, and the results from a series of psychometric tests presented to establish the conditions under which this approach is valid.

0 Introduction

The subjective effects of the listening environment on listening tests are well documented [1,2]. In order to avoid these variations, image localization tests are usually carried out under near anechoic conditions, with a single, well-defined listener location. In normal use, however, loudspeakers are listened to in rooms with reflective boundaries, and by more than one person at a time. Conventional loudspeakers, as specular acoustic radiators, are particularly susceptible to acoustic room modes and boundary interference, and have a limited "sweet-spot" for optimal listening. In contrast, diffuse acoustic radiators such as the Distributed Mode Loudspeaker (**DML**) [3], are cited as having "sympathetic boundary interactions" and no sweet-spot [4]. A series of psychometric listening tests were proposed to compare the ability of these two classes of radiator to localize a stereo image in an untreated room. The results of the first of these tests are presented here. The nature and acoustical properties of DMLs are discussed elsewhere, e.g. [5,6].

1 The test procedure

The format of the experiment followed the "2 Alternate Forced Choice" method (2AFC). This was a double blind test, following the format recommended by Lipshitz and Vanderkooy [7], the test sequence and signals being generated randomly in software [8]. The signal format was two channel, dual mono, 44.1Khz, 16 bit. The replay levels were adjusted to give a Sound Pressure level (SPL) at the subject's position of 70dBA (Background traffic noise = 40dBA).

* New Transducers Ltd, Huntingdon, UK. AES Member
** New Transducers Ltd, Huntingdon, UK. AES Student Member
*** University of Essex. AES Fellow

A plan of the room used for the experiments is included as figure 1, which also indicates the placement of loudspeakers and listener. The general form for the arrangement of the subject and acoustic radiators is given in figure 2. Two loudspeaker sources are used and are placed equidistantly, R metres (2.75m, typically) from each other and the subject. Thus the speakers are placed at an angle of 30° either side of the subject's straight-ahead position.

The test consisted of a sequential presentation of two bursts of pink noise. Each burst was three seconds long separated by a gap of one second. The bursts were of identical origin, only one had been amplitude panned to one of a predefined range of image positions. The presentation order was also randomised. This was repeated 32 times. The first six pairs of bursts were predefined in the form of a preamble or teaching session, the remaining 28 being the randomised test signals. Starting with a large displacement angle Φ, the preamble is there to familiarise the subject to what is required, and to assure them that the test is actually working.

Subjects were required to identify which of the two bursts had been panned, and record this on a response form. They were told that if they were unsure of the answer they were to guess. The results for these preamble pairs were noted on the response form for the subjects' benefit.

Five tests were conducted, identified as Test1 through Test5, each test representing a different loudspeaker system or placement. Test 1 was used to "try the waters" with a few selected volunteers, and its results are not included in the subsequent analyses.

2 Anecdotal evidence

Feedback from Test 1 allowed the authors to refine the method of recording responses. Most complaints were due to the fact that engineers hate reading instructions! - As a result, the instructions were read to each participant before each test. The biggest problem seemed to be that people don't like being forced to choose between two apparently identical presentations.

All the tests were amicably received, with the consensus of opinion being that the conventional hi-fi box speakers used in Test 3 produced the tightest image. Many listeners observed that images to the right of centre were easier to assess than those to the left of centre.

As a result of this verbal feedback, the authors were expecting that Test 3 results would stand out from the rest, leaving the original hypothesis open until off-axis listening tests are concluded. We were surprised, therefore, to discover that Test 3 gave the worst results, on inspection. Our suspicions were partially justified, when a comparison on listener performances showed that listener 10 in Test 3 was seriously at variance with the group average. Test 3a is therefore merely Test 3, with listener 10's results excluded. Identification of each test set-up is found in the results summary given in Figure 20.

3 Results

It is shown in the appendix that the number of correct results for a particular angle and loudspeaker follows the binomial distribution with probability $p_c = (p_r+1)/2$, where p_r is the probability that the loudspeaker can resolve that angle. Consequently, all the results are displayed with means and 2.5% to 97.5% confidence limits using the binomial distribution. Figures 3, 7, 11 and 15 reflect tests 2, 3a, 4 and 5, respectively. In addition, figure 19 shows the average results of all four tests, with each individual result overlaid. A crude ranking of the different loudspeaker systems can be formed by subtracting the number of results that were significantly worse than average from number of results that were significantly better than average, as shown in figure 20.

4 Analysis of results

Confidence testing against $p_r = 0$ (i.e. no one can tell, or everyone is guessing) prove almost nothing - nearly all the results are significant at 95% - see figures 4, 8, 12 and 16. We therefore propose two tests to determine which side of the *threshold of detection* the results lie.

The threshold of detection is taken to be when $p_r = 0.5$, hence at this threshold we expect 75% successful detection (i.e. $p_c = 0.75$). The cumulative probability that each test result is on the threshold is displayed, along with the two-sided 95% confidence limits, in figures 5, 9, 13 and 17. The null hypothesis, H_0, is that we are at threshold of detection - results above the 97.5% limit are definitely above the threshold; results below the 2.5% limit are definitely below the threshold. The angular resolution of each loudspeaker system by this method is tabulated in figure 20.

The second test seeks to equalise type I and type II errors [9]. The three hypotheses being tested are; H_0, we are below threshold of detection, p_r=0; H_1, we are at threshold of detection, p=0.5, H_2, we are above threshold of detection, p=1. This method of testing results in two critical probabilities, which reject first H_0, then H_1. Figures 6, 10, 14 and 18 show the test results scaled so that they cross unity at the critical probabilities. The results which reject H_0 but not H_1 are termed "may hear", and those which reject both H_0 and H_1 are termed "can hear". The angular resolution of each loudspeaker system by this method is tabulated in figure 20.

5 Summary of results

The anecdotal evidence of a bias in the left / right imaging of the room is revealed in the results, with those from test 4 (figure 14) showing it particularly well. This may have been due to a window on the left-hand wall, which over-looks a noisy roundabout.

In contrast to the anecdotal evidence, however, the results show that the conventional speaker system performed the worst, despite being tested from the "sweet spot". Obviously a small image is not necessarily a well-located image. One might say it was lying convincingly, while the others were vaguely truthful!

6 A critical review of the experiment

The major weakness of the experimental procedure used is the variability of bandwidth between the systems under test. Although the total A-weighted sound level was equalised, systems with higher bandwidth would have lower level at the frequencies most useful for localization. With this limitation in mind, the authors sought to ensure that systems were at least similar. In particular, the hybrid cone-DML system used in Test 5 was full range, and directly comparable to the box speaker used in Test 3.

The experiment also fails to show whether the good localization results for the DML systems were due to the diffuse radiation, or to some other feature. Suggestions as to what these may be include uniform directivity [10], and the lack of cross-over problems. It is hoped that anechoic measurements will help to answer this uncertainty.

7 Future work

It is the aim of the authors to refine and expand the testing in a number of ways. These include;

i) using band-limited signals to reduce the effects of frequency response variations and / or frequency-response matched DML and cone systems
ii) repeating the experiments using binaural recordings taken in an anechoic chamber.
iii) repeating the experiments with participants seated away from the "sweet-spot"
iv) using delay instead of / in addition to amplitude for image synthesis
v) using different size DML systems and different conventional systems

We also wish to replace the simple forced choice experiment with a more direct estimation of source direction. The disadvantage of current test is that it only determines that something is *different*, not whether it is *right*. This will be most important for iii) above.

8 Conclusions

Results of a series subjective listening tests seem to confirm the hypothesis that localization precision as a function of room acoustics is minimized by the use of diffuse acoustic radiators such as DML panels. In the tests completed to date, the diffuse radiators performed at least as well as a quality two-way cone-in-box loudspeaker system, even in the "sweet-spot" which should favour the latter. The authors would like to thank their colleagues at NXT for their participation in these tests.

References

[1] M. Barron, *"The Subjective effects of first reflections in concert halls – the need for lateral reflections"*, Journal of Sound and Vibration (1971) **15** (4), (475-494).
[2] Sean E. Olive, Peter L. Schuck, Sharon L. Sally and Marc E. Bonneville, *"The Effects of Loudspeaker Placement on Listener Preference Ratings"*, JAES vol. 42, No. 9, pp 651-669,
[3] A generic term used by New Transducers Limited (NXT) to describe loudspeakers working according to their teaching.
[4] H. Azima & N. Harris, *"Boundary Interaction of Diffuse Field Distributed-Mode Radiators"*, 103rd AES Convention, pre-print #4635
[5] N. Harris & M. Hawksford, *"The Distributed Mode Loudspeaker (DML) as a broad-band acoustic radiator"*, 103rd AES Convention, pre-print #4526.
[6] G. Bank & N. Harris, *"Distributed Mode Loudspeakers - Theory and Practice"*, paper given at AES conference "The Ins and Outs of Audio", in London, England; March 1998
[7] S. P. Lipshitz, J. Vanderkooy, *"The Great Debate: Subjective Evaluation"*, JAES Vol.29, No. 7/8, July/August 1981 .
[8] CoolEdit digital audio editor for Windows, by D.Johnston. Syntrillium Software Corporation.
[9] L. Leventhal, *"Type 1 and Type 2 Errors in the Statistical Analysis of Listening Tests"*, JAES vol. 34, No. 6, pp437-453
[10] D. Queen, *"The Effect of Loudspeaker Radiation Patterns on Stereo Imaging and Clarity"*, JAES vol. 27, No. 5, pp368-379

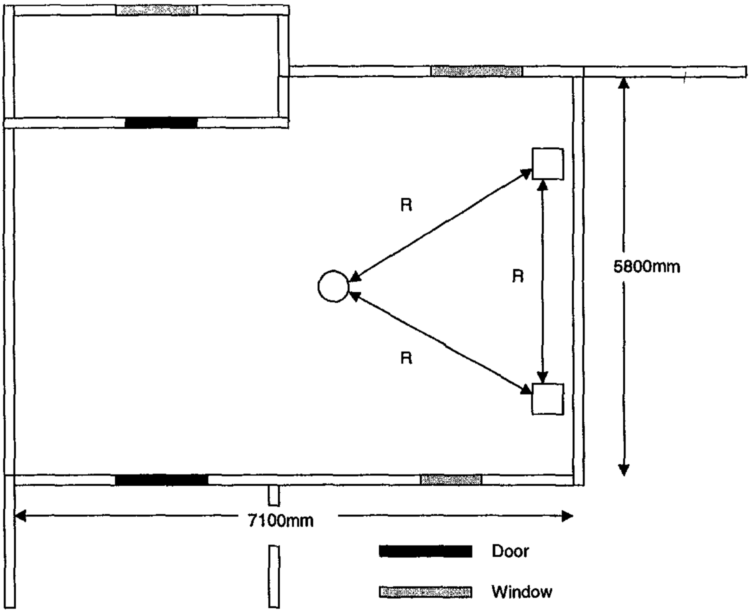

Figure 1. Schematic of the room used for the listening experiments.

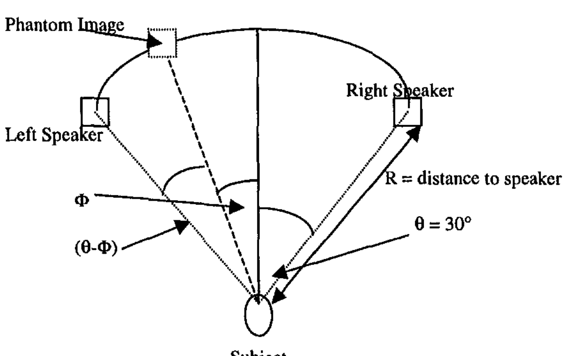

Figure 2. Diagram showing the relative positions of source, subject and 'speakers.

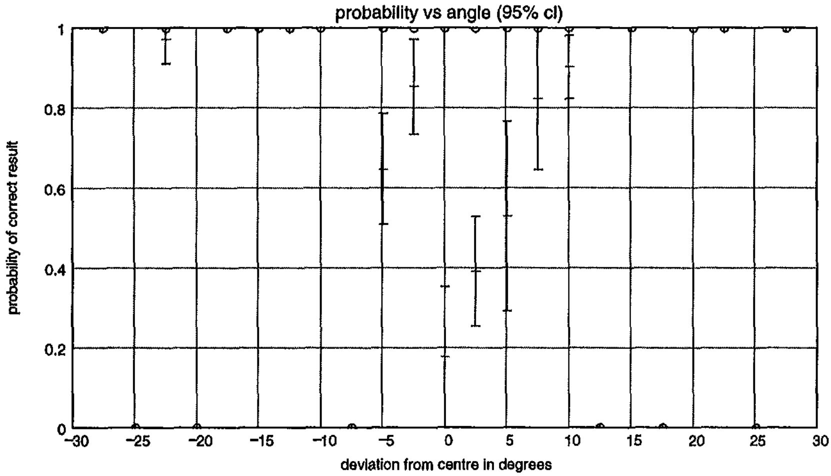

Figure 3. Mean results with 95% confidence limits for Test 2.

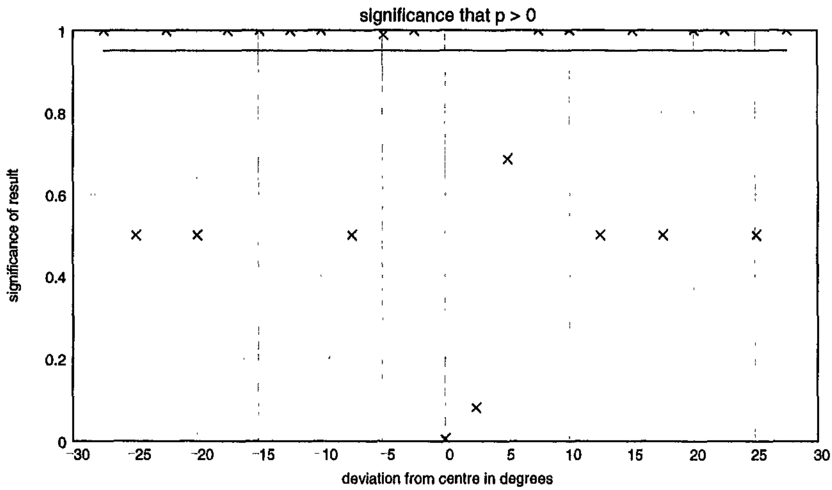

Figure 4. One-sided significance of results being different to p_r=0 for Test 2.

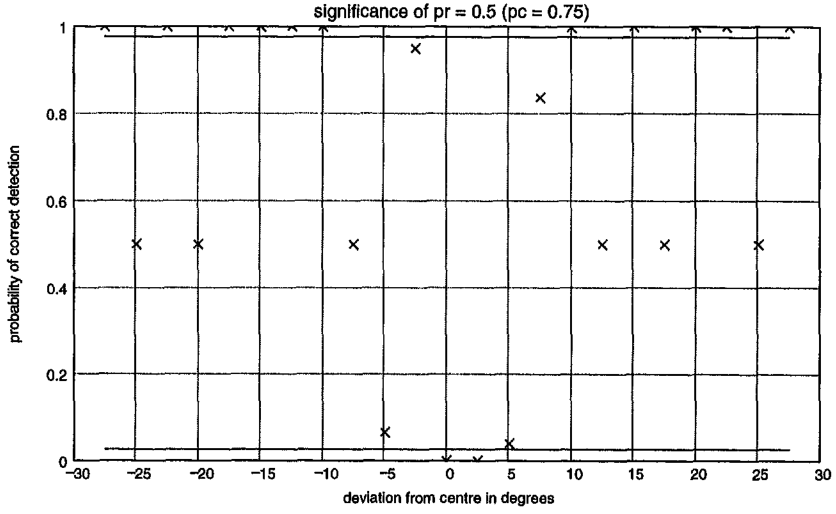

Figure 5. Two-sided significance of results being different to p_r=0.5 for Test 2.

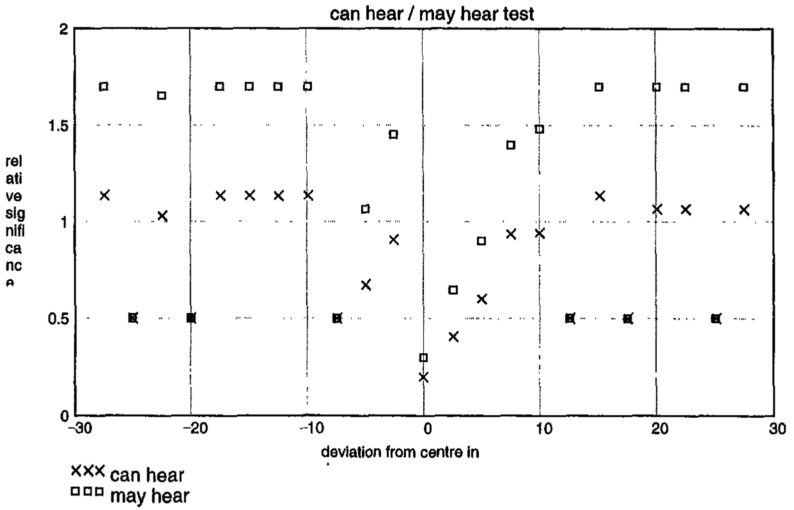

Figure 6. Significance of results being different to p_r=0.5 for Test 2.

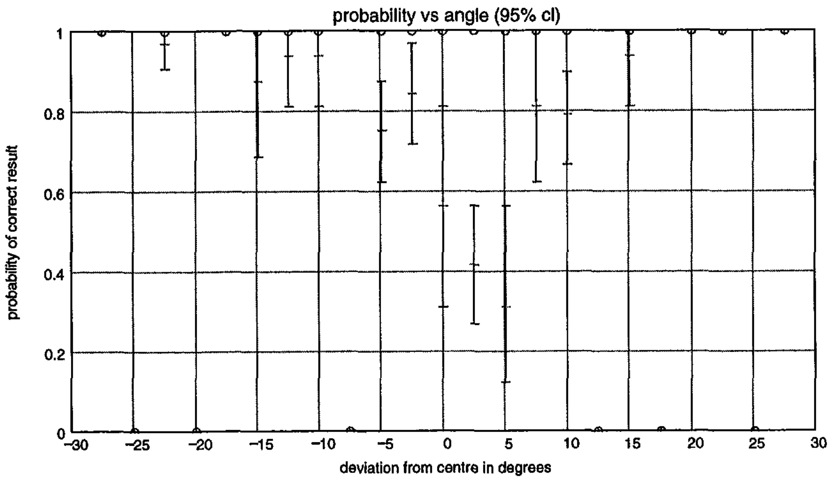

Figure 7. Mean results with 95% confidence limits for Test 3a.

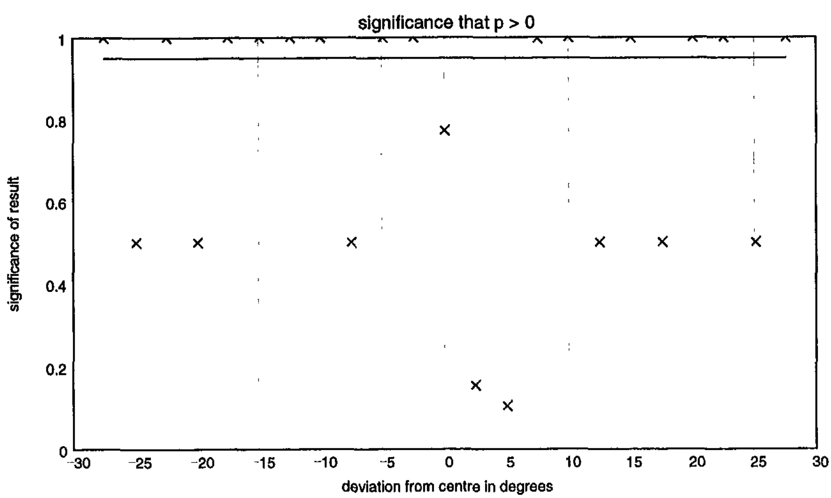

Figure 8. One-sided significance of results being different to p_r=0 for Test 3a.

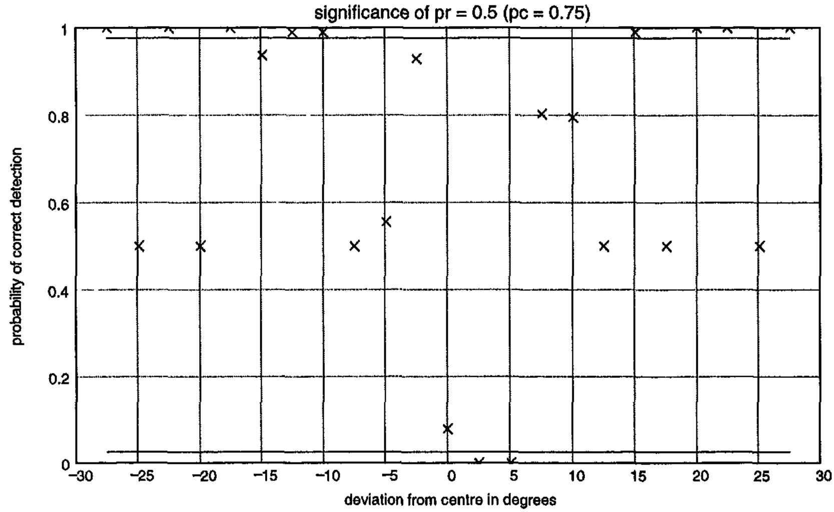

Figure 9. Two-sided significance of results being different to p_r=0.5 for Test 3a.

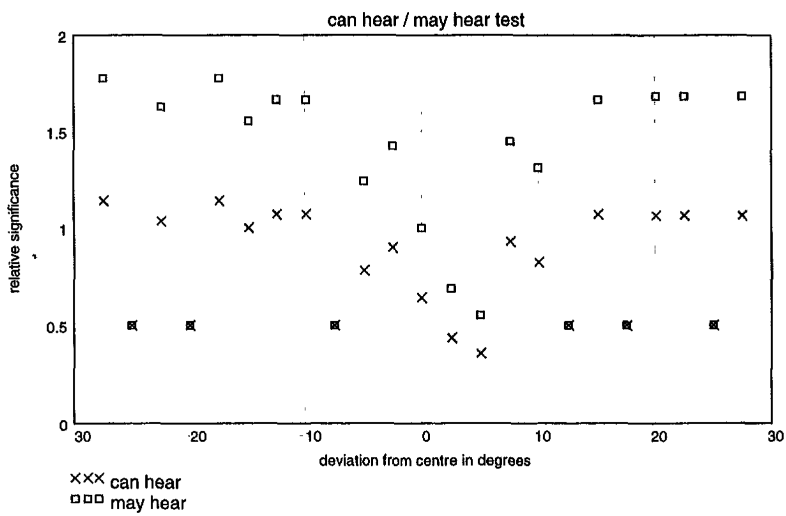

Figure 10. Significance of results being different to p_r=0.5 for Test 3a.

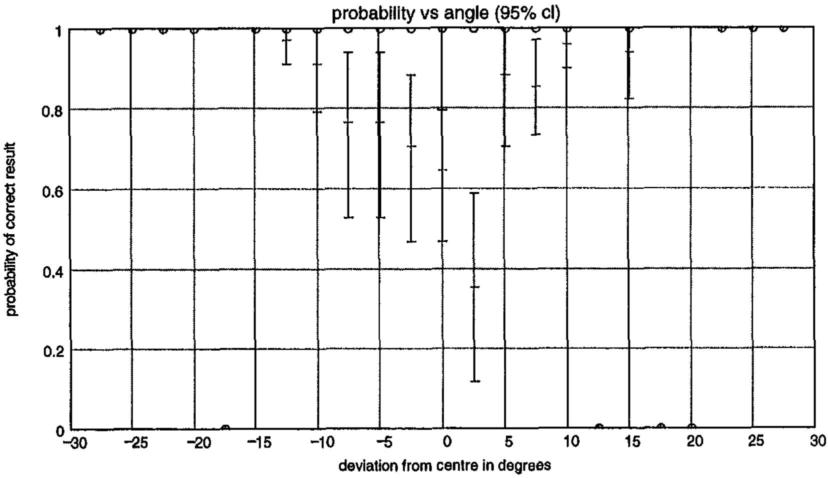

Figure 11. Mean results with 95% confidence limits for Test 4.

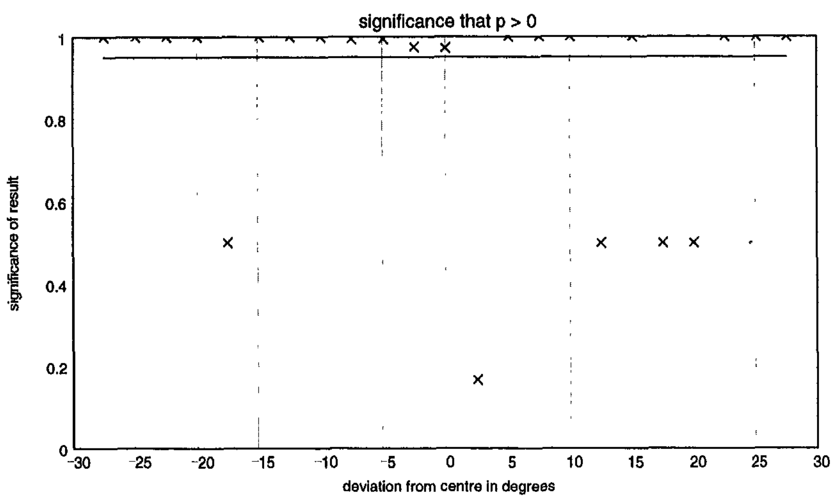

Figure 12. One-sided significance of results being different to p_r=0 for Test 4.

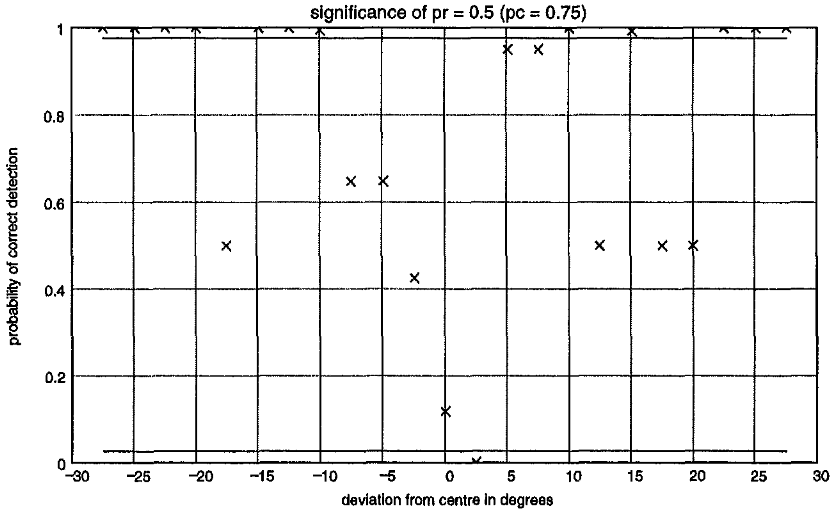

Figure 13. Two-sided significance of results being different to p_r=0.5 for Test 4.

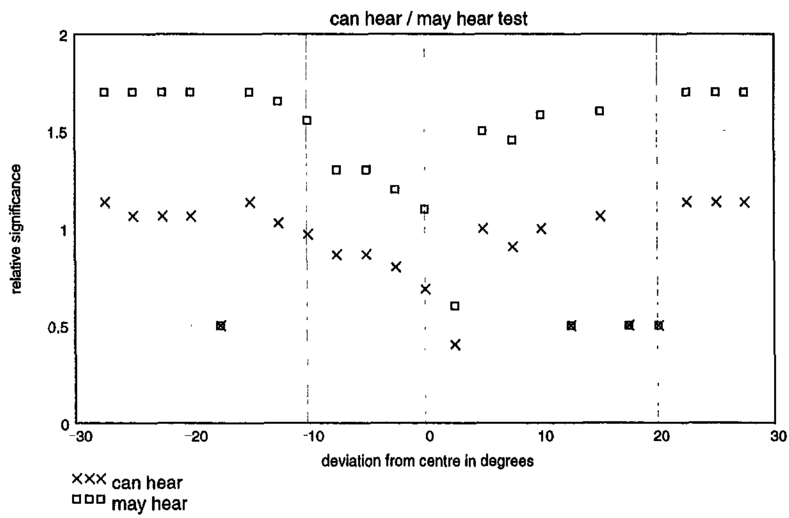

Figure 14. Significance of results being different to p_r=0.5 for Test 4.

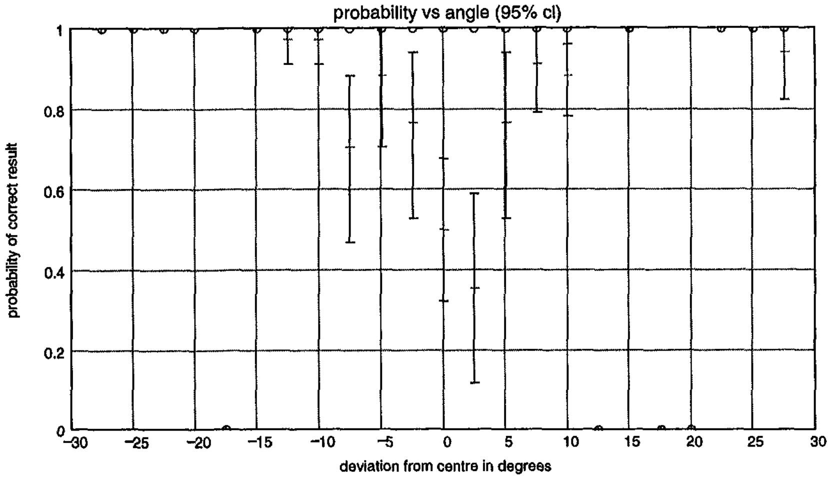

Figure 15. Mean results with 95% confidence limits for Test 5.

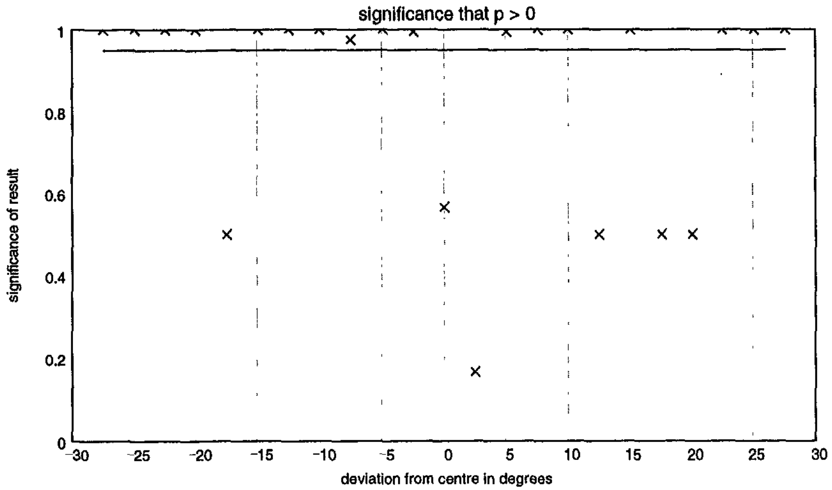

Figure 16. One-sided significance of results being different to p_r=0 for Test 5.

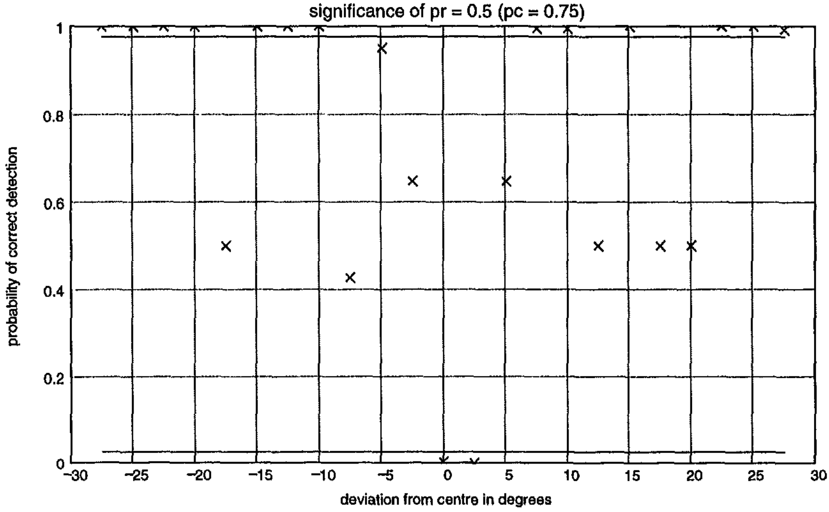

Figure 17. Two-sided significance of results being different to p_r=0.5 for Test 5.

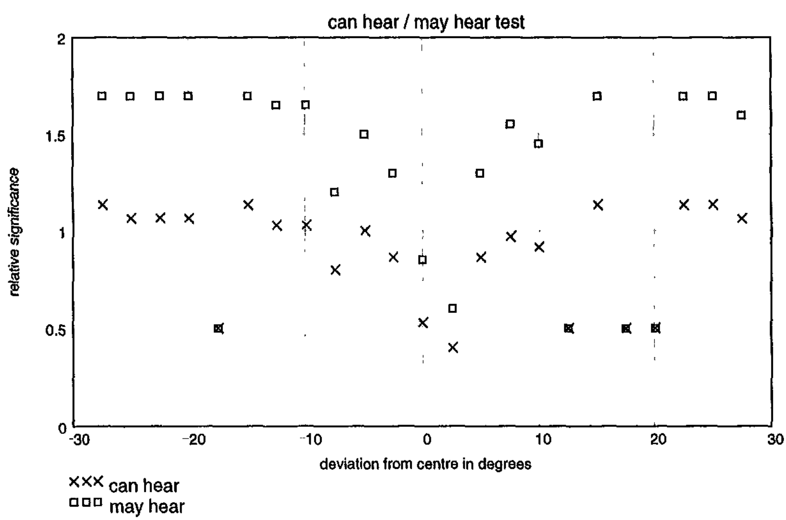

Figure 18. Significance of results being different to p_r=0.5 for Test 5.

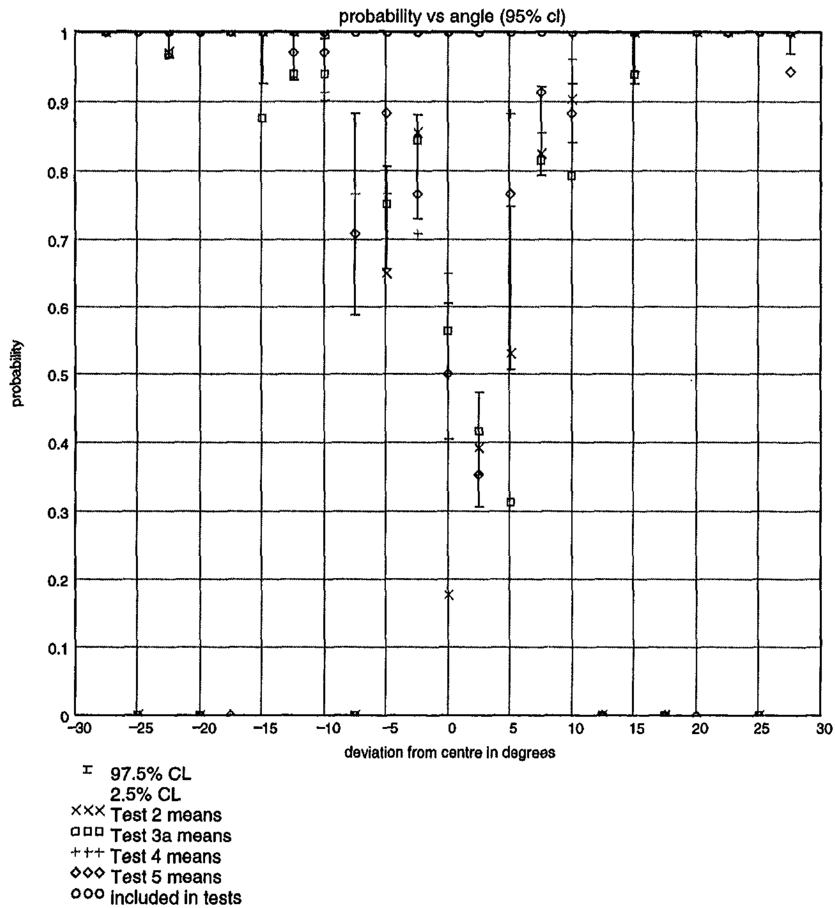

Figure 19. Mean results with 95% confidence limits for all tests.

experiment	description		two-sided p=0.5								
		97.5%	2.5%	2.5%	97.5%		mean left	mean rignt	difference	mean	rank
2	Tiled DML in wall	-10	0	0	10		-5	5	10	5	2
3a	Mission two-way box	-10	2.5	5	15		-3.75	10	13.75	6.875	4
4	Tiled DML off wall	-10	2.5	2.5	10		-3.75	6.25	10	5	2
5	Tiled DML + woofer hybrid	-10	0	2.5	7.5		-5	5	10	5	2

			can hear / may hear test								
experiment	description	can	may	may	can		mean left	mean rignt	difference	mean	rank
2	Tiled DML in wall	-7.5	-2.5	5	10		-5	7.5	12.5	6.25	2
3a	Mission two-way box	-10	0	7.5	12.5		-5	10	15	7.5	4
4	Tiled DML off wall	-12.5	0	5	5		-6.25	5	11.25	5.625	1
5	Tiled DML + woofer hybrid	-10	-2.5	5	10		-6.25	7.5	13.75	6.875	3

experiment	description		ranking by significance		
		better	worse	delta	rank
2	Tiled DML in wall	0	2	-2	3
3a	Mission two-way box	0	3	-3	4
4	Tiled DML off wall	2	1	1	1.5
5	Tiled DML + woofer hybrid	2	1	1	1.5

Notes:

The "tiled DML" was demonstrated at the 103rd AES convention in New York
The "tiled DML + woofer hybrid" was demonstrated during the 1998 CES in Las Vegas
The "Mission two-way box" is the 752 model. Mission is a sister company of NXT

Figure 20. Overall results summary for listening tests

Appendix

Assume a binomial distribution, with p(able to resolve) = p_r.
Then in N trials, the probability of n correct detections, p_n is given by

$$p_n(n) = \binom{N}{n} \cdot p_r^{\,n} \cdot q_r^{\,N-n} \quad \text{where } q_r = 1 - p_r, \text{ and } \binom{N}{n} = {}_N C_n = \frac{N!}{(N-n)! \cdot n!}$$

If we assume that those who don't correctly detect just guess fairly, then 50% of them will get the right answer, i.e. p(guess) = q(guess) = 1/2. So, for n correct detections, the conditional probability of m lucky guesses is

$$p_{n|m}(n,m) = \binom{N-n}{m} \cdot \left(\frac{1}{2}\right)^m \cdot \left(\frac{1}{2}\right)^{N-n-m} = \binom{N-n}{m} \cdot \left(\frac{1}{2}\right)^{N-n}$$

So the probability, p_k, of k correct responses is just the sum of all probabilities where n+m=k, i.e

$$p_k(k) = \sum_{i=0}^{k} p_n(i) \cdot p_{n|m}(i, k-i)$$

$$p_k(k) = \sum_{i=0}^{k} \binom{N}{i} \cdot p_r^{\,i} \cdot q_r^{\,N-i} \cdot \binom{N-i}{k-i} \cdot \left(\frac{1}{2}\right)^{N-i}$$

$$p_k(k) = \sum_{i=0}^{k} \frac{N!}{(N-i)! \cdot i!} \cdot p_r^{\,i} \cdot q_r^{\,N-i} \cdot \frac{(N-i)!}{(N-k)! \cdot (k-i)!} \cdot \left(\frac{1}{2}\right)^{N-i}$$

$$p_k(k) = \left(\frac{1}{2}\right)^N \cdot \sum_{i=0}^{k} 2^i \cdot \frac{N!}{(N-k)! \cdot (k-i)! \cdot i!} \cdot p_r^{\,i} \cdot q_r^{\,N-i} = \frac{N!}{(N-k)! \cdot k!} \cdot \left(\frac{1-p_r}{2}\right)^{N-k} \cdot \left(\frac{1+p_r}{2}\right)^k$$

It is seen by inspection that we have another binomial distribution, with modified probability p_c given by

$$p_c = \frac{1+p_r}{2} \qquad q_c = 1 - p_c = \frac{1-p_r}{2} \qquad p_k(k) = \binom{N}{k} \cdot p_c^{\,k} \cdot q_c^{\,N-k}$$

$$E(k) = N \cdot p_c = \frac{N}{2} \cdot \left(1 + p_r\right) \qquad \mathrm{var}(k) = N \cdot p_c \cdot q_c = \frac{N}{4} \cdot \left(1 - p_r^{\,2}\right)$$