Australian Journal of Audiology 1979 1:2 pages 83 – 88

BIAS IN ASSESSING BINAURAL ADVANTAGE

DENIS BYRNE
and
HARVEY DILLON

National Acoustic Laboratories, 5 Hickson Road, Sydney, Australia.

Abstract

Attention is drawn to a bias which may occur in assessing binaural advantage in research studies and in individual clinical cases. It applies to any procedure in which binaural advantage is defined as the difference between binaural performance and the performance of whichever ear performs better. Binaural performance should be compared with the performance of the better ear but in some instances this does not happen because test variability will result in a higher score being obtained with the poorer ear. When these higher scores from the poorer ear are substituted for the lower scores of the better ear the mean "better" ear score is inflated and binaural advantage is underestimated. A method is presented for calculating the degree of bias, and procedures for minimizing bias are discussed.

A large body of research has been directed towards assessing the possible advantages of binaural hearing aid fitting. There is now extensive evidence to support this practice which has, especially over the last few years, become widely accepted. Nevertheless, there are still many audiologists who are not convinced of the benefits of binaural fitting or who regard any such benefits as insufficient to justify the extra trouble and expense involved. We do not intend, in this article, to discuss binaural fitting as such, but simply to draw attention to a source of bias which may lead to underestimation of binaural advantage. This bias is liable to occur over a wide range of experimental studies and in the assessment of possible binaural advantage in the individual clinical case. It has been mentioned with reference to binaural summation at threshhold (Shaw, Newman & Hirsh, 1947) but it does not seem to be recognized as a problem in assessing other types of binaural advantages. This is unfortunate since the degree of bias is related to the test-retest variability of the particular measure of binaural advantage. It is, therefore, especially significant for tests having a high degree of test-retest variability, and this category includes the most commonly used test of binaural advantage, namely speech discrimination testing.

The bias in question may occur in any procedure in which binaural advantage (or lack of it) is defined as the difference between binaural performance and the monaural performance of the better ear. This practice is reported in some early studies of binaural hearing aid fitting (e.g. Di Carlo and Brown, 1960) as well as in some recent studies (e.g. Lankford and Faires, 1973; Seigenthaler, 1978; Grimes, Mueller and Sweetow, 1979). In principle, this definition of binaural advantage is completely reasonable since presumably the alternative to binaural aid fitting is to fit whichever ear provides the better monaural performance. The bias arises because, in practice, the better ear cannot be identified consistently and in some instances, the better ear score is discarded in favour of a higher score which, by virtue of test-retest variability, happened to be obtained with the poorer ear. The following examples should give a general indication of the nature and extent of bias arising in this fashion.

First, let us suppose that a particular subject has equal discrimination for all three test conditions namely, binaural, monaural left ear, monaural right ear. In other words, his true discrimination score, defined as the mean score of a large (theoretically infinite) number of tests, is equal for all conditions. However, if each condition were tested only once, test-retest variability would usually produce different scores for each of the three conditions. If we repeated this procedure (i.e. testing each condition once) for a large number of times, for this subject (or for different subjects who all had equal discrimination across the three conditions) we would expect that each condition would obtain the highest score on approximately one-third of the trials. On this basis we would find a "binaural advantage" for one case in three and a "monaural advantage" for two thirds of the cases. Thus, when averaged over the whole group of tests, the "better" ear monaural performance would appear to be superior to binaural performance although they were, in reality, equivalent. Similarly, where there is, in fact, a binaural advantage, this may not be shown because it may be offset by the bias described above.

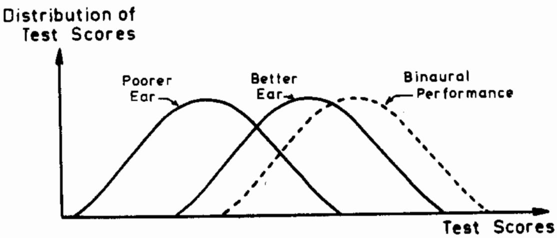

Figure 1.

An example of the distribution test scores for poorer ear, better ear and binaural performance.

This bias may be examined further by referring to Figure 1 which shows a hypothetical distribution of scores for the three test conditions. These could be considered as the scores for one subject tested many times or alternatively as scores for a group of subjects, tested once for each condition. (In the latter instance, the better ear could be the left ear for some subjects and the right ear for others). To assess binaural advantage we should compare the distribution of scores for binaural performance with the distribution for the better ear monaural performance. However, since the better and poorer ear distributions overlap, there will be some trials in which the higher score will be obtained with the poorer ear which will therefore, be misidentified as the "better" ear. Thus, some of the scores from the better ear distribution will be replaced by higher scores from the poorer ear distribution. The frequency with which the "better"ear will be misidentified, and hence the extent of bias which is introduced, will depend on the difference between the true scores of the better and poorer ears and the spread of the scores for each condition. Misidentification will occur quite often if, as in the figure, the true scores are relatively close and the distributions have a wide spread. This would be fairly representative of the situation which occurs when binaural advantage is assessed by speech discrimination testing, since large inter-aural differences in discrimination are rather exceptional, and since test-retest variability typically shows a standard deviation of 6 to 7% for a 50 item test of moderate difficulty (e.g. Shore et al, 1960; Beattie & Edgerton, 1976). The chance of misidentification will, of course, be reduced if the difference between the better and poorer ear is larger or if test variability is less.

To calculate precisely the extent of bias it would be necessary to know the true scores of the better and poorer ears for each subject and the distribution of scores which would be obtained if each subject were tested numerous times. In the absence of such data it is possible to estimate the approximate degree of bias by making certain assumptions which are consistent with other data. To provide a simple but realistic example, let us assume that each subject's scores are normally distributed for each ear; that the difference between the true (i.e. mean) scores for the two ears is 7% and that the distributions for both ears have standard deviations of 7%. Thus, the distribution curves will intersect at the point which is 0.5 standard deviations below the mean of the better and 0.5 standard deviations above the mean of the poorer ear. If we conducted a large series of trials, in which each of the two conditions was tested once, we would expect the poorer ear to produce the higher score in 24% of trials (see Appendix). If we calculated a mean score for the "better" ear, defined as whichever ear produced the higher score on each trial, this would be 1.3% higher than the true mean score of the better ear (see Appendix). This degree of bias cannot be disregarded, considering that binaural advantage as measured by speech discrimination tests, can only be expected to amount to about 6% to 8% (Dermody & Byrne, 1975). Furthermore, the amount of bias may well be greater in some studies than in this example, either because the true inter-aural differences are smaller or because the test variability is greater. The latter may occur for various reasons, such as the use of tests with fewer items. For example, if the true scores of both ears were the same, and the standard deviations of both distributions remained at 7%, the mean "better" ear score would be inflated by 4%. If in addition to the true scores being the same, the standard deviations were increased to 10%, the degree of bias would amount to 5.6%.

Table I

Bias introduced by selecting the better of two scores for each trial (i.e. amount by which the mean of the "better" scores will exceed the true mean of the better ear).

DIFFERENCE IN MEANS (NO. OF STANDARD DEVIATIONS)	AMOUNT OF BIAS (NO. OF STANDARD DEVIATIONS)
0	0.56
0.5	0.34
1.0	0.19
1.5	0.10
2.0	0.05
2.5	0.02

Table I shows the amount of bias that will be introduced for a given difference in inter-aural performance. Provided this difference is expressed in terms of the number of standard deviations of the distribution of scores for each ear, the bias is independent of the actual values of the means or standard deviation. The values shown on the table are based upon the equations derived in the appendix when applied to the case of a normal distribution of test scores. For a test containing N items with an overall probability of discrimination P, the normal distribution is good approximation of the binomial distribution provided both NP and N(1−P) are greater than 10 (Hays, 1969). When the number of test items is small, or if very high or low scores are obtained, the expected bias may be obtained by using Tables II through V (for 10, 25, 50 and 100 test items respectively). Note that the lower and difference scores used in obtaining the bias from these tables should be the mean and difference in means of the distributions on which they are based. As these will usually be unknown they may be replaced by their best estimate, the actual scores obtained in a single test.

The bias described in this article poses a problem for both the research worker and the clinician who needs to assess binaural advantage. An alternative to comparing binaural performance with "better" ear monaural performance is to compare it to the average of the two monaural conditions (e.g. Stearns and Lawrence, 1977). However, this could introduce a serious bias in the opposite direction and usually could not be justified except perhaps for some experiments with normal hearing listeners where it might be reasonable to assume that there would be little inter-aural difference in true performance. Another option is to compare binaural performance with the performance of the ear, selected on the basis of audiological test data and other general considerations, as being the more suitable for aid fitting (e.g. Jerger, Carhart and Dirks, 1961). While it could be said that this is a realistic approach, it must be admitted that there is some risk that the selected ear may not provide the better (true) performance on the test being considered and thus binaural advantage may be inflated. We believe that, for most purposes, the best approach is to compare binaural and better ear monaural performance but that the bias described above should be recognized and minimized. The most effective way to reduce this bias would be to test each condition (i.e. binaural, left ear, right ear) several times for each subject. The "better" ear would be defined as the ear having the higher mean score and this would be compared with the mean score for the binaural condition. The advantages of this procedure, compared to using a single trial (i.e. one test under each condition) for each subject, are that the "better" ear is less likely to be misidentified and that better estimates of true performance are obtained for both the monaural and binaural conditions. Statistically the use of repeated tests is equivalent to presenting the same total number of items in a single trial. However, there may be practical advantages in using repeated tests, such as avoiding the risk of fatigue which could occur if a single, lengthy test were used.

The use of several test for each condition may also provide a rough estimate of the test-retest variability applying in the individual and this may possibly be of value in assessing the likelihood of bias. For example, if all of several scores obtained with one ear were better than any of several scores obtained with the other, we could be confident that the "better" ear had been correctly identified.

A comparison of corresponding entries in Tables II through V will show the reduction in bias that occurs when longer tests are used. If scores of around 50%, for example, are individually obtained from both ears, the expected bias will be 9, 6, 4 and 3 percentage points for 10, 25, 50 and 100 item tests respectively.

Table II

Bias incurred (in percentage points) as a result of selecting the maximum of two scores based on 10-item tests.

N = 10			
LOWER SCORE (%)	DIFFERENCE IN SCORES (PERCENTAGE POINTS)		
	0	10	20
10	5	2	1
20	7	4	2
30	8	4	2
40	9	5	2
50	9	5	2
60	9	4	2
70	8	4	1
80	7	2	-
90	5	-	-

Table III

Same as Table II but for 25-item tests.

N = 25						
LOWER SCORE (%)	DIFFERENCE IN SCORES (PERCENTAGE POINTS)					
	0	4	8	12	16	20
10	3	2	1	1	0	0
20	4	3	2	1	1	0
30	5	3	2	1	1	0
40	5	4	2	2	1	0
50	6	4	2	1	1	0
60	5	4	2	1	1	0
70	5	3	2	1	0	0
80	4	3	1	0	0	-
90	3	1	0	-	-	-

Table IV

Same as Table II but for 50-item tests.

	N = 50			
LOWER SCORE (%)	DIFFERENCE IN SCORES (PERCENTAGE POINTS)			
	0	5	10	15
10	2	1	0	0
20	3	1	1	0
30	4	2	1	0
40	4	2	1	0
50	4	2	1	0
60	4	2	1	0
70	4	2	1	0
80	3	1	0	0
90	2	0	0	0

Table V

Same as Table II but for 100 item tests.

	N = 100		
LOWER SCORE (%)	DIFFERENCE IN SCORES (PERCENTAGE POINTS)		
	0	5	10
10	2	0	0
20	2	1	0
30	3	1	0
40	3	1	0
50	3	1	0
60	3	1	0
70	3	1	0
80	2	0	0
90	2	0	-

Assuming that the better ear monaural performance has been assessed without excessive bias, there remains the problem of determining whether the difference between binaural performance and better ear monaural performance is significant. This will not be discussed here, except to join others in pointing out that the commonly used procedure of accepting as significant any difference exceeding a fixed percentage (typically 6% to 8%), is invalid. The fact is, that the difference required to reach any specified level of confidence will depend on the number of test items and the values of the true scores. This issue is discussed in detail by Thornton and Raffin (1978) who present tables for determining critical differences.

It is appreciated that, in some instances, the clinician or research worker may find it impractical to conduct sufficient testing that he can be confident that no significant bias remains in his assessment of binaural advantage. If this is so, an allowance for bias can be made using the figures shown in the tables. If the appropriate table entry is subtracted from the better of the two scores the resulting score will be the best estimate of the true better ear performance. Although such a corrected score would, in most cases, be closer to the 'true' score than the obtained, uncorrected score, it should be realized that this will not be so in every individual case because test-retest variability can sometimes result in an underestimate of the better ear score despite the bias in the opposite direction. Thus, although there is a bias towards underestimating binaural advantage for the majority of cases there will still be some cases where binaural advantage is overestimated because of test variability.

In summary, the assessment of binaural advantage may present problems which, in some circumstances, are difficult to overcome. However, at the very least, the audiologist should be aware of the possiblility of bias and should avoid being too hasty in concluding that there is no binaural advantage when, in fact, such an advantage may exist but is not evident because it has been reduced to insignificance by a bias which has inflated the 'better' ear monaural performance.

References

Beattie, R.C. and Edgerton, B.J. 1976. Reliability of monosyllabic discrimination tests in white noise for differentiating among hearing aids. *J. Speech Hear. Res.* 16, 464–476.

Dermony, P., and Byrne, D., 1975. Variability in speech test performance for binaural hearing aid evaluations. National Acoustic Laboratories, Report No. 62, Sydney, August, 1975.

Di Carlo, L. M., and Brown, W. J. 1960. The effectiveness of binaural hearing for adults with hearing impairments. *J. Aud. Res.* 1, 35–76.

Grimes, A.M. Mueller, H.G. and Sweetow, R.W. 1979. Monaural versus binaural aided discrimination in hearing-impaired children. *Audiol & Hear. Education* 5, 6–11.

Hays, W. L. 1969. Statistics. Holt, Rinehart and Winston, London.

Jerger, J., Carhart, R. and Dicks, D. 1961. Binaural hearing aid and speech intelligibility. *J. Speech Hear. Res.* 4, 137–148.

Lankford, S. E., and Faires, W. L. 1973. Objective evaluation of monaural versus binaural fitting. *J. Aud. Res.* 13, 263–267.

Seigenthaler, B. 1978. The non-universal binaural hearing advantage. XIV International Congress of Audiology, Acapulco, Nov. 1978.

Shaw, W. A., Newman, E.B. and Hirsh, I.J. 1947. The difference between monaural and binaural threshholds. *J. Exp. Psychol.* 37, 229–242.

Shore, I. Bilger, R.C. and Hirsh, I.J. 1960. Hearing aid evaluation: variability of repeated measurements. J. Speech Hear. Dis. 25, 152–170.

Stearns, W.P. and Lawrence, D.W. 1977. Binaural fitting of hearing aids. *Hearing Aid J.* April, 12 and 51–53.

Thornton, A.R. and Raffin, M.J. 1978. Speech-discrimination scores modeled as a binomial variable. *J. Speech Hear. Res.* 21, 507–518.

APPENDIX

We may ask how often will a score from the poorer ear be higher than a score from the better ear. Consider the two normal distributions shown below in Figure 2.

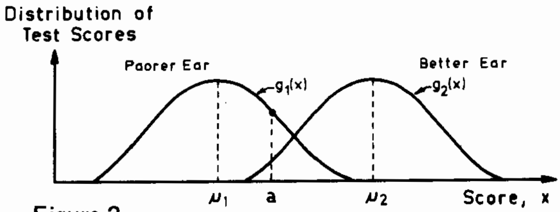

Figure 2.

Distribution of test scores for two ears with differing abilities in an auditory task (e.g. speech discrimination).

If particular values, x_1 and x_2 are drawn from $g_1(x)$ and $g_2(x)$ respectively, the poorer ear will be incorrectly selected whenever $x_1 > x_2$. The probability of this occuring can be found by considering the difference distribution $g(x_2 - x_1)$ which is shown in Figure 3.

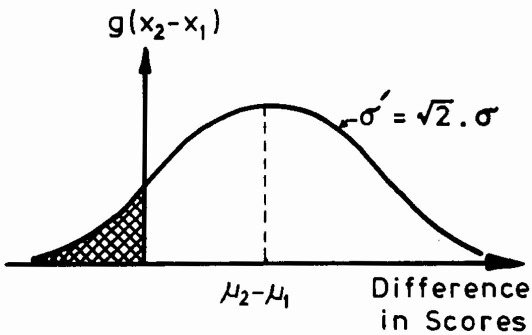

Figure 3.

The probability distribution of the difference between the two random variables whose individual distributions are shown in Figure 1.

The standard deviation of this distribution σ', is equal to $\sqrt{2}$ times the s.d. of the original distributions σ, which are assumed to be equal (Hays, p. 236). The probability of $x_1 > x_2$ is equivalent to the probability of having $x_2 - x_1 < 0$ which in turn is equal to the hatched area in Figure 2. This can be found from the cumulative normal distribution table using a normalised ordinate of $(N_2 - N_1)/\sigma$

Given that the incorrect score is sometimes selected as coming from the distribution with the higher mean, we may ask how large will be the error caused by this. The following derivation finds the expected value of the maximum of two scores, each drawn from a different distribution.

With reference to Figure 2, let a particular value of x_1 drawn from distribution $g_1(x)$ be "a". The probability that any value of x_2 drawn from $g_2(x)$ will be greater than "a" is then given by:

$$P(x_2 > a) = \int_a^\infty g_2(x)\,dx = A \qquad ....(1)$$

$$\text{Thus, } P(x_2 > a) = 1-A.$$

Since x_2 will be the selected score **A** of the time, and "a" will be the selected score 1−A of the time, the expected value of the maximum of x_2 and "a" will be:

$$E\left[\max(a, x_2)\right] = a.(1-A) + E\left[x_2|_{x_2>a}\right].A. \qquad ...(2)$$

Notice that the expected value of x_2 given that x_2 is greater than "a" is not equal to the expected value of x_2, since all values lower than "a" are excluded. In fact the distribution of $E\left[x_2 \middle| x_2 > a\right]$ will have the form shown in Figure 4.

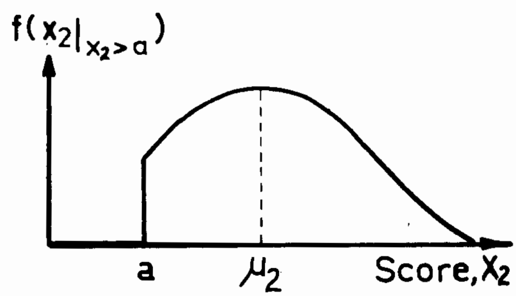

Figure 4.

The probability distribution of the variable x_2, given that it is greater than "a".

This distribution is simply related to the original distribution $g_2(x)$ by the need for both distributions to have an area of unity.

Thus $f(x_2|_{x_2>a}) = \begin{cases} 0 & ; \quad x_2 < a \\ \frac{1}{A} g_2(x) & ; \quad x_2 > a \end{cases}$

(since x_2 is the selected score of A the time)

The expected value of a variable x from a distribution f(x) is given by:

$$E[x] = \int_{-\infty}^{\infty} x \cdot f(x) \cdot dx \qquad \text{(Hays, P. 171)}$$

$$\text{Thus} \quad E[x_2|_{x_2>x_1}] = \int_{a}^{\infty} x \cdot \frac{1}{A} g_2(x)\, dx \qquad \ldots(3)$$

Combining equations 1, 2 and 3 then gives:

$$E[\max(a, x_2)] = a\left(1-\int_a^{\infty} g_2(x)\, dx\right) + \int_a^{\infty} x \cdot g_2(x)\, dx \,. \qquad \ldots(4)$$

This expected value is, of course, a function of the particular value "a". We need to average over all values of "a", each one weighted by its probability of occurrence, $g_1(a)$.

Replacing the particular value "a" by the general value x_1 then gives:

$$E[\max(x_1,x_2)] = \int_{-\infty}^{\infty} E[\max(a, x_2)] \cdot g_1(a) \cdot da$$

$$= \int_{-\infty}^{\infty} \left\{a\left(1-\int_a^{\infty} g_2(x)\, dx\right) + \int_a^{\infty} x \cdot g_2(x)dx)\right\} \cdot g_1(a)\, da \qquad \ldots.(5)$$

For the case of x_1 and x_2 drawn from discrete distributions with the average probabilities P_1 and P_2, and each based on N trials, equation (5) can be rewritten as:

$$E[\max(x_1,x_2)] = \sum_{j=0}^{N} \left\{ \frac{j}{N} \left(1-\sum_{k=j}^{N} g_2(k) \right. + \sum_{k=j}^{N} \frac{k}{N} \cdot g_2(k) \right\} g_1(j)$$

Tables II through V are based on this equation when applied to the binomial distribution.