On the minimum audible angle—A decision theory approach

William Morris Hartmann
Department of Physics, Michigan State University, East Lansing, Michigan 48824

Brad Rakerd
Department of Audiology and Speech Sciences, Michigan State University, East Lansing, Michigan 48824

(Received 15 March 1988; accepted for publication 13 January 1989)

The minimum audible angle (MAA) technique is a well-known psychoacoustical paradigm often used in the study of localization of sound. A difficulty with this paradigm, however, is that, in terms of decision theory, it is subject to two quite different interpretations. Although it is normally regarded as involving a discrimination task, the present work suggests that it is more likely to be an absolute identification task. Because of this difference in interpretation, it appears that previous work has overestimated the ability of listeners to localize sources of sound.

PACS numbers: 43.66.Qp, 43.66.Ba, 43.66.Pn [WAY]

INTRODUCTION

Thirty years ago, Mills (1958) developed a paradigm, called the minimum audible angle (MAA), designed to measure the accuracy with which human listeners can localize a source of sound. The method has since been used in studies of localization, both in the horizontal (azimuthal) plane (Perrott, 1969; Harris and Sargent, 1970) and in the vertical plane (Wettschureck, 1973). It has been applied to infants (Morrongiello, 1987), to the hearing impaired (Hausler *et al.*, 1983), and to underwater listeners, both humans (Feinstein, 1973) and dolphins (Renaud and Popper, 1975). The method has been extended to headphone listening to measure lateralization thresholds (Tobias and Zerlin, 1959; Mills, 1960).

The paradigm involves three source locations, as shown in Fig. 1(a). The center location is the standard (S), and it is placed at some angle Θ_0 with respect to the listener's forward direction. The second location is displaced from the standard by $\Delta\Theta = -A$, and the third is displaced from the standard by $\Delta\Theta = +A$ $(A > 0)$. For a study of localization in the azimuthal plane, the second and third source locations are equally to the left (L) and to the right (R) of the standard. Below, we shall assume an azimuthal geometry; our arguments are, however, quite general.

In an MAA trial, the listener hears two tones. The first is always from the standard. The second may come from the left or from the right. The listener's task is to say whether the second tone came from the left or the right of the standard. The method is one of constant stimuli. The parameters Θ_0 and A are held fixed for the duration of an experimental session, and the basic datum is the percentage of correct responses made by the listener for the given parameters. That value of A which corresponds to 75% correct responses can be taken to be the minimum audible angle.[1] For azimuthal localization of low-frequency tones (e.g., 500 Hz) near the forward direction $(\Theta_0 = 0)$, the minimum audible angle is found to be about 1 deg. This value is probably the most widely quoted number of any number in the localization literature; $\text{MAA}(\Theta_0 = 0) = 1$ deg.

The purpose of the present paper is to register a complaint against the minimum audible angle method. There is, perhaps, nothing wrong with the method within its own context. It is a well-defined procedure, and it leads to reproducible results. It can be used legitimately to compare localization accuracies in different conditions, for example, for different positions of the standard or for tones of different frequency.

The difficulty with the minimum audible angle method is that its interpretation, in terms of statistical decision theory, is ambiguous. Therefore, if one wants to relate the MAA to a hypothetical spread on the listener's internal decision axis, or if one wants to relate the MAA to other measurements of localization accuracy, for example, source-identification performance (Stevens and Newman, 1936; Oldfield and Parker, 1984; Butler, 1986), then there is no single clear way to proceed.

The difficulty is easily described. Imagine an MAA experiment set up for threshold performance. The left and right source positions are each separated by 1 deg from the standard. The experimenter supposes that he is going to find out whether a listener is sensitive to the 1 deg of separation

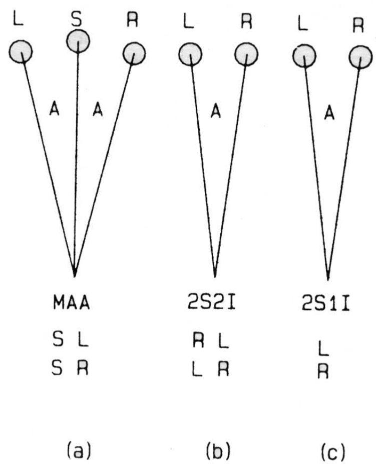


FIG. 1. Three experiments designed to measure localization accuracy: (a) the minimum audible angle, (b) two source two interval, and (c) two source one interval. The listener is at the vertex, and the sources L, S, and R are mutually separated by angle A. Below is shown the two possible stimulus sequences for each experiment, so called left trials and right trials.

 0001-4966/89/052031-11$00.80

between the standard and the other two sources. However, after some experience with the stimuli, the listener may learn to identify the left and right source positions absolutely, and these, of course, are separated by 2 deg. The listener does not have to attend to the tone presented from the standard. The listener can choose to pay attention only to the second tone and then decide whether it came from the left or right position. Indeed, as will be shown in Sec. I, the predictions of a plausible decision theory are that it is to the listener's advantage to choose just such a strategy. What, then, is the smallest detectable angular separation? Is it 1 deg or is it 2 deg? This difficulty and our resolution of it are the topics of the rest of this paper.

I. A DECISION THEORY FOR THREE EXPERIMENTS

The power of decision theory is that it is capable of unifying the results of different kinds of experiments in terms of a few basic parameters. If the parameters are determined from one experiment, then the theory predicts the results of a second. The present section proposes a decision theory to unify the results of different localization experiments. Its formal ancestor is the theory of intensity perception by Durlach and Braida (1969), but the application to localization includes some unique features.

The decision theory model postulates a decision axis, with coordinate values x, as an internal representation of location. Stimuli are mapped onto the decision axis, but not with perfect reproducibility. Because of internal noise, a stimulus with a given location Θ produces a distribution along x. The distribution is Gaussian with mean μ and variance σ^2. The mean increases monotonically with physical angle Θ. Along the decision axis there are also criterion points used by the listener to make decisions.

What is unusual about sound localization is that the listener has access to a precise reference for a decision axis and for the criterion points along it. This reference is established by the visual system, with a localization acuity several orders of magnitude better than the auditory system. As a result, it is reasonable to associate the listener's decision axis with the physical scale of location. This was the procedure of the MIT group (Searle *et al.*, 1976; Shelton and Searle, 1978). Our approach (Hartmann, 1983a) is slightly different in that it allows for the possibility of sensory bias along the decision axis. The mean of the auditory localization distribution corresponding to a given source is shifted from the visual referent by the "sensory bias" (see Appendix A).

We specifically consider three kinds of experiments, shown in Fig. 1. Figure 1(a) shows the MAA experiment, as described above. As noted at the bottom of the figure, a "left" MAA trial consists of the sequence standard–left. A "right" trial is the sequence standard–right.

Figure 1(b) shows an alternative discrimination experiment, "two sources two intervals" (2S2I). This experiment is similar to the MAA experiment, but there is no standard. The sources, L and R, are separated by angle A. The listener hears two tones, the sequence is either right–left (a left trial) or left–right (a right trial). The listener's task is to say which.

Figure 1(c) shows a simple identification experiment,

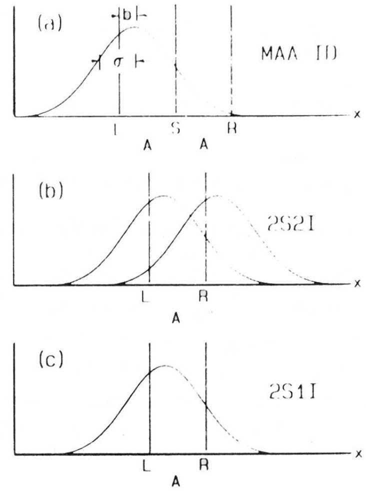


FIG. 2. Probability distributions of the internal representation of location for the three experiments. Referents for the source locations L, S, and R are shown by vertical lines. Figures (a)–(c) correspond to experiments in Fig. 1(a)–(c), respectively.

"two sources one interval" (2S1I) (Banks and Green, 1973; Balogh *et al.*, 1982). The source positions, L and R, are again separated by angle A. The listener hears only a single tone and has to say whether it came from the left or the right source.

We discuss decision theory models for these three experiments in reverse order. Figures 2(a)–(c) illustrate the operations of the models; they correspond to the three experiments of Figs. 1(a)–(c), respectively.

For this discussion, we need the cumulative normal function $C(X/\sigma)$, the integral of a normalized Gaussian, viz.,

$$C\left(\frac{X}{\sigma}\right) = \int_{-\infty}^{X} dx \frac{1}{\sigma\sqrt{2\pi}} e^{-x^2/(2\sigma^2)} \quad (1)$$

$$= \tfrac{1}{2} + \tfrac{1}{2}\,\mathrm{erf}[X/(\sigma\sqrt{2})], \quad (2)$$

where erf is the error function; e.g., $C(-\infty) = 0$. $C(0) = 0.5$, $C(1) = 0.84$, and $C(\infty) = 1$.

A. The 2S1I experiment

Figure 2(c) shows the internal distribution for the 2S1I experiment when the signal comes from the left source. Shown, too, are the referents that correspond to left and right source positions. The peak of the distribution, corresponding to the most probable value of x, does not occur at the position of the left source, but is shifted by the sensory bias b, defined as positive to the right. Because left and right sources are close together in a difference limen experiment, we assume that the sensory bias is the same for both. Finally, to maximize performance, within the constraint imposed by the bias, the listener chooses the left source if the value of x is closer to the left referent than to the right.

Using P_{cL} to represent the fraction of correct responses when the left source is presented and P_{cR} to represent the fraction of correct responses when the right source is presented, we have

$$P_{cL} = C[(A/2 - b)/\sigma] \quad (3a)$$

and

$$P_{cR} = C[(A/2 + b)/\sigma], \quad (3b)$$

where sensory bias b is measured in degrees, and σ is the standard deviation of the distribution, also measured in degrees.

B. The 2S2I experiment

Figure 2(b) shows the distributions for the values of x for the two tones presented in the 2S2I experiment. The listener's optimum strategy here is to treat this task as a discrimination experiment and to take the difference between the two values of x to form a decision variable. The decision variable is then normally distributed with a standard deviation of $\sigma\sqrt{2}$. Because the sensory bias is constant, the result of the differencing operation is to remove this bias from the calculation. Therefore, the long-term averages of P_{cL} and P_{cR} should be identical. One can expect, however, that an experiment will find that the two averages are not exactly identical, and we cope with this by introducing a "response bias" β, which is positive if there is a tendency to say left–right more often than right–left (see Appendix A). The equations for P_{cL} and P_{cR} are then given by

$$P_{cL} = C[(A - \beta)/(\sigma\sqrt{2})] \quad (4a)$$

and

$$P_{cR} = C[(A + \beta)/(\sigma\sqrt{2})]. \quad (4b)$$

C. The MAA experiment

The MAA experiment can be considered in two ways. First, it can be modeled as a discrimination experiment (MAA-DIS), which is the way that Mills (1958) thought about it. If, for instance, the stimulus sequence is "standard–left," then the figure for the distributions is identical to Fig. 2(b), with the symbol S replacing the symbol R. The fraction of correct responses to a standard–left sequence is given by P_{cL} in Eq. (4a). Similarly, the equation for P_{cR}, corresponding to the standard–right sequence, is Eq. (4b). Thus the equations for MAA-DIS are just the same as for the 2S2I task.

Alternatively, the MAA experiment can be modeled as an identification experiment (MAA-ID), which is the novel interpretation that motivates our work. The listener attends only to the second tone, and the distribution of x for that tone is shown in Fig. 2(a), for the case of a standard–left stimulus sequence. The referents for left and right sources are separated by $2A$. The listener makes a correct decision for this sequence if the value of x lies to the left of center. Therefore, percents correct are given by Eqs. (3) with A replaced by $2A$; i.e.,

$$P_{cL} = C[(A - b)/\sigma] \quad (5a)$$

and

$$P_{cR} = C[(A + b)/\sigma]. \quad (5b)$$

To illustrate the predictions of the decision theory, we consider a special case where there is no sensory bias ($b = 0$)

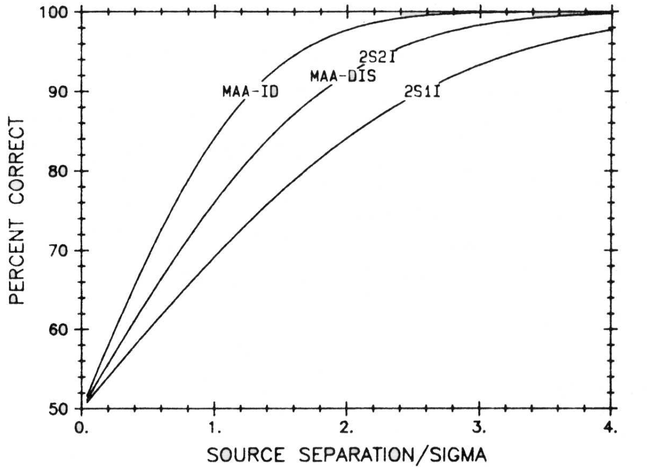


FIG. 3. Four psychometric functions predicted by the decision theory models of Sec. I. The functions for 2S2I and for MAA-DIS models are identical.

and no response bias ($\beta = 0$). From the formulas above, we calculate psychometric functions for the four models and plot them in Fig. 3. The vertical axis shows the percent correct responses, and the horizontal axis shows the separation of the sources, measured in units of the standard deviation of the internal distribution. As the source separation increases, the percent of correct responses increases.

Figure 3 shows that, for a given source separation A, the worst performance occurs in the 2S1I experiment; it is the hardest task. Next in order come the 2S2I experiment and the MAA experiment where the listener uses the discrimination strategy (MAA-DIS); both are described by the same curve. Finally, the best performance occurs in the case of the MAA experiment where the listener uses the identification strategy (MAA-ID). For equivalent performance, the MAA-ID model predicts that the difference limen measured in an MAA experiment will be better than that measured in a 2S2I discrimination experiment by a factor of the square root of 2. This observation, more than anything else, prompts the suspicion that the MAA experiment may not really be a discrimination experiment. It may be an absolute identification experiment instead, because the identification strategy should lead to better performance. But, to apply this strategy, the listener must ignore the standard tone of the MAA experiment, and it is not clear whether or not listeners do this. In order to find the answer, and in order to test the entire fabric of the decision theory models proposed above, we did some experiments.

II. THE EXPERIMENTS

We performed the three experiments described above, the minimum audible angle experiment (MAA), the two-source two-interval experiment (2S2I), and the two-source one-interval experiment (2S1I). The experiments were done in an anechoic room $11 \times 15 \times 8$ ft, IAC 107840.

A. Stimuli

To the extent possible, the stimuli were similar to those used by Mills (1958). The signals were 40-dBA, 500-Hz, sine tones with rise and decay durations of 70 ms and a fullon duration of 930 ms. In MAA and 2S2I experiments, where there were two tones per trial, the interstimulus interval was 1 s. The principal difference between our experiments and those of Mills was that Mills used a single moveable loudspeaker as a source. Our experiments used three (or two) different fixed speakers, mutually separated by angle A. The speakers were 12 ft from the seated subject, directly in front ($\Theta_0 = 0$), and were at ear height. The speaker system used is called "the Hydra," and it is described in Appendix B.

B. Lighting

Each of the three experiments was done in two ways, in the light and in the dark. In the light, the listener could see the sources, which provided an external reference for an internal localization axis. In the dark, the listener could not see anything at all. (Mills, with a moveable source, did his experiments in the dark.) In sum, there were actually six experiments in our study.

C. Subjects

There were three subjects, all male with normal hearing. Subjects B and W were the two authors, with previous experience in localization experiments. Subject K was an undergraduate volunteer with no previous experience. He was unaware of the hypotheses under test.

D. Procedure

An experimental run, during which all parameters were constant, consisted of 50 trials. Half the trials were left trials, half were right trials, defined per Fig. 1. The order of trials was randomly chosen by the computer that ran the experiment and collected the response data. To indicate his response, left or right, the listener pressed a button on a response box. The experiment was self-paced; a trial began 0.6 s after the listener had made a response to the previous trial. There was no feedback. After the 50 trials of a run, the listener was free to come out of the anechoic room and to look at the data.

The above procedure was found to be adequate for all of the six experiments except for one, namely, for 2S1I in the dark. Here, we found it necessary to give the subjects an orientation so that they were not completely lost on the initial trials. Therefore, the run began with a series of ten tones, alternating left and right; then, experiment trials were run in the normal way. After 25 trials, the listener heard another such series of orientation tones; the listener then completed the run with 25 more trials.

One goal of the experiments was to obtain psychometric functions to compare with the theoretical predictions in Fig. 3. Therefore, we used four different values of the intersource separation A: 3, 2, 1.5, and 1 deg. For each value of A and for each of the six experiments, each listener did four runs (200 trials).

The order of the experiments was not entirely random. A listener did all of the runs at an angular separation of 3 deg before doing the runs at $A = 2$ deg, etc., with the final runs done at 1 deg. Apart from a strict ordering on A, there was no particular rule for ordering the experiments except that we tended to do two runs of a given experiment consecutively, but never more than two runs.

After a listener had completed the requisite number of runs at a particular value of A, we examined the data for learning effects. If a learning effect appeared in the four runs for a given experiment, the listener did more runs of that experiment until his performance stabilized. The final data accepted for any experiment were those of the last four runs of that experiment.

The advantage of the procedure with ordered values of A is that it gave us the best chance to minimize any learning effects on the comparison of one experiment type with another, because the different experiment types were all done contemporaneously for a given A. The disadvantage is that any residual learning effect tends to flatten the psychometric function. Because our primary goal is to compare the different experiment types, the choice of procedure was an obvious one.

III. RESULTS

A. Psychometric functions

An overall measure of performance can be obtained by averaging P_{cL} and P_{cR}, respectively, the values of percent correct given left and right trials. This average psychometric function is inconsistent with the model equations above in that the width of the function does not exactly estimate the parameter σ, because the bias is treated only approximately. In practice, this inconsistency leads to only a small error. Figure 4 shows the psychometric functions, plotting average percent correct versus A, where A is the angular separation of the sources. Figure 4 can be compared with the predictions of the decision theory shown in Fig. 3. The theory predicts that performance in 2S2I should be better than the performance in 2S1I; therefore, in Fig. 4, the symbols 2 should always lie above the symbols 1. This is observed in 23 out of 24 possible comparisons. Thus the theory and the data are in qualitative agreement.

The models predict that, if the listener uses the *discrimination strategy* for the MAA experiment (MAA-DIS), then performance in MAA should be identical to that in 2S2I. Therefore, in Fig. 4, symbol M should coincide with symbol 2. This, however, is observed only rarely. The models predict that, if the listener uses the *identification strategy* for the MAA experiment (MAA-ID), then performance in MAA should be better than in 2S2I; i.e., symbol M should lie above symbol 2. This is observed in 22 of 24 possible comparisons, and in no case does the reverse ordering occur. Where equivalent performances can be found for 2S2I and MAA experiments, a factor of the square root of 2, as predicted by the theory, does not seem an unreasonable estimate for the ratio of the angular separations.

The reliability of the above comparisons depends upon the errors in the experimental psychometric functions. These can be estimated by comparing the individual runs of 50 trials with their average. Averaged across all listeners, in both light and dark, the standard deviation is 5%. The expected standard deviation of a distribution of differences across experiments is then about 7%. The percent correct for

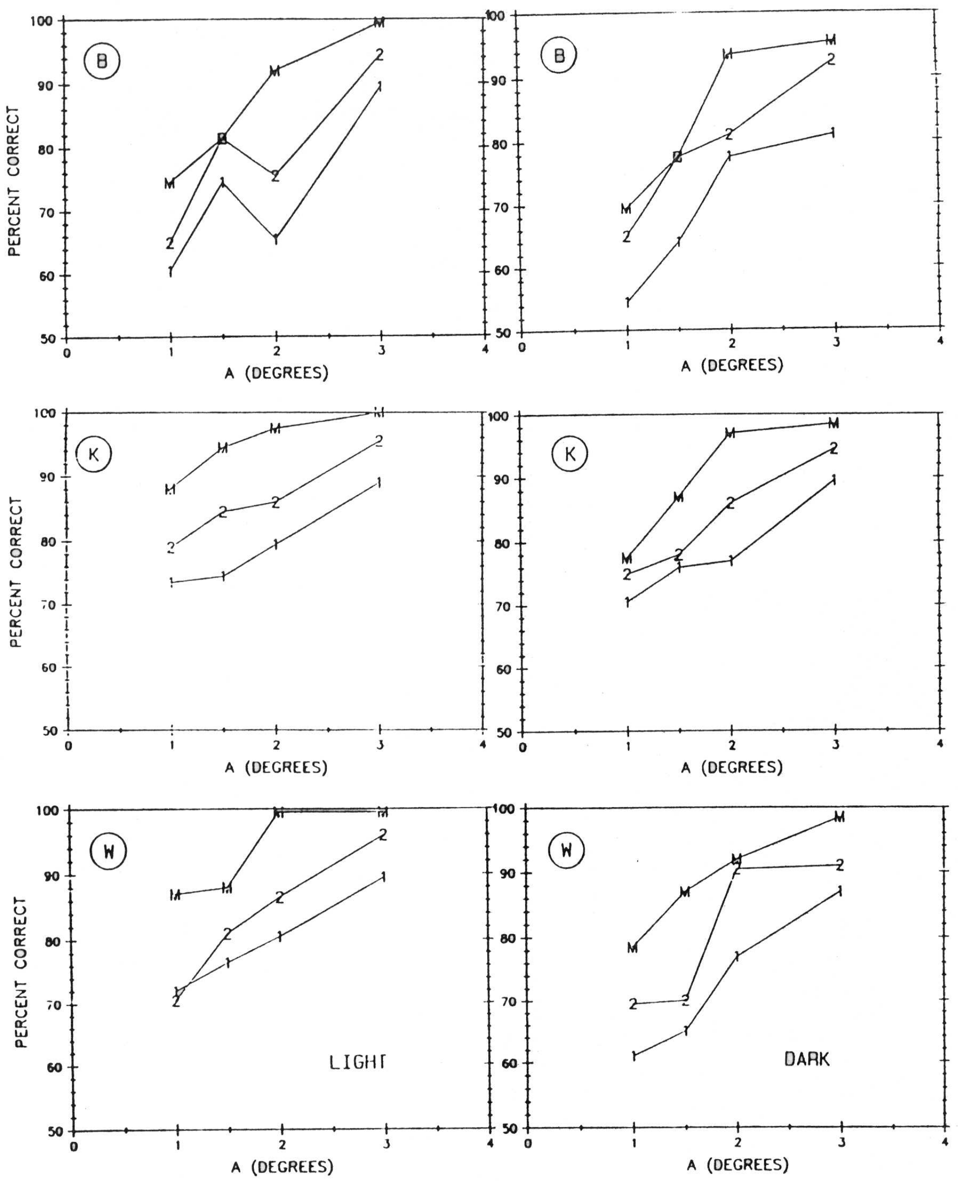


FIG. 4. Psychometric functions for listeners B, K, and W showing the percent correct responses given source separation A. Left panels show the results of experiments in the light; right panels are from experiments in the dark. Each panel shows three experiments: symbol 1 is 2S1I, symbol 2 is 2S2I, and symbol M is MAA. The experimental procedure was optimized for a comparison across the experiments, while risking a possible flattening of the functions; see text.

MAA is greater than the percent correct for 2S2I by 7% or more in 14 out of 24 possible comparisons. Half of the remaining ten comparisons are for an angle of 3 deg, where performance is usually so good that large differences cannot occur. Therefore, the data are consistent with the identification model for the MAA experiment but not with the discrimination model.

If, as we conjecture, the minimum audible angle experiment is really an identification task, then performance on the MAA experiment should correlate with performance on the 2S1I experiment, a prototypical identification task. Specifically, performance in the MAA experiment with a source separation of A should be equal to performance in the 2S2I experiment with a source separation of $2A$. Our choice of angular separations gives us two opportunities to test this hypothesis: MAA at 1 deg vs 2S1I at 2 deg, and MAA at 1.5 deg vs 2S1I at 3 deg. Table I shows these comparisons for the overall percent correct. The estimates of error given in parentheses are the standard deviations over the four runs. The hypothesis seems to be supported by the data: (1) The discrepancies in percent correct between the two experiments are within the error bars in 9 of a possible 12 comparisons; (2) the discrepancies are positive as often as they are negative; and (3) a statistical comparison (paired t test) shows no significant difference in the outcomes of the two experiments ($t = 0.40$, $p > 0.69$). This comparison between MAA and 2S1I experiments is further evidence that the MAA experiment may be an identification task.

B. Internal distributions

The most important thing about an internal distribution is its width. If the distribution is normal, per Eqs. (3)–(5), then the width is characterized by standard deviation σ. For each of the decision theory models there are two equations with two unknowns. Given P_{cL} and P_{cR} from experiment, the equations can be inverted to find the width σ and the bias.

TABLE I. Test of the hypothesis $P_{c,MAA}(A) = P_{c,2S1I}(2A)$. Entries are overall percent correct responses for comparable MAA and 2S1I experiments. If the MAA experiment is an identification task, then the two numbers in each row should be the same. The percent correct is an average for four runs for listeners B, K, and W in two lighting conditions. Numbers in parentheses are standard deviations ($N-1=3$ weight).

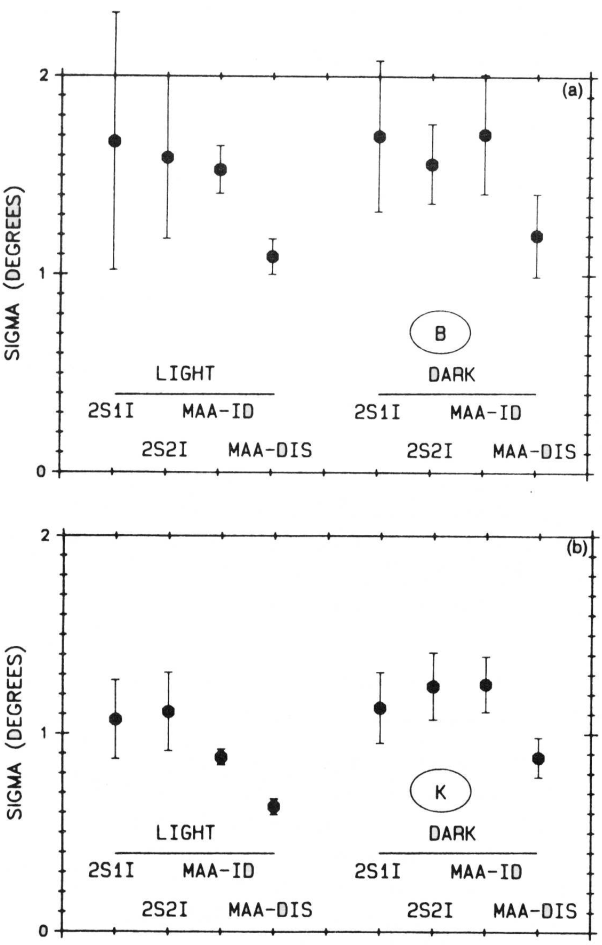
	MAA ($A=1$)	2S1I ($A=2$)	MAA ($A=1.5$)	2S1I ($A=3$)
B (light)	74.5 (3.8)	65.5 (3.3)	81.5 (3.8)	89.5 (0.9)
B (dark)	69.5 (5.7)	77.5 (3.8)	77.5 (4.6)	81.0 (12.2)
K (light)	88.0 (5.8)	79.5 (9.3)	94.5 (2.6)	89.0 (5.2)
K (dark)	77.5 (3.0)	77.0 (8.3)	87.0 (1.7)	89.5 (4.3)
W (light)	87.0 (7.0)	80.5 (0.9)	88.0 (7.9)	89.5 (0.9)
W (dark)	78.5 (3.0)	77.0 (3.3)	87.0 (4.1)	87.0 (3.3)

Unlike the calculation in Sec. III A, this treatment of the data is entirely consistent with the model equations.

If the models are correct and complete, then they should unify the data. The values of σ obtained from the inversion procedure should be the same for all the different angular separations and for all the experiments. Of course, the MAA-DIS and MAA-ID models cannot both be correct.

The values of the width obtained by inverting the equations and using our data as input do not show a systematic dependence on A, and, from the work of Searle *et al.* (1976), none would be expected for the small range in these experiments (maximum 6 deg). Therefore, the widths presented here are averaged over the four A values in a given experiment. The results, for the three listeners, are shown in Fig. 5. The error bars have an overall length equal to twice the standard deviation, as computed over the values of A ($N-1=3$ weight). These error bars show the standard deviation of the estimate of the standard deviation of the internal distribution.

Figure 5 shows that, for the six experiments, there is usually some region of overlap between the widths computed from models 2S1I and 2S2I and MAA-ID. The figure also shows that there is very little overlap with the widths computed from the MAA-DIS model. The latter are always the



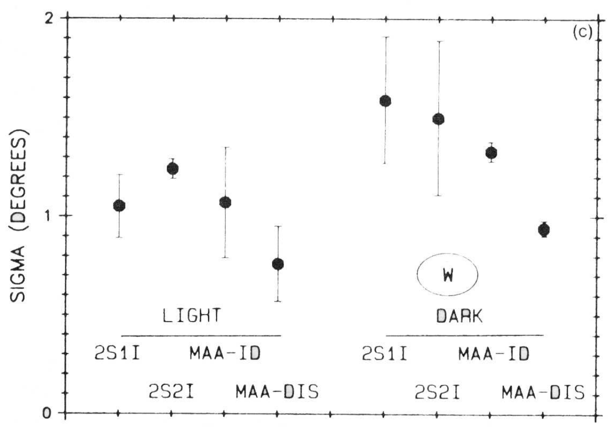


FIG. 5. Values of the width of the internal distribution, sigma, for three listeners. Per Sec. I, the 2S1I and 2S2I experiments are analyzed as identification and discrimination, respectively; the MAA experiment is analyzed in both ways. Separate analyses are done for experiments in the light and in the dark. The length of the error bars is twice the standard deviation computed over four different angular separations.

lowest points on the graph for any experiment. Therefore, the conclusion from the calculated widths in Fig. 5 agrees with the conclusion from the psychometric functions in Fig. 4: Modeling the minimum audible angle experiment as an absolute identification task is consistent with other experiments, namely, with 2S1I and 2S2I; modeling the minimum audible angle as a discrimination task is inconsistent.

The figure shows that the error bars tend to be smallest for the MAA experiments. This does not indicate any superiority of the MAA experiment; instead, it is an inevitable result of the calculations. As the percentage of correct responses becomes large, the values of the width determined by inverting Eqs. (3)–(5) become increasingly insensitive to the value of percent correct. All six experiments were done at the same values of A, and in each case the highest percent correct occurred for the MAA experiment. It is not surprising then that, when an average is computed over angles A, the insensitivity should appear as smaller error bars for the MAA experiment.

Our best estimate for the width of the internal distribution is found by averaging the widths obtained from the 2S1I, 2S2I, and MAA-ID model analyses of the data. The results are shown in Table II. These can be compared with the minimum audible angle, defined by the MAA-DIS analysis, also shown in Table II. The table shows that the minimum audible angle method overestimates the listener's accuracy by about a factor of 1.5. This factor is in good agreement with the factor predicted by the decision theory, a factor of the square root of 2.

IV. CONCLUSIONS

The central idea of this work is that measurements of auditory localization acuity obtained by different experimental methods can be unified. It is proposed that decision theory provides such a unifying schema. The decision theory represented by Eqs. (3)–(5) recognizes two components to localization error, a width of the internal distribution and a bias (see Appendix A).

The decision theory makes the prediction that the minimum audible angle paradigm of Mills (1958) is optimally an absolute identification task and not a discrimination task. This is contrary to the original intent of the paradigm and contrary to the usual interpretation of it.

We performed minimum audible angle experiments as well as two other kinds of experiments, two-source one-interval (2S1I) and two-source two-interval (2S2I), for comparison. The decision theory model is unequivocal about the latter two experiments: 2S1I is an identification task; 2S2I is optimally a discrimination task. All experiments were done both in the light and in the dark for a total of six experiments.

The experimental results were treated in two ways. As a first approximation, we plotted psychometric functions for the overall percent correct. These functions were consistent with the conjecture that the MAA experiment is an identification task and inconsistent with the idea that the MAA experiment is a discrimination task. Second, we did an exact analysis, inverting the equations of the decision theory models to find the widths of the internal distributions along the sensory axis. Again, the results favored the interpretation that the MAA experiment is an identification task.

The significance of our results for localization acuity itself is this: Experimentally, it is found that the MAA for 500-Hz tones in the forward direction is about 1 deg. For instance, the average MAA in our experiments (Table II) is 0.9 deg. One interpretation of this result is that the human listener can discriminate between sources that are separated by 1 deg. This is a particularly compelling conclusion because, in the discrimination model for the MAA, the width of the internal distribution turns out to be almost exactly equal to the MAA itself, $\sigma = \text{MAA}/0.95$. This interpretation is, however, incorrect. The listener very likely treats the MAA experiment as an identification task, and, therefore, the observed value of the MAA is smaller than the actual limit of spatial resolution. Our work suggests that the MAA underestimates the width (overestimates the acuity or resolving power) by a factor of about 1.5. The observed factor of 1.5 compares well with the expected value from the decision theory, from Eqs. (4) and (5), a factor of the square root of 2.

Because the results of psychoacoustical experiments often show considerable variability, a factor of 1.5 may not appear to be of much importance. In the case of localization acuity, however, the situation may be different. Our work using a source-identification paradigm (Hartmann, 1983a,b; Rakerd and Hartmann, 1985, 1986), with a span of sources between 24 and 36 deg, consistently finds widths σ equal to about 2 deg. This can be compared with apparently equally persistent estimates from MAA experiments of 1 deg. We do not think that the discrepancy can be ignored. The present work shows that part of the discrepancy, a factor of 1.5, can be attributed to the incorrect interpretation of the MAA experiment. The remaining part of the discrepancy, a factor of 1.3, may possibly be attributed to the increase in width with increasing source span suggested by the MIT group. Figure 4 in Searle *et al.* (1976) suggests that a factor of 1.3 for spans of about 30 deg is reasonable.

A practical conclusion that can be drawn from the above is that, if one wants to measure the ability of listeners to discriminate between sound locations, then one might better use the method that we have called two-source two-interval instead of the MAA method. There is little question about how to interpret the 2S2I task in terms of decision theory, and the results are more likely to agree with the results of other experiments, such as one-interval identification experiments.

TABLE II. Columns σ show the widths of the internal distributions estimated by averaging over three experiments: 2S1I, 2S2I, and MAA (identification analysis). Columns MAA-DIS show the minimum audible angle found in the MAA experiment (discrimination analysis).

Listener	σ(deg) Light	σ(deg) Dark	MAA-DIS (deg) Light	MAA-DIS (deg) Dark
B	1.6	1.7	1.1	1.2
K	1.0	1.2	0.6	0.9
W	1.1	1.5	0.8	0.9

ACKNOWLEDGMENTS

We are grateful to Dr. Adrian Houtsma for helpful discussions. This work was supported by the NIH and the NIMH.

APPENDIX A: BIASES

The decision theory equations (3)–(5) include biases, a sensory bias b for the identification analysis, and a response bias β for the discrimination analysis. These two forms of bias enter the equations in similar ways; both have dimensions of degrees of arc.

1. Sensory bias

A bias appears in a source-identification experiment when a listener's average identification of the position of a given source does not agree with the physical position of that source. Systematic biases, common to all listeners, occur in source-identification experiments done in rooms. These can easily be as large as tens of degrees, and they can be understood from the waveforms present at the listener's two ears (Rakerd and Hartmann, 1985, 1986). The corresponding bias measured in anechoic conditions, however, cannot be so easily explained. Presumably, this bias represents an imperfect registration between the auditory localization system and the visual system. Therefore, we refer to such a bias as a sensory bias. The sensory bias is expected to be small compared to the bias that can be observed in rooms, and not systematic.

Because the sensory bias appears explicitly in the decision theory equations, the presence of a sensory bias in experimental data does not affect the values of the width of the internal distribution as computed from those data. Including the sensory bias leads to experimental values of the width that are smaller than they would be if calculated from equations that permit no bias effect. Formally, the sensory bias is equivalent to a shift in criterion points. This has been called response bias by Braida and Durlach (1972), but in this paper we use the term response bias for the effect described below.

2. Response bias

The response bias appears in a discrimination analysis. Unlike sensory bias, the response bias cannot be understood as a shift between auditory and visual systems because the decision variable in a discrimination experiment is assumed to be a difference between sensory variables for successive stimuli. Our 2S2I experiment will serve as an illustration of response bias. In the absence of response bias, the listener responds left (viz., right–left) whenever the (difference) decision variable is negative. In the case of a positive response bias β, however, the decision variable must be more negative than $-\beta$ in order for the listener to respond left.

There are two reasons for including response bias in the discrimination analysis. The first is that the MAA procedure introduced by Mills effectively allows for a response bias. The minimum audible angle of Mills is not changed if there is a horizontal shift of his psychometric function, the percent right judgments as a function of the angle between source and standard. Any comparison with the data of Mills should, therefore, include a response bias in the analysis. The second reason is that it is necessary for the discrimination analysis to have two parameters in order to make a fair width comparison with the identification analysis, which has two parameters.

3. Observed biases

The biases observed in our experiments, as calculated from the inversions of Eqs. (3)–(5), are not large. A bias of 0.46 deg was the largest ever found (subject K, 2S2I in the dark, $A = 1$ deg). We found no systematic dependence of the biases on source separation angle A, except that biases tended to become small when overall performance became very good.

Both sensory biases and response biases tended to be positive as often as they were negative. A measure of the size of the bias can be obtained by averaging the absolute values over subjects and over angles A. The average absolute biases were: for 2S1I a sensory bias of 0.17 deg in both the light and the dark, and for 2S2I a response bias of 0.23 deg in both the light and the dark. Contrary to what might have been expected, response biases were not generally smaller than sensory biases. For MAA, the sensory bias calculated in MAA-ID and the response bias calculated in MAA-DIS are numerically identical; we found biases of 0.16 deg in the light and 0.27 deg in the dark.

The experimental biases appeared to be mostly random. A possible systematic component is suggested by the following calculation: There were six experiments, three subjects, and four angular separations for a total of 72 comparisons. Of these, there were a total of 32 cases where each of the four runs exhibited a nonzero bias. Of these 32, there were 13 cases in which the bias had the same sign on all four runs (7 of 20 identification cases, 6 of 12 discrimination cases). If the biases were entirely random, one would expect that four biases have the same sign one-eighth of the time, namely, in only 4 cases out of 32. It was not possible, however, to associate this apparent systematic character with any parameter of the experiments. In summary, the two kinds of bias are necessary ingredients in our calculations, but we cannot explain why they occur.

4. Implications of the bias

The MIT decision theory (Searle *et al.*, 1976; Shelton and Searle, 1978) does not include bias. Therefore, estimates of width (inverse acuity) based upon our model will be somewhat smaller than estimates based upon the MIT model. The data analyzed in the 1976 paper mostly came from identification experiments; therefore, the issue is one of sensory bias. Our experience with sensory bias is that it can be markedly increased by directing the listener's gaze away from the center of the source array (Hartmann, 1983b). As a result, we would expect that discrepancies between our model and the MIT model become important when the angular span of the sources is large. Then the listener cannot be facing all the sources at once. If widths are calculated from experimental data using a model that does not include bias, then one would expect the widths to increase significantly as the span of the sources increases. The MIT papers found that this increase does indeed occur; the effect is regarded as the localization analog of the "range effect" observed in intensity discrimination. We suggest that at least a part of that increase is due to increased sensory bias as the span of the sources increases.

APPENDIX B: THE HYDRA

The anechoic room used in these studies is small, and, therefore, the source positions in an MAA experiment must be closely spaced. With the sources 12 ft (3.66 m) from the listener, an angular separation of 1 deg corresponds to a source separation of only 2.5 in (6.4 cm). If three different loudspeakers are used for the sources, they must all be small, no more than 2.5 in. in diameter. But speakers this small are inefficient radiators for 500-Hz sine tones. When driven hard enough to make a 40-dBA tone at 12 ft, these small speakers produce detectable harmonic distortion. Further, the character of the distortion tends to be peculiar to each speaker, so that a listener can learn to identify a source by its characteristic tone color, an obvious experimental flaw.

The Hydra, hardware and protocol, solves the above problem by using three speakers for each source. Because three sources are needed for an MAA experiment, the Hydra has three elements. Each Hydra element is made from a length of standard 1.5-in. PVC drain pipe with a double wye (Genova double wye) on its far end. A double wye is a fitting with an output and three inputs. It is shaped like the "peace symbol," from the 1960s, with its three inputs separated by 45 deg and the output directly opposite the center input. The key to the Hydra design is that each input flange exactly fits the rim of a Radio Shack miniature speaker, nominally 2 in. (5.1 cm) in diameter, Archer number 40-245. A matching clean-out plug (Genova 71615) forms a tight seal at the rear of the speaker. The clean-out plug is drilled for electrical connections to the speaker. With the wye in a vertical orientation, each Hydra element has a width of 2.25 in. (5.7 cm).

The front end of the main tube is fitted with a series of three PVC pressure couplings of decreasing diameter to reduce the final inside diameter to 3/4 in. (1.9 cm). Therefore, at 12 ft, the aperture subtends 0.3 deg of arc. The largest of these couplings can slide within the main tube for fine tuning of the overall length. The main tube has a length of 20.5 in. (52 cm), which puts the second resonance of the entire system at 500 Hz, the frequency of the desired signal. Tuning the system in this way considerably improves the ratio of signal to harmonic distortion. For two speakers, it eliminates measurable distortion altogether. Four speakers have measurable third harmonic; three have both third and fifth. The largest measured distortion product is 43 dB down from the signal at 500 Hz.

The final assembly consists of three Hydra elements, with their front ends open toward the listener, separated by angle A. There are, therefore, a total of nine speaker heads. In operation, the selection of speaker heads is randomized. When an experimental trial calls for a left source, for example, one of the three speakers in the left Hydra element, selected at random, is turned on. A bank of relays, outside the anechoic room, selects the speakers under computer control.

Because there are nine speakers, it is very difficult for a listener to learn to recognize a source by its tone color. An additional element in the protocol is that the drive level is randomized for each tone, over a range of ± 1 dB. This procedure further randomizes the ratio of distortion to signal, while not seriously disrupting the listener's concentration on source location.

APPENDIX C: ADDITIONAL EXPERIMENTS

The experiments reported in this appendix are identical, in principle, to those in the body of the paper. There were, again, three experiment types, MAA, 2S2I, and 2S1I, and all three were done both in the light and in the dark. The stimulus tones were also identical.

This series of experiments was done before the invention of the Hydra. Stimuli were presented via three 5-in. speakers mounted on a board with a 3/4-in. hole allowing the sound

TABLE CI. Percent correct responses for left and right trials for four listeners in three experiments with two angular separations in the light and in the dark.

Exp	Deg	B L	B R	J L	J R	C L	C R	M L	M R
Light									
MAA	2.3	92	100	99	99	98	98	98	100
MAA	1.9	90	100	98	100	95	99	87	96
2S2I	2.3	83	87	97	100	93	92	94	94
2S2I	1.9	84	69	94	94	92	97	89	89
2S1I	2.3	88	87	86	96	90	78	87	82
2S1I	1.9	74	81	83	86	78	86	74	82
Dark									
MAA	2.3	89	97	96	99	96	96	93	98
MAA	1.9	82	88	85	97	84	99	94	92
2S2I	2.3	84	83	84	87	89	94	83	85
2S2I	1.9	67	82	88	85	92	98	73	89
2S1I	2.3	70	77	89	89	76	76	75	80
2S1I	1.9	62	71	76	77	79	75	66	70

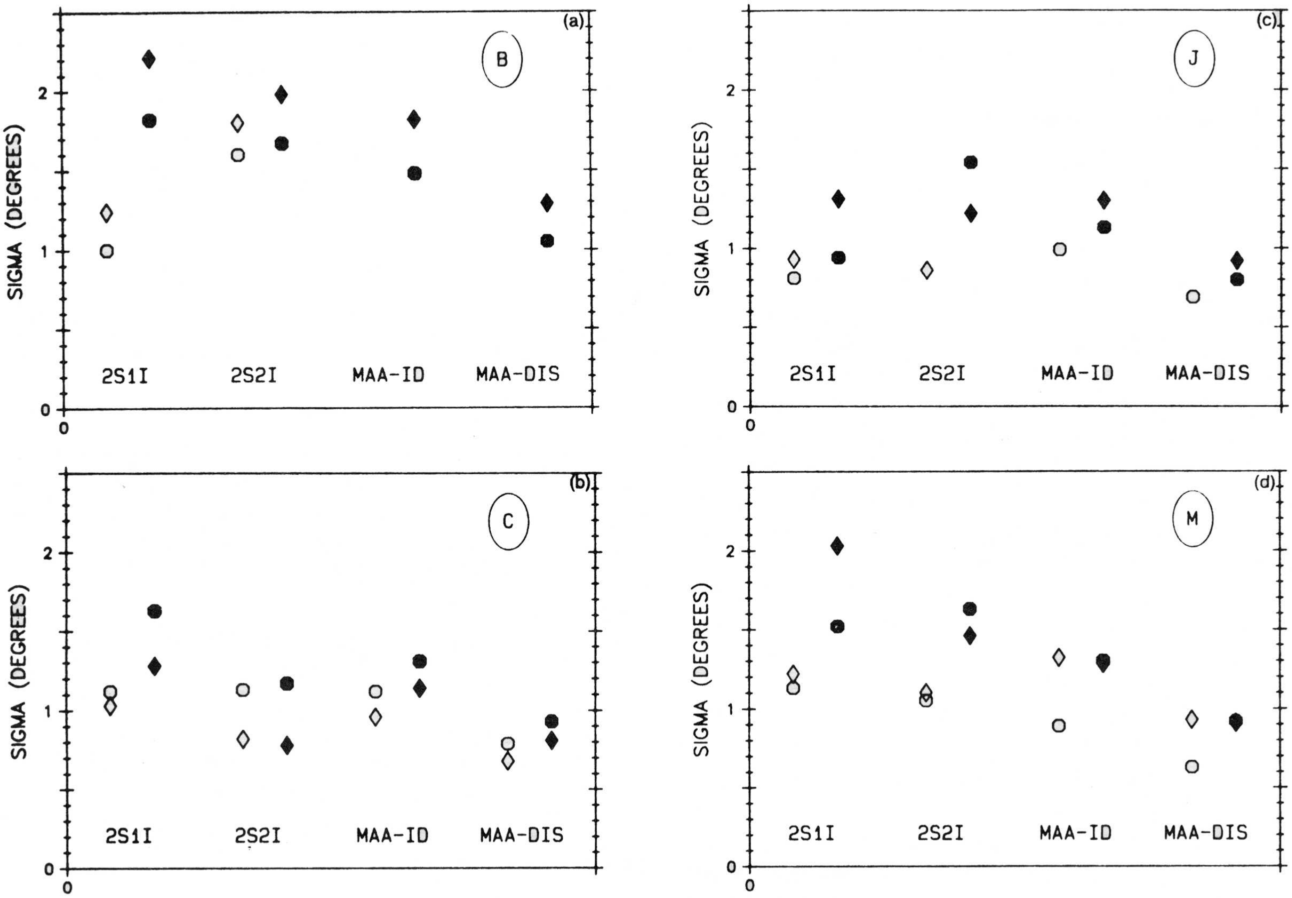


FIG. C1. Values of the width of the internal distribution, sigma, for experiments of Appendix C with four listeners. Open symbols are for experiments in the light; closed symbols are for experiments in the dark. Diamonds show results for $A = 1.9$ deg; octagons show results for $A = 2.3$ deg. As in Fig. 5, the computations for the MAA data are done using both the identification model and the discrimination model.

to pass. The speaker assemblies were placed as closely together as possible, and two different values of angle A, 2.3 and 1.9 deg, were obtained by seating the listener at 13 and 16 ft, respectively, from the speakers.

The results of the experiment are shown in Table CI, which gives percent correct responses given left trials and right trials. The table shows a high level of performance for several of the experiments, an expected result given the relatively wide separation of the sources. Because of the high performance, we had doubts about the usefulness of the data. As it turned out, however, the data tend to confirm the conclusions reached in the body of the paper. This can be seen in Fig. C1, which shows the widths of the internal distributions as computed from Eqs. (3)–(5). Because it is not possible to invert the equations when performance is 100% correct, some data points are missing from the figure.

Figure C1 shows widths that are comparable to those of Fig. 5. Thus the decision theory appears to be reasonably successful in unifying all the data. There is a tendency for the open symbols, corresponding to experiments in the light, to cluster, except for MAA-DIS, where the values of the width are the smallest. There is a similar tendency among the closed symbols, corresponding to experiments in the dark. This supports the conclusion that the identification model is a better interpretation of the MAA experiment than the discrimination model. The evidence here is not so strong as in the body of the paper, and that is probably because the calculated values of the width become less sensitive to the input data as the overall percent correct becomes large.

[1]The procedure actually used by Mills (1958) was to draw a psychometric function plotting the percent judgments right as a function of the angle of the source with respect to the standard. Half the distance between source angles for 25% and 75% right judgments was defined as the minimum audible angle. This procedure is equivalent to the introduction of a response bias, as noted in Appendix A.

Balogh, S. W., Schuck, J. R., and Leventhal, D. B. (**1982**). "A study of schizophrenics' ability to localize the source of a sound," J. Nerv. Ment. Dis. **169**, 484–487.

Banks, M. S., and Green, D. M. (**1973**). "Localization of high- and low-frequency transients," J. Acoust. Soc. Am. **53**, 1432–1433.

Braida, L. D., and Durlach, N. I. (**1972**). "Intensity perception II. Resolution in one-interval paradigms," J. Acoust. Soc. Am. **51**, 483–502.

Butler, R. A. (**1986**). "The bandwidth effect on monaural and binaural localization," Hear. Res. **21**, 67–73.

Durlach, N. I., and Braida, L. D. (**1969**). "Intensity perception I. Preliminary theory of intensity resolution," J. Acoust. Soc. Am. **46**, 372–383.

Feinstein, S. H. (**1973**). "Minimum audible angle underwater: A replication under different acoustic and environmental conditions," J. Acoust. Soc. Am. **54**, 879–881.

Harris, J. D., and Sargeant, R. L. (**1970**). "Sensory behavior of naval personnel: Monaural/binaural minimum audible angle of auditory response," U.S. Naval Submarine Medical Center, Groton, CT, Rep. 607.

Hartmann, W. M. (**1983a**). "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391.

Hartmann, W. M. (**1983b**). "Localization of sound in rooms–The effect of visual fixation," Proc. 11th ICA, Rev. d'acoustique **3**, 139–142.

Hausler, R., Colburn, S., and Marr, E. (**1983**). "Sound localization in subjects with impaired hearing. Spatial-discrimination and interaural-discrimination tests," Acta-Otolaryngol. Suppl. 1–62.

Mills, A. W. (**1958**). "On the minimum audible angle," J. Acoust. Soc. Am. **30**, 237–246.

Mills, A. W. (**1960**). "Lateralization of high-frequency tones," J. Acoust. Soc. Am. **32**, 132–134.

Morrongiello, B. A. (**1987**). "Infants' localization of sounds in the median sagital plane: Effects of signal frequency," J. Acoust. Soc. Am. **82**, 900–905.

Oldfield, S. R., and Parker, S. (**1984**). "Acuity of sound localization: A topography of auditory space I—Normal hearing conditions," Perception **13**, 581–600.

Perrott, D. R. (**1969**). "Role of signal onset in sound localization," J. Acoust. Soc. Am. **45**, 436–445.

Rakerd, B., and Hartmann, W. M. (**1985**). "Localization of sound in rooms, II: The effects of a single reflecting surface," J. Acoust. Soc. Am. **78**, 524–533.

Rakerd, B., and Hartmann, W. M. (**1986**). "Localization of sound in rooms III: Onset and duration effects," J. Acoust. Soc. Am. **80**, 1695–1706.

Renaud, D. L., and Popper, A. N. (**1975**). "Sound localization by the bottlenose porpoise Tursiops truncatus," J. Exp. Biol. **63**, 569–585.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). "Model for auditory localization," J. Acoust. Soc. Am. **60**, 1164–1175.

Shelton, B. R., and Searle, C. L. (**1978**). "Two determinants of localization acuity in the horizontal plane," J. Acoust. Soc. Am. **64**, 689–691.

Stevens, S. S., and Newman, E. B. (**1936**). "The localization of actual sources of sound," Am. J. Psychol. **48**, 297–306.

Tobias, J. V., and Zerlin, S. (**1959**). "Lateralization threshold as a function of stimulus duration," J. Acoust. Soc. Am. **31**, 1591–1594.

Wettschureck, R. G. (**1973**). "Die absoluten Unterschiedsschwellen der Richtungswahrnehmung in der Medianebene beim natürlichen Hören sowie beim Hören über ein Kunstkopf-Übertragungssystem," Acustica **28**, 197–208.