Interaural level differences and the level-meter model

William M. Hartmann[a] and Zachary A. Constan
Department of Physics and Astronomy, Michigan State University, East Lansing, Michigan 48824

(Received 23 May 2001; revised 19 June 2002; accepted 20 June 2002)

The interaural level difference (ILD) plays a significant role in sound localization. However, the definition of ILD for noise is open to some interpretation because it is not obvious how to deal with the inevitable level fluctuations. In this article, the ILD is interpreted as an energylike (time-integrated) measure of stimulus level, independent of other stimulus details—particularly interaural correlation. This concept is called the ''level-meter model.'' The model was tested by measuring human ILD thresholds for noise stimuli that were interaurally correlated, or anticorrelated, or uncorrelated. An additional test (not involving lateralization) measured the threshold for level discrimination based on loudness. According to the level-meter model, all four thresholds should be the same. The experimental results showed that the predictions of the level-meter model held good to within about half a dB, although thresholds for level discrimination were systematically higher than ILDs. Among the ILDs themselves, thresholds were slightly higher for uncorrelated noise. The latter result could be explained by replacing the level-meter model with a loudness-meter model, incorporating temporal integration. The same model accounted for the bandwidth dependence of the threshold.
[DOI: 10.1121/1.1500759]

PACS numbers: 43.66.Pn, 43.66.Ba, 43.66.Fe [LRB]

I. INTRODUCTION

An important contribution to the human ability to localize a sound is the difference in level between the two ears, or interaural level difference (ILD). The object of the work described in this article was to gain insight into the way the auditory system uses ILD information in a broadband noise signal. Specifically, we wanted to know whether it matters if the noises in the two ears are mutually coherent or not. Although coherence is necessary for the use of interaural time differences, it is not obvious that coherence plays any role in the use of ILD information.

To bring the question into clear focus, we formed a null hypothesis about the way that the binaural system measures the ILD, namely the ''level-meter model.'' According to the level-meter model the binaural system measures the power in the left ear with one meter, measures the power in the right ear with a second meter, compares those two power measurements, and uses the difference to localize or lateralize. Binaural coherence is unimportant to the operation of this system because the system only measures the average levels from the two ears. This model would predict that the ability to detect small differences in interaural level should be independent of whether the noise is interaurally correlated or uncorrelated. Implicit in the level-meter model is an idealized integration time, long enough to average over stimulus fluctuations so that the precision of the level measurement is limited only by the noise duration.

The alternative to the level-meter is a model in which the binaural system does not reduce levels in left and right ears to single individual magnitudes prior to comparison. For instance, a system might tend to track the difference between left and right signals as a function of time. Such an ILD-tracking system is sensitive to the difference between correlated and uncorrelated noise because of inherent fluctuations in noise power. For perfectly correlated noise $(\text{cross correlation}=1)$, the signals in the two ears are identical except for a fixed level difference. The instantaneous ILD is constant even though the noise fluctuates. By contrast, if the noise signals in the two ears are perfectly uncorrelated $(\text{cross correlation}=0)$, then the fluctuations in left and right ears are independent, and the instantaneous ILD varies as well. The ILD fluctuation leads to additional variance, causing this ILD-tracking model to predict that the ability to detect small differences in interaural level should be better for correlated noise than for uncorrelated.

In their study of monaural versus binaural discrimination, Jesteadt and Wier (1977) made a similar distinction between models. Their *independent threshold* model, like the level meter model, is insensitive to interaural coherence. Their *information integrating* model, like the tracking model, is sensitive, possibly giving different results for interaurally correlated or uncorrelated noise.

Apart from models of measurement, there is the matter of perception. A perfectly coherent (diotic) noise is normally perceived as a compact image near the center of the head. Introducing a small ILD moves the image to the left or right but retains the compact character if the ILD is less than about 8 dB (Blauert, 1983). By contrast, a binaurally uncorrelated noise forms a fuzzy image that fills the head. *A priori*, it seems likely that listeners should be able to lateralize the compact image more successfully than the fuzzy image. Thus, both the ILD-tracking model and common experience suggest that ILD sensitivity should be greatest for correlated noise, contrary to the prediction of the level-meter model.

The role of binaural coherence in connection with ILD

---

[a] Electronic mail: hartmann@pa.msu.edu

sensitivity has been studied before. In unpublished work, Grantham and Ahlstrom (1982) measured ILD sensitivity for broadband noise and narrow-band noise centered on several frequencies. For two out of three listeners, ILD thresholds tended to be smaller for diotic-except-for-level (correlated) noise than for completely uncorrelated noise. The advantage observed for correlated noise was greater when noise bursts were brief, 30 ms, than when noise bursts were 500 ms in duration. Similarly, an abstract by Neutzel (1982) reported slightly smaller ILD thresholds for broadband correlated noise compared to uncorrelated noise. The experiments of the present article expand on these works, including additional conditions.

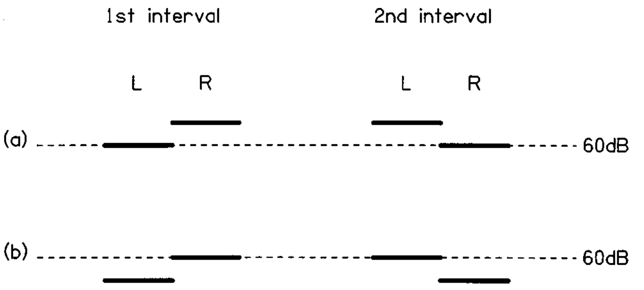

FIG. 1. There were two options (a) and (b) for the level variation on a *right–left* trial as shown. The two options were used randomly and equally often. The total power (left + right) was the same on both time intervals, but the ILDs were reversed from the first interval to the second. There were similar options for *left–right* trials.

II. EXPERIMENT 1—BROAD BAND

Experiment 1 measured the human ability to detect small interaural differences in the level of broadband noise under three conditions of noise coherence:

(1) Diotic-except-for-level: Noises were identical in the two ears except for a difference in level. Diotic noise is called *No* in the traditional notation of masking level difference (MLD) studies; the cross-correlation function for zero lag is 1.

(2) Anticorrelated: Noises were identical in the two ears except for a difference in level and an overall factor of -1 for one ear. Thus each spectral component in the left ear was 180 degrees out of phase with the corresponding component in the right ear. Anticorrelated noise is called $N\pi$ in traditional MLD notation; the cross correlation for zero lag is -1.

(3) Uncorrelated: Noise sources for left and right ears were independent. Uncorrelated noise is called *Nu* in traditional MLD notation; the ensemble-averaged cross correlation is 0 for all values of the lag.

The ability of listeners to detect ILDs was probed in a lateralization experiment using headphones. If lateralization, based on ILD information, is only a matter of comparing levels in the left and right ears, as conjectured in the level-meter model, then lateralization thresholds should be the same for all three noise conditions. Furthermore, the model suggests that lateralization cannot be more accurate than the initial independent level measurements in the two ears. As will be shown in Sec. IV, the model predicts that threshold ILDs should be equal to the difference limen in intensity, as also suggested years ago by von Békésy (1930). It is an indication of the power of the level-meter model that it unifies such disparate percepts as lateralization and intensity discrimination. To test this idea, we added a fourth condition,

(4) DLI (difference limen in intensity): Noises in left and right ears were always identical.

A. Method

1. Lateralization experiment

In the lateralization experiment, noise was presented on two successive intervals. The two intervals had ILDs of equal magnitude but opposite sign. The sign reversal was expected to cause the image on the second interval to be either to the right or to the left of the image on the first, corresponding to ''left–right'' or ''right–left'' trials, respectively. The listener was required to indicate the direction of the change in a two-alternative forced-choice task.

On every interval, one ear received the standard level of 60 dB SPL. As shown in Fig. 1, a right–left trial could be made either by making the right-ear level larger on the first interval or making the left-ear level smaller on the first interval. Both options were used with equal probability according to a random schedule. Whatever the ILD on the first interval, its sign was reversed to make the ILD for the second. Similarly, two corresponding sequences were randomly used for left–right trials.

The magnitude of the ILD was varied in a staircase fashion. The staircase was one-up three-down, targeting the 79% correct point on a psychometric function. For every incorrect response the interaural level difference (ILD) was increased by an increment. After three successive correct responses, the ILD was decreased by an increment. The magnitude of the increment was caused to vary according to the staircase value, with the goal of reaching threshold quickly and providing accuracy in the vicinity of threshold. When the experimental run began, the initial ILD was 2 dB, and the ILD increment was 0.5 dB. As the staircase progressed, the ILD increment depended on the ILD. When the ILD was greater than 1.5 dB the increment was 0.5 dB. When the ILD was between 1.0 and 1.5 dB, the increment was 0.2 dB. When the ILD was less than 1.0 dB, the increment was 0.1 dB. The minimum possible ILD was 0.1 dB.

A staircase run continued until the staircase had changed direction 14 times. The first four turning points were discarded and the average and standard deviation for the remaining ten turning levels became the data for the run. The duration of runs ranged from 2 to 4 minutes.

2. DLI experiment

The stimulus for the DLI experiment was a minor variation on the lateralization experiment. To make the DLI experiment, the stimulus for the right ear from the lateralization experiment was sent to both left and right ears. Therefore, the experiment was diotic. The listener's task was to say whether the first or second interval was the louder. Otherwise, the DLI experiment was identical to the lateralization experiments.

3. All experiments

The experiments of four types, (1) lateralization of correlated noise, (2) lateralization of anticorrelated noise, (3) lateralization of uncorrelated noise, and (4) DLI for diotic noise, were done in random order, except that no experiment type was done more than twice in succession. After a few training runs, listeners completed runs for which data were collected. Listeners received no feedback. From the experience of Grantham and Ahlstrom (1982), and from our own pilot experiments, we expected the distinctions among the four types of experiments to be subtle. Therefore, we insisted that the runs lead to rather tight staircases with small variance. Runs with a standard deviation among turning levels less than 0.3 dB were considered ''tight'' and were accepted. Runs with larger standard deviation were repeated. Final data were based on the last six tight runs for each listener on each of the experiment types, a total of 24 runs, or 144 runs for the six listeners.

B. Stimuli

All stimuli were white, Gaussian, broadband noises generated by Zener diodes biased near the breakdown knee. The spectrum was verified by a spectrum analyzer. The Gaussian character was checked by repeated sampling and plotting a histogram of the values. For the uncorrelated noise experiment independent noise generators were used. The noises were low-pass filtered at 10 kHz (−48 dB/oct). The spectrum level was 20 dB, making the stimulus level 60 dB SPL. Stimuli were turned on with matched voltage-controlled amplifiers, controlled by a common gating signal with a rise/fall time of 30 ms. Therefore, lateralization was not affected by interaural onset differences. The gate was triggered by a Tucker-Davis TG6 module to establish precise timing. After a 300-ms warning interval, marked by a green pilot lamp, there were the two stimulus intervals, 500 ms in duration, separated by a 500-ms silent gap. After the stimulus intervals, a red pilot lamp requested a response from the listener. There was no time limit for the response. The experiment was self-paced with the next trial beginning 300 ms after a response. The 500-ms stimulus duration placed a theoretical limit on our ability to test the level meter model.

Listeners were seated in a double-walled sound-attenuating room (Acoustic Systems model RE-244) and received the stimuli via Sennheiser HD 480 headphones. They made their responses by pressing one of two buttons on a response box.

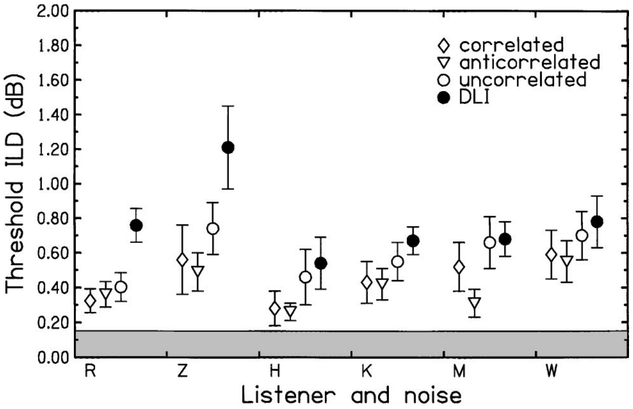

FIG. 2. Results of experiment 1: Threshold ILDs are shown for three noise types: Diamonds for binaurally correlated noise, triangles for anticorrelated noise, and open circles for uncorrelated noise. Filled circles show difference limens for the level of diotic noise (DLI). The experimental method could not measure thresholds below 0.15 dB.

C. Listeners

There were six listeners, R, Z, H, K, M, and W. Listeners R, Z, H, and M were males, ages 21, 29, 60, and 19. Listeners K and W were female, ages 21 and 19.

D. Results

For each listener and experiment type, the mean of the six runs and the standard deviation were found. These data appear in Fig. 2. Of particular interest was a comparison among the different experiment types. This comparison, averaged over listeners, is shown in the top line of Table I.

Figure 2 shows that all thresholds fell between 0.2 and 1.3 dB. Thus, they were all similar. However, some systematic differences are also visible. For every listener the largest threshold was the DLI. Also, among the ILDs (open symbols), every listener's largest threshold was for uncorrelated noise. The ILD threshold values for correlated and uncorrelated noise from Table I are 0.451 and 0.585 dB, respectively. These can be compared with values of 0.33 and 0.40 dB obtained by Neutzel (1982) using a four-interval procedure.

An analysis of variance was conducted on the four thresholds: lateralization of noise (correlated, anticorrelated, and uncorrelated) and the difference limen in intensity. The test indicated significant differences among the four, $F(3,15)=16.7$, $p<0.01$. *Posthoc* comparisons of the individual means found that the DLI was greater than all three threshold ILDs ($p_{\max}<0.03$). Further, the threshold ILD for anticorrelated noise was less than that for uncorrelated noise, $F(1,15)=10.42$, $p=0.03$. These paired comparisons, like all the others in this article, include a Bonferroni correction factor (Dunn, 1961) for multiple comparisons.

TABLE I. Threshold ILDs in dB from experiments 1–3, averaged across listeners for each stimulus type.

Experiment	Bandwidth (kHz)	Correlated	Anticorrelated	Uncorrelated	DLI
1	10	0.451	0.398	0.585	0.773
2	1	0.679	0.733	0.835	1.021
3	1 (rnd)	0.756	0.766	0.880	1.152

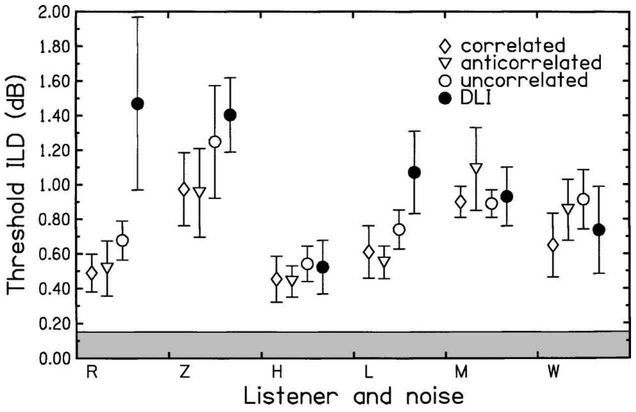

FIG. 3. Same as Fig. 2 but showing the results of experiment 2.

An analysis of variance on the three ILD thresholds alone indicated a significant difference among these noise types, overall, $F(2,10) = 16.67$, $p < 0.001$. *Posthoc* Bonferroni-protected comparisons of the individual means showed significant differences between the correlated and uncorrelated noises, $F(1,10) = 16.07$, $p = 0.007$, and between the anticorrelated and uncorrelated noises, $F(1,10) = 31.30$, $p < 0.001$.

III. EXPERIMENT 2

Experiment 1 found a weak effect of noise coherence on the ability to lateralize a noise. Because the noise was broadband, 0–10 000 Hz, it is not entirely clear how to interpret this result. The human binaural system is sensitive to interaural coherence in envelope fluctuations over a broad frequency range (Henning, 1974). However, the system is insensitive to coherence in the fine structure above about 1500 Hz and cannot use interaural time differences (ITD) in the fine structure to lateralize above this frequency. In Experiment 1, most of the noise power was outside this low-frequency range. By contrast, listeners' sensitivity to ILD is essentially independent of frequency (Yost, 1981).

The goal of experiment 2 was to provide a more severe test of the concept that ILD sensitivity is independent of noise type. The noise was low-pass filtered so that all the noise power was below 1 kHz. It was expected that interaural coherence would be more important in experiment 2 because both fine-structure ITD and envelope ITD contribute to lateralization and to perceived auditory source width across the entire frequency range.

A. Method

Experiment 2 was identical to experiment 1 in method except that the low-pass filter cutoff was decreased from 10 to 1 kHz, again -48 dB/oct. The noise level in the headphones remained 60 dB. In another change, listener K was replaced by listener L, a male age 20.

B. Results

The results of experiment 2 are shown in Fig. 3, comparable to Fig. 2 for experiment 1. Averages over listeners are given in Table I. The comparison shows that lateralization performance in experiment 2 declined compared to experiment 1. For the five listeners common to experiments 1 and 2, the average threshold across the four experiment types was 0.82 dB on experiment 2 compared to 0.55 dB on experiment 1. For every listener and every experiment type, threshold either increased or stayed about the same when the bandwidth was reduced from 10 to 1 kHz. Evidently all listeners had made good use of power above 1 kHz in experiment 1.

Experiment 2 was done with the intention of magnifying differences among the different experiment types. However, Fig. 3 shows that quite the reverse occurred. Unlike experiment 1, there is no stimulus type that leads to the greatest threshold for all listeners, and there is no universal agreement within the ILDs (open symbols) either.

An analysis of variance performed on the four experiment types showed that the difference among thresholds was marginally significant, $F(3,15) = 3.35$, $p = 0.045$. However, paired comparisons indicated no significant differences, $p_{\min} \geqslant 0.06$.

In an analysis of variance on the three ILD thresholds alone, the three noise types were found to be significantly different overall, $F(2,10) = 4.25$, $p < 0.05$. This was chiefly due to a marginally significant mean difference between the thresholds for correlated and uncorrelated noise, $F(1,10) = 8.22$, $p = 0.05$. The other individual comparisons did not approach significance. In the end, it was only barely possible to challenge the level-meter model on the basis of the results of experiment 2.

IV. EXPERIMENT 3

There is a remote possibility that the lateralization tasks in experiments 1 and 2 were not really binaural. Inspection of the protocol in Fig. 1 shows that it would theoretically have been possible for a listener to monitor a single ear and perform the task based on a loudness comparison among the two intervals. As will be seen in the discussion section that follows, this would have been a poor strategy because the experiments showed that level discrimination was less successful than lateralization. However, this strategy remains a theoretical possibility. The purpose of experiment 3 was to provide a check on experiment 2 by making it very unnatural to perform the task using a monaural strategy.

A. Method

Experiment 3 was identical to experiment 2 except that the standard level (always 60 dB in experiments 1 and 2) was randomized on every experimental interval. The distribution of levels was rectangular, centered on 60 dB, and $R = 5$ dB in width. According to Green (1988), a level randomization of R dB should lead to a threshold level of $R[1 - \sqrt{2(1-P_C)}]$, where P_C is the percentage of correct responses on a two-alternative task. For $P_C = 0.79$, the expected threshold for decisions based on level alone is 1.8 dB. The level randomization was applied to the lateralization tasks only. It could not logically be applied to the diotic level discrimination task, and thus the DLI task remained identical to experiment 2. Listeners were the same as in experiment 2.

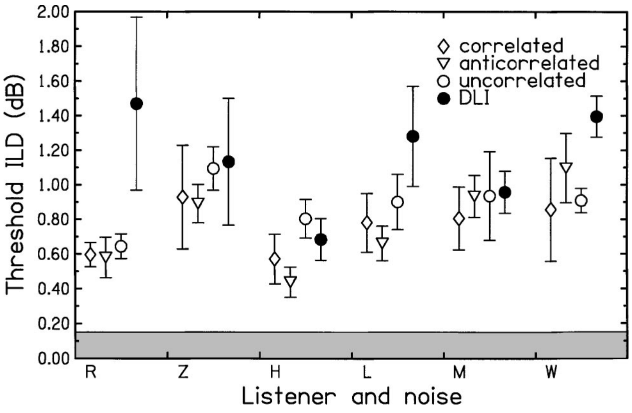

FIG. 4. Same as Fig. 2 but showing the results of experiment 3.

B. Results

The results of experiment 3 are shown in Fig. 4. They can be compared with Fig. 3 for experiment 2 (fixed standard). Averages over listeners appear in Table I. The largest value ever seen for any listener was 1.1 dB, notably less than 1.8 dB expected for judgments based on level in a single ear. We conclude that listeners did not make their decisions based on level alone, but used interaural level differences to do the task.

It was expected that experiment 3 would lead to the same results as experiment 2. The data suggest that the results were indeed similar. Of the 18 ILD thresholds (6 listeners × 3 noise conditions), 11 increased and 5 decreased when the standard was randomized. No change was outside the error bars. The only difficulty encountered in experiment 3 was that listener *R*, who was clearly having problems with the DLI task in experiment 2 (Fig. 3), consistently failed to satisfy our criterion for tight DLI staircases on the identical task in experiment 3. Therefore, his DLI data point for experiment 2 was used also for experiment 3.

Table I shows that the order of thresholds among experiment types in experiment 3 was the same as in experiment 2. An analysis of variance performed on the four experiment types showed that the difference among thresholds was significant, $F(3,15)=6.53$, $p<0.01$. Comparisons of paired differences indicated only two significant differences, namely DLI vs ILD for correlated noise, $p=0.01$, and DLI vs ILD for anticorrelated noise, $p=0.01$.

An analysis of variance performed on the three ILD thresholds showed that trends among the different noise types for experiment 3 were similar to those for experiment 2. Subjects were, however, somewhat more variable in exhibiting those trends. Accordingly, the overall analysis of variance on the noise results approached, but did not reach significance, $F(2,10)=2.66$, $p=0.12$.

To further compare experiments 2 and 3, the averages over listeners are given in Fig. 5. The plots are nearly parallel. The most likely explanation for the small consistent difference between the thresholds for experiments 2 and 3 is that the level randomization on experiment 3 slightly disrupted listener concentration.

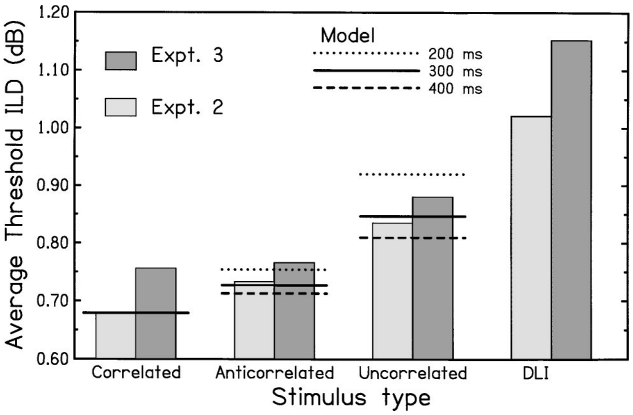

FIG. 5. Thresholds, averaged over the six listeners, for experiments 2 and 3. The vertical scale has been greatly expanded compared to Figs. 3 and 4 to show details better. Dotted lines show the predictions of the loudness-level meter model for integration times of 200 and 400 ms. The dashed line shows the prediction for 300 ms.

V. DISCUSSION

A. The level-meter model

The level-meter model predicts that the thresholds for the four tasks in the experiments should all be the same. According to this model the signals in left and right ears are not compared on a moment-to-moment basis; only an average representation of the signal levels is compared—average neural driven firing rate, for example. Because the level measurements by the two ears are independent in this model, it is immaterial whether the signals are correlated or not. Therefore, one expects the threshold ILDs to be the same for all three noise types. The level-meter model also predicts that the difference limen in intensity (DLI) should be the same as the threshold ILDs.

The argument for this equivalence is as follows: Suppose that a level measurement in a single ear has a standard deviation σ. The distribution of interaural level differences then has a standard deviation of $\sigma\sqrt{2}$. Therefore, if a power increment, Δ, is added to one ear, the ratio of internal ILD to standard deviation for a single interval is $\Delta/(\sigma\sqrt{2})$. Because the ILD, Δ, is added to one interval and subtracted from the other in the lateralization task, the means of the interval distributions differ by 2Δ, and the standard deviation of the difference distribution becomes 2σ. Therefore, a comparison between first and second intervals leads to a d' of $2\Delta/(2\sigma)$, or Δ/σ.

In the DLI experiment, the level difference between the two intervals is Δ, and the ratio of internal level difference to the standard deviation for a differential measurement is $\Delta/(\sigma\sqrt{2})$. However, this is the result for a single ear. In this experiment, both ears receive the same stimulus. As shown by Jesteadt and Wier (1977), d' for diotic DLIs is better than for monaural by the square root of 2. Therefore, the ratio of internal difference to standard deviation is Δ/σ, the same result as obtained for lateralization.

A brief argument that leads to the same conclusion is that for both ILD and DLI tasks, both ears are subjected to a level difference of Δ between the first and second intervals.

Therefore, a level meter should behave the same on both types of experiments.

B. ILD vs DLI

Although the thresholds in Figs. 2–5 are all approximately the same, there are systematic differences for different experiment types. These differences show that the level-meter model cannot be exactly right. The largest systematic difference is between the DLI and the ILD threshold for correlated noise. The ILD is notably smaller.

The smaller ILD threshold suggests the involvement of a specific binaural comparison process. For example, neurons in the lateral superior olive that are sensitive to finite ILDs are not involved in the DLI experiment because ILDs are zero in that experiment. Such neurons are activated by the ILD experiments, and information from them could be responsible for the lower threshold for ILD compared with DLI.

Alternatively, our theoretical comparison of the DLI and ILD tasks may be incomplete because the two tasks impose different memory requirements. The DLI task requires the comparison of two sequential loudnesses; an ILD task requires the comparison of two sequential lateral positions.

C. The loudness-meter model

According to the level-meter model, binaural coherence should play no role in ILD thresholds. The experiments of this article have shown that this concept is approximately true. However, it does not appear to be entirely true. Our best estimate for the deviations from the simple level-meter model appear as the averaged data for experiment 2 in Fig. 5. There, ILD thresholds are shown to be lowest for correlated noise, highest for uncorrelated noise, and intermediate for anticorrelated noise, though the difference between correlated and anticorrelated noise was not found to be statistically significant.

We discovered that a straightforward modification of the level-meter model can account for the small deviations observed experimentally. A clue actually appeared in the work by Grantham and Ahlstrom (1982). Their measurements showed that ILD thresholds were similar for correlated and uncorrelated noise when the noise bursts were 500 ms in duration, but thresholds were usually larger for uncorrelated noise when the noise bursts were only 30 ms in duration. This result indicates that listeners are able to average over the additional variability in uncorrelated noise given the opportunity to do so, but that the additional variability leads to increased thresholds when listeners are prevented from integrating. This insight further suggests that for long stimulus durations, as used in the experiments of this article, the thresholds for uncorrelated noise are limited by the finite human auditory integration time. The auditory integration time is an element in the perception of loudness.

The remainder of this section introduces the loudness-meter model as an improved alternative to the level-meter model. The loudness-meter model includes two elements of loudness, namely temporal integration and compression. It also includes half-wave rectification of the signal, a physiologically based feature that is common in current auditory models. This section shows that with reasonable values of the integration time, the additional variance for uncorrelated noise is just large enough to account for the threshold differences seen experimentally.

The loudness-meter model, from signal to decision criterion, is as follows:

Let the signal in ear e ($e=L$ or R) be written as

$$x_e(t)=\sum_{n=1}^{N} A_n^e \cos(2\pi f_n t+\phi_n^e), \tag{1}$$

where A_n^e and ϕ_n^e are the amplitude and phase of component n in ear e. The amplitude is given by a Rayleigh distribution (e.g., Hartmann, 1997) and the phase is rectangularly distributed over 360 degrees. On different experimental intervals, the amplitudes and phases of the components will be different, and these differences lead to the stimulus variability.

The excitation rate is a half-wave rectified and compressed version of the signal,

$$r_e(t)=\{\mathcal{H}[x_e(t)]\}^{\eta}, \tag{2}$$

where $\mathcal{H}$ is the half-wave operator, and η is the compression exponent. The half-wave rectification will be directly responsible for the fact that anticorrelated noise leads to greater variability than correlated noise. The compression exponent is taken to be $\eta=0.6$, consistent with the observed rule for loudness (Fletcher, 1953; Stevens, 1955).

The excitation attributable to ear e, obtained by the listener on an experimental observation interval, is the integral of the excitation rate, $r(t)$,

$$E_e=\int_0^T dt\, r_e(t), \tag{3}$$

where T is the integration time. In the limit that the integration time is infinite, the system functions like the ideal level meter and the variability in E_e depends only on the duration of the excitation.

Finally, the left-or-right decision criterion, based on ILD, is the excitation difference relative to the mean,

$$\Delta=2\,\frac{E_R-E_L}{E_R+E_L}. \tag{4}$$

The advantage in normalizing Δ by the mean in this way is that it makes the decision criterion independent of the overall level of the experiment. The sign of criterion Δ determines whether the image is perceived to be on the left or the right. The magnitude of Δ determines whether the image is close to the midline or off to the side.

Because of the stimulus variability, the values of E_R and E_L are different on different experimental intervals, and the value of Δ varies from trial to trial. For correlated noise, however, any variation in E_R is perfectly mirrored in E_L. Because of the normalization, Δ becomes a function of the stimulus ILD only and there is no stimulus variability. If noises are not perfectly interaurally correlated, there is a trial-to-trial variation in Δ, characterized by a stimulus variance, σ_S^2.

D. The loudness-meter model and coherence

1. The sensitivity index

Average $\overline{\Delta}$ is the expected value of the decision variable, Δ. It is the ensemble average, i.e., the average over all possible signals, as specified by the sets of amplitudes and phases $\{A^L\}\otimes\{A^R\}\otimes\{\phi^L\}\otimes\{\phi^R\}$.

In an experiment with stimuli drawn from the ensemble, the variance of Δ is $\mathrm{var}(\Delta)=\sigma^2$. Together with $\overline{\Delta}$, the variance determines the sensitivity index d' (Green and Swets, 1966),

$$d'=\overline{\Delta}/\sigma. \tag{5}$$

The variance is composed of two parts, an internal noise σ_N^2 and a stimulus noise σ_S^2. If these two forms of noise are assumed to be additive and independent, the total variance is given by

$$\sigma^2=\sigma_N^2+\sigma_S^2. \tag{6}$$

2. Detailed calculation

A key to the calculation is that when the stimulus is coherent (diotic-except-for-level), there is no stimulus variance, i.e., $\sigma_S=0$. Because the stimulus variance is zero, all the total variance is internal, and the measured ILD threshold, L_R-L_L, gives an estimate for σ_N. That calculation is simplified by using an analytic form for the ensemble average $\overline{\Delta}$ in terms of the interaural level difference (L_R-L_L) in dB,

$$\overline{\Delta}=2\,\frac{10^{\eta(L_R-L_L)/20}-1}{10^{\eta(L_R-L_L)/20}+1} \tag{7}$$

or

$$\overline{\Delta}=2\tanh[0.057\,56\,\eta(L_R-L_L)].$$

Equation (7) is easily proved for noise that is diotic-except-for-level. To prove it for other conditions requires that the ensemble average of several nonlinear functions be equal to the functions of their ensemble-averaged arguments, not generally a valid step. However, numerical experiments using ensembles of 2000 waveforms showed that Eq. (7) holds good to much better than 1% accuracy.

The loudness-meter model was tested on the average results for experiment 2. The calculation was a three-step process. The first step required finding the internal noise, σ_N, from the threshold ILD for correlated noise using Eqs. (7) and (5) with $d'=1.16$. The ILD threshold of 0.68 dB led to $\sigma_N=0.0405$. Because of the normalization in Eq. (4) both Δ and σ are dimensionless.

The second step required the computation of the stimulus variance σ_S^2 as a function of integration time. Calculations of the variance of Δ, based on 2000 waveforms for a given integration time, led to the values of σ_S in Table II. As it turned out, the variance for uncorrelated noise was larger than the variance for anticorrelated noise by about a factor of 2. It was not *a priori* evident that this would be the case, but it always was, for any value of integration time we studied (25 to 400 ms).

TABLE II. Values of internal and stimulus noise computed for experiment 2.

T (ms)	σ_N	$\sigma_{S=\pi}$	$\sigma_{S=u}$
200	0.0405	0.019 38	0.036 90
300	0.0405	0.015 34	0.030 04
400	0.0405	0.012 85	0.026 24

The third step added the internal and stimulus variances according to Eq. (6) and used the total to predict threshold ILDs (L_R-L_L) from Eqs. (5) and (7). The predicted thresholds for integration times of 200 and 400 ms are given by the dashed and dotted lines in Fig. 5. All these useful integration times are less than the duration of our experimental stimuli, 500 ms, indicating that the experimental method did not impose an important duration limitation. The best agreement between model and experiment is obtained with an integration time of 300 ms, shown in Fig. 5 by the solid line. An integration time of 300 ms agrees with the value obtained by Plomp and Bouman (1959) for low-frequency stimuli such as ours, although it should be noted that their integration window was exponential whereas ours was rectangular. With an integration time of 300 ms, the agreement between the model and experiment is excellent.

E. The loudness-meter model and bandwidth

As shown by Table II, the noise that limits performance in experiment 2 is mostly internal noise. In the spirit of the loudness-meter model, it ought to be possible to improve the signal-to-noise ratio by increasing the bandwidth of the stimulus so that more channels of the binaural system are used. Because experiments 1 and 2 showed that ILD thresholds always decreased when the bandwidth was increased from 1 to 10 kHz, there is preliminary evidence that such an effect might be operating. For instance, the ILD threshold for correlated noise decreased from 0.68 to 0.45 dB with the increased bandwidth. It is interesting to ask whether this decrease can be predicted by the loudness-meter model.

The number of auditory channels involved in processing can be taken to be the number of critical bands within the range of the stimulus. A top frequency of 1000 Hz corresponds to 15.5 Cam, and a top frequency of 10 000 Hz corresponds to 35.2 Cam.

The ratio of the number of channels is 35.2/15.5 or 2.27 and the square root is 1.51. From Table II, the internal noise in experiment 2 is $\sigma_N=0.0405$, and dividing it by 1.51 leads to an expected value of $\sigma_N=0.0269$ for experiment 1. According to Eqs. (5) and (7) this leads to an expected threshold ILD of 0.45 dB, in exact agreement with experiment 1.

Several details of the loudness meter model for lateralization invite comparison with details from the study of loudness *per se*. Table II shows that the ILD experiments require that the internal noise needs to be several times the stimulus noise. This result agrees with the conclusions of loudness discrimination experiments (e.g., Raab and Goldberg, 1975). As noted above, the ILD experiments with correlated noise show improved performance with increasing bandwidth that agrees with the statistical ideal. Ideal bandwidth dependence was found by Green (1960). By contrast, Raab and Goldberg (1975) found only a weak dependence on bandwidth, and deBoer's (1966) model for discrimination actually considers the internal noise to be independent of bandwidth. Buus (1990) concluded that discrimination would improve with increasing bandwidth, as expected, for low-level noise but not for high. Applied to the bandwidth results in the present article, the Buus conclusions suggest that discrimination would improve when bandwidth is increased from 1 to 10 kHz, but not by as much as the ideal energy detector because of the upward spread of excitation in the 1-kHz case. What prevents us from reaching a firmer conclusion on this matter is that previous work has employed briefer noise bursts than ours and band-pass noises.

F. Alternative calculations and caveats

The compression exponent η was chosen to be 0.6 in the loudness-meter model because this is the value generally found in loudness experiments. This value of the exponent, combined with an integration time of 300 ms, led to good agreement with experimental data. However, alternative exponents are possible. We made a parametric study employing exponents of 0.2, 0.4, 0.6, 0.8, 1.0, 1.4, and 2.0. Although the values of Δ from Eq. (7) varied considerably, those variations were tracked by corresponding changes in σ_S from the numerical study of the variance of Δ. In the end, the calculations proved remarkably insensitive to the compression exponent. For instance, an exponent of 0.4, as suggested by the recent masking-level-difference study by Bernstein *et al.* (1999), led to good agreement with the measured thresholds when the integration time was increased to about 400 ms. Further, the predictions using an exponent of 0.4 and an integration time of 300 ms were not unreasonable in view of the uncertainty in the experimental thresholds.

A second, and more radical, reformulation of the model eliminated both the stages of half-wave rectification and compression. Instead, the calculation simply squared the input signals, $x_e(t)$, to obtain a power, and integrated the power to make a level-meter model with finite integration time, i.e., an *integrating level-meter model*. The purpose of this calculation was to allow us to separately assess the roles of temporal integration and of the other elements modeling the auditory system. The results of the calculation are easily described. Most important, the integrating level-meter model predicts no difference between correlated and anticorrelated noise—both have zero stimulus variance. Thus, this model does not fit the data in Fig. 5 as well as the level meter model, but it cannot be ruled out experimentally because the difference in ILD thresholds for correlated and anticorrelated noise did not reach statistical significance in any of our three experiments. Otherwise, the integrating level meter model fits the data for uncorrelated noise with an integration time of 240 ms, a not unreasonable value. Thus, in the end, the only aspect of the loudness meter model that is unequivocally indicated by our measured ILD thresholds is the temporal integration.

VI. CONCLUSIONS

The guiding hypothesis for this article was the level-meter model, which says that the human binaural system deals with interaural level differences (ILDs) for noise stimuli by making separate measurements of excitation from left and right ears and comparing those measurements at a higher level. According to the model, temporal fine structure of the noise is unimportant. Only the average excitation is important. Accordingly, the model predicts that ILD thresholds are the same whether the noise in the two ears is diotic-except-for-level (correlated), inverted (anticorrelated), or completely independent (uncorrelated). The model further predicts that ILD thresholds should be the same as intensity difference limens for diotic noise.

Experiments with noise having bandwidths of 10 and 1 kHz were performed to test the level-meter model. The results of the experiments can be briefly summarized by saying that the level-meter model works rather well. Although individuals differed from one another, thresholds for the four different experiment types agreed to within about half a dB. Therefore, for most practical purposes, one can rely on the level-meter concept.

At a finer degree of detail, the experiments showed some departure from the level-meter model. The largest difference was between the difference limen in intensity (DLI) and the ILD threshold for correlated noise. This difference might be due to specific binaural processes that contribute to ILD sensitivity but are excluded from diotic tasks like the difference limen. It might result from the different memory requirements of the two tasks.

Of greater interest were the differences among the ILD experiments themselves, with different noise coherence. Here, the differences were smaller. These differences could be explained by replacing the level-meter model by a loudness-meter model, incorporating temporal integration. The model further assumed that the total variance was the sum of internal noise and stimulus noise. Predictions of the model were in good agreement with the experimental results. The model also successfully accounted for the observed bandwidth dependence of ILD thresholds. An experimental observation that was not addressed by the model was that roving the level of the noise tended to increase thresholds (experiment 3 versus experiment 2). Averaged over listeners, the change was roughly the same for all experiment types, and it was never more than 0.2 dB. We tentatively attributed the change to a slight distraction caused by the rove.

To guard against misinterpretation, it is important to note that all the sounds considered in this article had similar amplitude spectra in left and right ears. A naive level-meter imposes no such requirement. It would lateralize equally well given a 100-Hz tone in one ear and a 5000-Hz tone in the other. The auditory system would not behave like this level meter because it would not fuse these tones into a single image. Although questions of spectral similarity have been deliberately avoided in this article, one imagines that one coherence-independent level meter per critical band might serve as an adequate model when different ears receive different spectral shapes.

There is recent animal physiological work that also suggests that the processing of ILDs in broadband noise is independent of the binaural coherence of the noise. Egnor (2001) examined the posterior part of the ventral nucleus of the lateral lemniscus—the first stage of ILD processing in the barn owl. She found no difference in neural activity when coherent noise was replaced by incoherent noise. Further, the owls retained their vertical plane behavioral response when the coherence of the noise was changed. By contrast, replacing coherent noise by incoherent noise eliminated the ability of barn owls to localize in the horizontal plane (Saberi *et al.*, 1998). Because barn owls appear to use ILDs and ITDs separately to localize in separate planes, the insensitivity to coherence in ILD may not be surprising.

Tollin and Yin (2002) studied the spatial receptive field (SRF—neural firing rate as a function of azimuth angle) of LSO neurons in cat. For the single neuron investigated, they showed that the shape of the SRF was unchanged when broadband coherent noise was replaced by incoherent noise. The peak was reduced by 15% to 25%, however. To summarize, both these studies of other species are broadly consistent with the coherence insensitivity demonstrated in the present article.

The results of this work have significance for the localization of steady-state sounds in rooms. Because of reflections from room surfaces, the signals to the two ears are always incoherent to some degree. In a large reverberant environment they are uncorrelated above 500 Hz (Lindevald and Benade, 1986). It is not possible to use interaural time differences on binaurally uncorrelated signals because there is no common feature to time. By contrast, the present experiments show that binaural coherence has an almost negligible effect on the use of interaural level differences. It may be that standing waves in the room cause the ILDs to be misleading about the true location of the source, but the nervous system is entirely capable of making use of this information or disinformation. This result tends to focus attention on the ILD as an important element in sound localization in a room.

ACKNOWLEDGMENTS

This work was supported by a grant from the National Institute of Deafness and other Communicative Disorders. We are grateful to Dr. Brad Rakerd and Mary Jo Hidecker for help with the statistical evaluation of the results and to Dr. Leslie Bernstein, Dr. Steve Colburn, and an anonymous reviewer for useful comments.

---

[1] In the search for a replacement for listener K in experiment 2, three other listeners (male students C and X, and middle-aged female J) were tested and repeatedly failed to meet our criterion for tight staircases when the bandwidth was 1 kHz. Especially curious was the case of listener C. Listener C had among the lowest thresholds for a bandwidth of 10 kHz. He also had low thresholds for a bandwidth of 1 kHz for all noise conditions except correlated noise. However, C had great difficulty in lateralizing correlated noise with a bandwidth of 1 kHz. According to C's own description, he perceived a "hole," which is an absence of noise. The hole moved opposite to the ILD-induced image. The effect was time dependent; C could perform well for most of a run before the hole would capture his attention, causing him to answer incorrectly 100 percent of the time, even at large ILDs. The authors introduced stimuli of shorter duration in order to reduce the salience of an ITD cue, but to no effect. Louder stimuli and the introduction of an interaural delay also failed to improve performance. The hole appeared less often when the low-pass cutoff was increased to 2 kHz, and it disappeared completely for a cutoff above 3 kHz. The authors have no explanation for any of this.

[2] Critical band numbers on the *Cam* scale refer to critical bands measured in Cambridge using notched-noise measurements (Glasberg and Moore, 1990; Moore, 1995). The Cam units are identical to the units regrettably called ERBs in much of the *fin de siecle* psychoacoustical literature.

Bernstein, L. R., van de Par, S., and Trahiotis, C. (**1999**). "The normalized interaural correlation: Accounting for $NoS\pi$ thresholds obtained with Gaussian and 'low-noise' noise," J. Acoust. Soc. Am. **106**, 870–876.

Blauert, J. (**1983**). *Spatial Hearing*, translated by J. S. Allen (MIT, Cambridge, MA).

Buus, S. (**1990**). "Level discrimination of frozen and random noise," J. Acoust. Soc. Am. **87**, 2643–2654.

deBoer, E. (**1966**). "Intensity discrimination of fluctuating signals," J. Acoust. Soc. Am. **40**, 552–560.

Dunn, O. J. (**1961**). "Multiple comparisons among means," J. Am. Stat. Assoc. **56**, 52–64.

Egnor, R. (**2001**). "Effects of binaural decorrelation on neural and behavioral processing of interaural level differences in the barn owl (Tyto alba)," J. Comp. Physiol., A 589–595.

Fletcher, H. (**1953**). *Speech and Hearing in Communication*, edited by J. B. Allen (Acoustical Society of America, New York, 1995) (reprint).

Glasberg, B. R., and Moore, B. C. J. (**1990**). "Derivation of auditory filter shapes from notched noise data," Hear. Res. **47**, 103–138.

Grantham, D. W., and Ahlstrom, J. B. (**1982**). "Interaural intensity discrimination of noise as a function of center frequency, duration, and interaural correlation," J. Acoust. Soc. Am. Suppl. 1 **71**, S86.

Green, D. M. (**1960**). "Auditory detection of a noise signal," J. Acoust. Soc. Am. **32**, 121–131.

Green, D. M. (**1988**). *Profile Analysis Auditory Intensity Discrimination*, Oxford Psychology Series No. 13 (Oxford, New York), p. 20.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York).

Hartmann, W. M. (**1997**). *Signals, Sound, and Sensation* (Springer-Verlag, New York), Appendix I.

Henning, G. B. (**1974**). "Detectability of interaural delay in high-frequency complex waveforms," J. Acoust. Soc. Am. **55**, 84–90.

Jesteadt, W., and Wier, C. C. (**1977**). "Comparison of monaural and binaural discrimination of intensity and frequency," J. Acoust. Soc. Am. **61**, 1599–1603.

Lindevald, I. M., and Benade, A. H. (**1986**). "Two-ear correlations in the statistical sound fields of rooms," J. Acoust. Soc. Am. **80**, 661–664.

Moore, B. C. J. (**1995**). "Frequency analysis and masking," in *Hearing, Handbook of Perception and Cognition*, edited by B. C. J. Moore (Academic, San Diego).

Neutzel, J. M. (**1982**). "Sensitivity to interaural intensity differences in tones and noise as measured with a roving level procedure," J. Acoust. Soc. Am. Suppl. 1 **71**, S47.

Plomp, R., and Bouman, M. A. (**1959**). "Relation between hearing threshold and duration for tone pulses," J. Acoust. Soc. Am. **31**, 749–758.

Raab, D. H., and Goldberg, I. A. (**1975**). "Auditory intensity discrimination with bursts of reproducible noise," J. Acoust. Soc. Am. **57**, 437–447.

Saberi, K., Takahashi, Y., Konishi, M., Albeck, Y., Arthur, B. J., and Farahbod, H. (**1998**). "Effects of decorrelation on neural and behavior detection of spatial cues," Neuron **21**, 789–798.

Stevens, S. S. (**1955**). "The measurement of loudness," J. Acoust. Soc. Am. **27**, 815–829.

Tollin, D. J., and Yin, T. C. T. (**2002**). "The coding of spatial location by single units in the lateral superior olive of the cat, I. Spatial receptive fields in azimuth," J. Neurosci. **22**, 1454–1467.

von Békésy, G. (**1930**). "Zur Theorie des Hörens: Über das Richtungshören bie einer Zeitdifferenze oder Lautstärkeungleichheit der beidseitigen Schalleinwirkungen," Phys. Z. **31**, 824–838, 857–868; cited by Blauert (1983), p. 162.

Yost, W. A. (**1981**). "Lateral position of sinusoids presented with interaural intensive and temporal differences," J. Acoust. Soc. Am. **70**, 397–409.