

Are There Limits to Binaural Additivity of Loudness?

Gerd Gigerenzer and Gerhard Strube
Ludwig-Maximilians-Universität München, Germany

In recent years there has been notable interest in additive models of sensory integration. Binaural additivity has emerged as a main hypothesis in the loudness-scaling literature and has recently been asserted by authors using an axiomatic approach to psychophysics. Restrictions of the range of stimuli used in the majority of former experiments, and inherent weaknesses of the axiomatic study by Levelt, Riemersma, and Bunt (1972) are discussed as providing reasons for the present investigation. A limited binaural additivity (LBA) model is proposed that assumes contralateral binaural inhibition for interaural intensity differences that exceed a critical level. Experimental data are reported for 12 subjects in a loudness-matching task designed to test the axioms of cancellation and of commutativity, both necessary to the existence of strict binaural additivity. In a 2×2 design, frequencies of 200 Hz and 2 kHz were used, and mean intensity levels were 20 dB apart. Additivity was found violated in 33 out of 48 possible tests. The LBA model is shown to predict the systematic nonadditivity in the loudness judgments and to conform to results from other studies.

Integrating sensory information from two or more perceptual dimensions or channels is a central feat accomplished by the human information-processing system, and much effort has been made to clarify the rules of sensory integration. In recent years, there has been notable interest in a broad variety of research questions regarding additive models of integration. Examples are provided in the studies by Krantz and Tversky (1975) on visual perception, Wilkening (1979) on perceptual development, and Anderson (1981) on information integration theory.

Two distinct methodological approaches have been taken to generate and test additive models of sensory integration. The traditional method is psychophysical scaling (Stevens, 1975). Although the practical necessity and usefulness of that approach is widely appraised, fundamental problems are involved in scaling or "direct measurement." Representatives of the second approach, fundamental measurement theories (Krantz, Luce, Suppes, & Tversky, 1971), have pointed out that scaling procedures rest on the assumption of some metric that is established

per fiat (fitting equations to data), whereas fundamental measurement theories explicitly state the conditions necessary for the joint existence of, for example, interval scales and an additive integration rule. A similar approach has been taken by "additive functional measurement" (Anderson, 1981); this may be regarded as a special case of that kind of fundamental measurement theory known as additive conjoint measurement (Schöenemann, Cafferty, & Rotton, 1973).

Binaural Additivity

The issue of binaural additivity of loudness provides an outstanding example of an additive integration hypothesis that was formulated at least as early as Urbantschitsch's (1893) study. Both the scaling approach and fundamental measurement theories have been applied to that issue, the latter even in probabilistic formulations (Falmagne, 1976; Falmagne, Iverson, & Marcovici, 1979).

The basic phenomenon is well known: A tone of given physical intensity appears louder when presented to both ears than when presented to only one ear. The question is, Can the perceived loudness of the binaural stimulus be expressed through some algebraic rule of combining relevant stimulus attri-

Requests for reprints should be sent to Gerd Gigerenzer, Institut für Psychologie der Universität München, Friedrichstrasse 22/D-8000 München 40, Germany.

butes (i.e., intensity or loudness) in the left and right auditory channels?

In the past, summation of sound energy was assumed to enhance binaurally perceived loudness, and it seems that von Békésy (1929) was the last to propose that explanation. Binaural summation of loudness rather than physical intensity was invoked more successfully (Fletcher & Munson, 1933). An ongoing debate exists as to whether binaural summation is perfect (Hellman & Zwislocki, 1963; Marks, 1978) or only partial (Reynolds & Stevens, 1960; Scharf & Fishken, 1970). Experimental data suggest that degree of summation is independent of tone frequency (even if frequency is different in both ears; Scharf, 1969) but generally dependent on the intensity and nature (e.g., wide-band noise vs. pure tones) of the stimulus (Durlach & Colburn, 1978).

Despite those variabilities and the variation observed in coefficients of loudness power functions (Hellman & Zwislocki, 1963), agreement is widespread as to the existence of binaural additivity of loudness in general. This means that loudness values established independently of each other at both aural channels are additively combined into a central loudness sensation. The term *binaural additivity* is used here in this sense.

With a scaling approach, additivity depends on the metric of the loudness scale. However, the validity of the output metric of subjects' judgments is not open to an empirical test and consequently does not allow for rigorous testing of binaural additivity (Levelt, Riemersma, & Bunt, 1972).

Axiomatic Studies

The fundamental measurement theory (Krantz et al., 1971) opened up a new and promising approach because it enabled experimenters to test specific hypotheses of binaural combination without the prerequisite of loudness scale construction.

In general, a combination rule is specified (e.g., an additive model). Again, it must be noted that additivity is used in a very general sense to mean that values are taken independently of each other, transformed, and finally combined to form a total value; summation, then, is a special case of additivity. In the next

step, axioms (i.e., conditions necessary for the model to hold) have to be defined and then may be tested empirically. As a result, the model being tested can be refuted or corroborated, and the degree of corroboration will depend on the severity of the test (Popper, 1963).

Levelt et al. (1972) were first to test the double cancellation axiom (a condition necessary for additivity) with data from paired-comparison binaural loudness judgments. They concluded that additivity exists. Falmagne (1976) tested the axioms of cancellation and commutativity (both necessary conditions for his probabilistic version of the additive model) and found the laws of additivity violated. But that study was based on only a few hundred trials taken by a single subject, and in a further experiment employing three subjects and some thousand trials each, evidence was accumulated for a specific additive model (Falmagne et al., 1979). One of the main results, however, is shown by Falmagne et al. to be consistent with at least one nonadditive model (see Equation 22 in Falmagne et al., 1979). After weighing the evidence, Falmagne et al. tentatively favor an additive model, in view of Levelt et al.'s results.

Aims of the Present Article

Generally, "the work of Levelt and his colleagues is rightfully considered as a key result in this area" (Falmagne, 1976, p. 67). The design of the Levelt et al. study, however, will be shown to have substantial shortcomings that prevented a truly empirical test of binaural additivity in their experiment.

A further reason to question binaural additivity is the use of a limited range of interaural intensity differences in the scaling literature (e.g., Hellman & Zwislocki, 1963; Reynolds & Stevens, 1960) and in axiomatic studies (Falmagne et al., 1979). Even if binaural additivity exists, it may be restricted to equal or nearly equal differences in intensity between the left and right channels. Therefore we conducted an experiment that used a sufficiently wide range of interaural intensity differences; results are discussed with respect to models for complete and for limited binaural additivity.

Critique of the Study by Levelt,
Riemersma, and Bunt

Levelt et al. (1972) formalized binaural additivity in terms of additive conjoint measurement. They tested two subjects in 18 circumstances involving a number of replications of a few basic conditions, and concluded that "the 18 data matrices strongly show the double cancellation property, which leads to the conclusion that, for our subjects and conditions, binaural additivity exists" (p. 60).

The theory of additive conjoint measurement states three testable conditions for additivity: transitivity, independence (single cancellation), and double cancellation (Krantz et al., 1971). Levelt et al., for economical reasons, tested neither transitivity nor independence but replaced the test of the latter by a "monotonicity assumption." Therefore, from the three generally testable conditions for additivity, double cancellation alone was subjected to empirical test.

Double cancellation is a property that necessarily follows from binaural additivity. If $a, b, c \in A$ are the intensities of pure tones

presented to the left ear, and $p, q, r \in P$ are the intensities of pure tones presented to the right ear, then $ap \in A \times P$ denotes a binaural pair. Let f and g be loudness functions for the left and right ears, respectively. The symbol \succeq stands for "is judged louder than or equal to." Under conditions of binaural additivity, a comparative judgment about tone pairs can be expressed as

$$ap \succeq bq \text{ iff } f(a) + g(p) \geq f(b) + g(q), \quad (1)$$

or, equivalently,

$$ap \succeq bq \text{ iff } f(a) - g(q) \geq f(b) - g(p). \quad (1a)$$

Double cancellation is expressed by

$$\text{if } ap \succeq bq \text{ and } br \succeq cp \text{ then } ar \succeq cq. \quad (2)$$

By expressing the binaural loudness judgments in Equation 2 in terms of differences of loudnesses, as given in Equation 1(a), we get the following:

if

$$f(a) - g(q) \geq f(b) - g(p)$$

and

$$f(b) - g(p) \geq f(c) - g(r),$$

then

$$f(a) - g(q) \geq f(c) - g(r). \quad (2a)$$

This equation clarifies the meaning of double cancellation as a special case of the well-known transitivity inherent in additive structures. Double cancellation means that if binaural additivity exists, binaural loudness differences have to be transitive.

To test double cancellation, Levelt et al. constructed a rank order of the stimulus pairs with respect to their binaural loudness, counting how often a given pair dominates (i.e., is judged louder than) another one. Because the 2,592 paired comparisons required to construct that rank order are not feasible for economical reasons, only 450 comparisons were realized in the experiment; the other comparisons were estimated using the aforesaid monotonicity assumption. Therefore, no less than 82.64% of Levelt et al.'s data were defined a priori.

Although monotonicity does not generally imply the double-cancellation property (Levelt et al., 1972, construct a counterexample for demonstration), the Levelt et al. a priori data structure does imply double cancella-

Table 1
Double Cancellation in the Predetermined Rank-Order Structure of Levelt et al.'s (1972) Data

Intensities (dB) at right ear	Intensities (dB) at left ear					
	20	30	40	50	60	70
20	0	2	4 ^a	6	8	10
30	2	6	10 ^b	14	18	22
40	4	10	16	22	28	34
50	6	14	22	30	38	46
60	8	18	28	38	48	58
70	10	22	34	46	58	70

Note. Of 2,592 (= $36 \times 36 \times 2$) paired comparisons necessary for the 36 binaural pairs and two orders of presentation, only 450 were obtained empirically by Levelt, et al. (1972). The rest were constructed by definition and are reconstructed here according to the procedure of Levelt et al. (1972, p. 58).

^a The value of 4 (e.g., for the pair [20, 40]) means that (20, 40) dominates per monotonicity assumption two binaural pairs (viz, [20,20] and [20, 30]), and because of two different orders of presentations, this number is multiplied by two.

^b 3×3 submatrices, like the one shown enclosed in dotted lines, test double cancellation, which means that if $(30, 30) \succeq (40, 20)$, and $(40, 40) \succeq (50, 30)$ then $(30, 40) \succeq (50, 20)$ must hold.

tion, which can be computed from the values and method given in Table 1.

Levelt et al.'s a priori structure contains more than 80% of the total information on which their subsequent test of double cancellation rests. The comparatively small amount of empirical data (17.36%) added to those values can scarcely be expected to distort their predetermined structure severely. A comparison of Levelt et al.'s a priori structure with their results helps to make that point clear.

Table 1 shows the number of domination instances (ranks) for all pairs that can be computed from the a priori values alone. In the 400 instances (i.e., the 400 3×3 submatrices embedded in their 6×6 matrix), double cancellation is never violated; it holds in 252 instances and is not open to test in 148 instances. The a priori structure yields a nearly perfect prediction of Levelt et al.'s data, which reported an average of 3.9 violations, 244.7 positive instances, and 151.4 no-test situations. Ostensibly, the small amount of true empirical data did not noticeably change the a priori structure. Note that even the proportion of no-test instances is nearly identical. We therefore conclude that the empirical status of the double-cancellation test as performed by Levelt and his co-workers is severely shaken. If the comparisons not determined a priori are computed as random, with a probability of .5 for any given pair *ap* dominating another pair, a matrix again results that is completely additive.

A second line of argument is independent of the artificial nature of Levelt et al.'s (1972) data. Even if the axioms of transitivity, independence, and double cancellation had been shown to stand up to empirical testing, we ought to be careful in evaluating the status of these necessary, but not sufficient, conditions as "practically sufficient," because it depends on the size of the design. Generally, the larger the design, the smaller the probability that additivity exists when the validity of the three necessary conditions has been established (Arbuckle & Larimer, 1976; McClelland, 1977). Although Arbuckle & Larimer's Monte Carlo studies do not report values for a 6×6 design (as was used by Levelt et al.), we may take their value for a

7×5 design as a minimum: No less than 68% of all cases in which independence and double cancellation are valid can still be shown to be nonadditive, a proportion that would be even higher for a 6×6 matrix of stimuli. For these reasons, the study by Levelt et al. does not provide strong support for binaural additivity of loudness.

A Model of Limited Binaural Additivity

Because strict binaural additivity has been brought into question, what alternatives may be thought of? Physiology suggests numerous ways of interaction between left and right ears (e.g., a detector mechanism for interaural time and intensity differences, which determine the degree of lateralization; Strube, 1982). More specifically, data from binaural/binaural loudness-matching experiments suggest an inhibitory effect of the louder tone on the contralateral, softer one; if the interaural intensity difference exceeds 6–12 dB (less for pure tones, more for noise; Irwin, 1965; Keen, 1972). Binaural loudness combination only appears linear with rather small interaural intensity differences (characteristically, when lateralization is not complete). Outside that range, binaural loudness gain becomes spurious, possibly because of contralateral binaural inhibition.

Because physiological details are not well known to date, and because the construction of approximating functions, apart from the general deficiencies already noted, is open to considerable variation (Treisman & Irwin, 1967), no effort is made to construct a best-approximation function. Instead, we prefer to confine ourselves to the delimiting of a tolerance field, which may be defined by assuming strict additivity (with a gain of 6–10 dB over mean intensity), until (at an intensity difference of twice the maximum gain) the louder tone alone is assumed to determine binaural loudness. Figure 1 shows the tolerance field for this *limited binaural additivity* (LBA) model.

Let $a, b \in A$ be the intensities of pure tones presented to the left ear, and $p, q \in P$ be the intensities of pure tones presented to the right ear. The model then states that

$$ap \gtrsim bq \text{ iff } \max[M(a, p) + \delta, \max(a, p)] \\ \geq \max[M(b, q) + \delta, \max(b, q)], \quad (3)$$

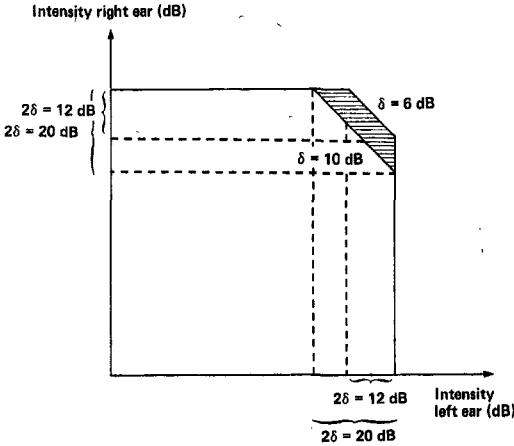


Figure 1. Equiloudness contours for the limited binaural additivity model stated in Equation 3. (The cross-hatched area is the tolerance field for deltas between 6 and 10 dB.)

with M designating the arithmetic mean and δ designating the gain. The symbol \geq stands for "is judged louder than or equal to."

Clearly this model is simplified because it assumes binaural symmetry and an abrupt, instead of gradual, increase in interaural inhibition. The LBA model, however, is easily demonstrated, specifies limits for binaural additivity, and is capable of computing predictions of experimental outcomes.

In addition, a quick comparison with Irwin's (1965) data shows a surprisingly good approximation for so simple a model, with correlations of .9 and higher. (Incidentally, Treisman & Irwin, 1967, reported correlations of .97 with various power functions, independent of the exact combination rule. This fact amply demonstrates that function approximation alone, apart from detecting gross errors, is not useful for testing models because significant deviations are rarely found.)

Testing Binaural Additivity

Our axiomatic formulation and method of testing binaural additivity is similar to Falmagne's (1976) random conjoint measurement model. Let $a, b \in A$ be the intensities of pure tones presented to the left ear, and $p, q \in P$ be the intensities for the right ear. The experimental paradigm of binaural matching requires the subject to adjust the

intensity b so that two alternately presented tone pairs (a, p) and (b, q) appear equally loud (intensities a, p , and q are fixed by the experimenter).

For the matching paradigm, the hypothesis of binaural additivity may be stated algebraically as follows:

$$ap \sim bq \text{ iff } f(a) + g(p) = f(b) + g(q). \quad (4)$$

The functions f and g may be regarded as monaural loudness functions for the left and right ear, respectively, accounting for possibly different loudness functions for each ear. Falmagne (1976) gives a probabilistic version of Equation 4, replacing b with a random variable $U_{pq}(a)$ and adding an error term $E_{pq}(a)$. This change makes the right side of Equation 4

$$f[U_{pq}(a)] = g(p) - g(q) + f(a) + E_{pq}(a). \quad (5)$$

Accordingly, the well-known necessary conditions in the algebraic framework of additive conjoint measurement are reformulated. The first of the two testable conditions (for all $a \in A$ and $p, q, r, s \in P$) is *cancellation*:

$$m_{pr}(a) = m_{pq}[m_{qr}(a)], \quad (6)$$

in which $m_{pr}(a)$ is the median of the distribution of values b that results from the subject first hearing a tone pair (a, p) , then a tone r at the right ear together with a tone b at the left ear, and adjusting b so that the loudness of (b, r) appears equal to the loudness of (a, p) .

This cancellation rule is the probabilistic counterpart of the double cancellation condition (Falmagne, 1976, p. 70), which relates to the transitivity of binaural loudness differences.

The second of the testable conditions necessary for binaural additivity is *commutativity*:

$$m_{pq}[m_{rs}(a)] = m_{rs}[m_{pq}(a)] \quad (7)$$

Commutativity refers to the property of addition that, when adding two numbers x and y , $x + y = y + x$ (i.e., the result of addition is order independent). Thus, if binaural additivity holds, it should make no difference if the resulting median is constructed by balancing against p and q first and then against r and s , or against r and s first and then against p and q .

The meaning of these conditions will also be clarified when we consider the testing procedure.

Testing Cancellation

Step 1. The subject adjusts an intensity b so that (a, q) sounds equally loud as (b, r) for fixed intensities $a, q,$ and r . Repeating this procedure k_1 times yields the median $b' = \hat{m}_{qr}(a)$, which is an estimator for the median $m_{qr}(a)$.

Step 2. The equivalent procedure, with fixed intensities $b', p,$ and $q,$ and adjusting c so that (b', p) appears equal to (c, q) , repeated k_2 times, yields the median $c' = \hat{m}_{pq}[m_{qr}(a)]$.

Step 3. Similarly, the adjustment of d (with fixed intensities $a, p,$ and r) so that (a, p) equals (d, r) , repeated k_3 times, gives the median $d' = \hat{m}_{pr}(a)$.

If binaural additivity holds, no difference must exist between the distributions of c and d with medians c' and d' .

Testing Commutativity

Step 1. The subject adjusts an intensity b so that (a, r) sounds equally loud as (b, s) for fixed intensities $a, r,$ and s . Repeating this procedure k_1 times gives a distribution with median $b' = \hat{m}_{rs}(a)$.

Step 2. An intensity d is adjusted by the subject so that (b', p) equals (d, q) . Repeating this k_2 times yields the median $d' = \hat{m}_{pq}[m_{rs}(a)]$.

Step 3. Similarly, c is adjusted so that (a, p) equals (c, q) . Repeating this procedure k_3 times gives the median $c' = \hat{m}_{pq}(a)$.

Step 4. An intensity e is adjusted so that (c', r) equals (e, s) . Repeating the procedure k_4 times yields the median $e' = \hat{m}_{rs}[m_{pq}(a)]$.

If binaural additivity holds, no difference must exist between the distributions of d' and e' with medians d' and e' .

Predictions of the Limited Binaural Additivity Model

If we apply the LBA model of Equation 3 to the matching paradigm, we get

$$ap \sim bq \text{ iff } \max[M(a, p) + \delta, \max(a, p)] = \max[M(b, q) + \delta, \max(b, q)]. \quad (8)$$

The model not only allows for the prediction of equality or direction of deviation of

Table 2
Predictions of Experimental Results from Strict and Limited Binaural Additivity Models

Comparison of medians		Range of medians		
Strict	Limited	Strict	Limited	
			$\delta = 6 \text{ dB}$	$\delta = 10 \text{ dB}$
Cancellation				
$c = d$	$c > d$	No prediction	$c = 71.25$ $d = 69.00$ $\Delta = +2.25$	$c = 77.25$ $d = 73.00$ $\Delta = +4.25$
Commutativity				
$d = e$	$d > e$	No prediction	$d = 69.75$ $e = 69.00$ $\Delta = +.75$	$d = 75.75$ $e = 72.00$ $\Delta = +3.75$

the medians in question but also provides an estimation of the absolute values of the medians.

Choosing the fixed intensities is a crucial point for the predictions of this model in relation to a model that is completely additive. Ostensibly, the LBA model reduces to an additive model for all cases in which the intensity difference of a binaurally presented tone pair does not exceed 2δ . In order to get different predictions from the completely additive and LBA models, we chose $a = 57, p = 69, q = 54,$ and $r = 42$ for the test of cancellation, and $a = 57, p = 69, q = 54, r = 48,$ and $s = 42$ for the test of commutativity (all values are in dB SPL for the loud condition; all values are 20 dB less in the soft condition).

For deltas of 6 dB and 10 dB, we get the prediction that

$$m_{pq}[m_{qr}(a)] > m_{pr}(a) \quad (9)$$

(cancellation violated), as well as

$$m_{pq}[m_{rs}(a)] > m_{rs}[m_{pq}(a)] \quad (10)$$

(commutativity violated).

Table 2 shows the predicted values for c and d (test of cancellation) and for d and e (test of commutativity).

Method

Subjects

Twelve subjects (4 female, 8 male [10 psychology students and the authors]) received eight sessions within 2

weeks (four conditions \times two axioms). The order of experimental conditions was randomly determined for each subject. All subjects had shown normal hearing ability in pretest sessions and had received training sessions for the matching task.

Apparatus and Procedure

Experimental apparatus and procedure were modeled after those in Falmagne's (1976) study. The main deviations were as follows: (a) Four experimental conditions were used instead of one, and (b) automatic control of the experiment served to shorten experimental sessions (no loss of time between trials) and to rule out errors by the experimenter.

Signals. 200-Hz and 2000-Hz sine-wave tones were gated at 2 Hz with a duty cycle of 50%. The reference levels for intensity were 70 dB and 90 dB SPL, respectively. Intensity of the fixed stimuli ranged from -21 dB to -48 dB with respect to the above reference levels. Intensities used for testing the cancellation axiom were $a = -33$ dB, $p = -21$ dB, $q = -36$ dB, and $r = -48$ dB. For testing commutativity, a , p , and q were used as before, but in this condition, $r = -42$ dB and $s = -48$ dB.

Apparatus. The stimuli were generated by a Newtronics-function generator and gated through an analog switch (using TI 604), then divided into a left and right channel. Precision attenuators (using AD 7110, 0 to -88.5 dB in 1.5 dB steps), standard amplifiers (Kenwood), and headphones (Sennheiser) completed the signal path. Oscilloscope (Gould) and sound-level meter (with artificial ear, Brüel & Kjaer) were used for monitoring. A microcomputer system (Synertek SYM-1, using 6502 CPU) controlled attenuators, logged data, computed medians (for testing cancellation, medians were

computed in Step 2; for testing commutativity, medians had to be computed in Steps 2 and 4 of the experiment), and interacted with the subjects by means of a small control panel. The control panel contained two light-emitting diodes (LEDs) for signaling trials and pauses and four buttons: (a) start trial/present reference pair, (b) louder, (c) softer, (d) OK, match—trial over. The subjects were seated in a sound-attenuated room (30 dB to 32 dB[A] residual noise; further attenuated, of course, by the headphones). Details of the instrumentation and the computer program have been published elsewhere (see Strube, 1981, pp. 172-187).

Procedure. Each trial consisted of the following: (a) the subject pressing the start button, whereupon the reference pair was presented for 4 sec, after which the stimulus pair to be matched was presented. (b) The subject could now adjust the loudness of the variable stimulus of the pair by pressing louder or softer buttons, and at any time switch to hearing the reference pair. (c) When the subjects determined that a match had been achieved, they terminated the trial by pressing the OK button.

Choosing optimal values for the number of trials in each step of the experiment requires a number large enough for the empirically obtained medians to be good estimations of the population medians, and small enough to keep the subject from becoming a victim of fatigue. Measurements from extended training sessions with our subjects indicated that values of $k_1 = 41$, $k_2 = k_3 = 20$ (cancellation) and $k_1 = k_3 = 41$, $k_2 = k_4 = 20$ (commutativity) fulfilled these requirements. Thus, for each of the four experimental conditions in our 2×2 design (frequencies of 200 and 2000 Hz, combined with intensity reference levels of 90 and 70 dB SPL), an experimental session consisted of 81 or 122 trials, respectively, plus three initial trials that were not evaluated, to create a total of 836 trials for each subject.

Table 3
Results of Testing the Axioms of Cancellation and of Commutativity

Subject	200 Hz				2000 Hz			
	Intensity low		Intensity high		Intensity low		Intensity high	
	Can.	Com.	Can.	Com.	Can.	Com.	Can.	Com.
1	3.780*	3.057*	2.231*	-.592	4.781*	2.341*	5.044*	3.956*
2	1.743	4.030*	1.309	5.371*	.554	5.274*	.622	5.456*
3	1.971*	2.091*	2.271*	4.632*	-.601	-.247	.980	2.392*
4	1.484	-1.299	.295	4.962*	-2.557*	.593	-1.799	4.279*
5	.419	-.028	2.760*	5.105*	-.198	2.528*	1.788	-1.108
6	-3.226*	3.890*	1.212	-.663	-1.195	-.196	-3.761*	3.441*
7	-1.580	-.292	.096	-.371	-.042	1.174	-1.073	-.906
8	4.750*	4.981*	-.393	-5.389*	1.936	-2.549*	.875	-3.877*
9	-.123	.090	2.129*	1.961*	4.022*	-.703	1.314	4.933*
10	-2.021*	4.962*	3.284*	1.765	.667	2.814*	-.744	4.032*
11	2.687*	-.478	-.236	-1.057	.554	-.240	.723	3.048*
12	4.415*	-2.505*	.155	-1.019	1.623	-.715	2.588*	2.692*
<i>M</i> <i>z</i>	2.350	2.309	1.364	2.740	1.561	1.615	1.776	3.343

Note. Can. = cancellation; Com. = commutativity. A significant *z* value means that there is a significant violation of additivity (i.e., a significant difference between medians c' and d' [cancellation], or d' and e' [commutativity], respectively). A positive value of *z* indicates that $c' > d'$ ($d' > e'$), a negative value, that $c' < d'$ ($d' < e'$).

* $p < .05$.

Results

Testing Strict Additivity

Binaural additivity prohibits the medians c' and d' (cancellation), as well as d' and e' (commutativity) to differ significantly from each other. A z statistic was computed according to the method of Raatz (cited in Liernert, 1973, p. 233). Table 3 shows the results for our 12 subjects under the four conditions. In 33 of these 48 cases, at least one of the necessary conditions is violated. Therefore, the hypothesis of unrestricted binaural additivity has to be rejected.

Fit of the Limited Binaural Additivity Model

Predicted mean values. A comparison of the mean values of c and d (or d and e , respectively; see Table 4) with the limits of the tolerance field generated by the LBA model (see Table 2) shows that the LBA model fits our data reasonably well; only three values lie slightly outside the predicted range (by .2, .2, and .7 dB, respectively).

Predicted differences. The predicted deviations ($c' > d'$; $d' > e'$) show up under all experimental conditions. The magnitude of the differences $d' - e'$ lie in the expected range, whereas the differences $c' - d'$ turn out to be smaller than predicted. More important, however, is that the direction of the differences is significantly in favor of the predicted direction: c' is greater than d' in 33 of the 48 cases, $\chi^2(1) = 6.75$, $p < .01$, d' is greater than e' in 28 of 48 cases (ns). Taking the significant z values of Table 3 renders a similar result: Of the 46 significant differences found, 38 are in the expected direction.

Effect of Experimental Conditions

Effect on deviations from binaural additivity. The bottom line of Table 3 shows mean absolute z values (absolute because additivity is violated by deviations in any direction). Analysis of variance (ANOVA) for the absolute z values gives a significant result only for intensity level on the z values obtained when testing commutativity, $F(1, 33) = 6.47$, $p < .025$. Binaural additivity is violated more under higher intensity) than under lower intensity. For both axioms, interactions of sim-

Table 4
Effects of Frequency and Mean Intensity Level

Intensity level (dB)	200 Hz	2000 Hz	Total
Cancellation			
70			
c	52.8	52.7	52.7
d	52.0	52.2	52.1
Δ	+ .8	+ .5	+ .6
90			
c	54.2	52.1	53.1
d	53.2	51.7	52.5
Δ	+1.0	+ .4	+ .6
Total			
c	53.5	52.4	52.9
d	52.6	51.9	52.3
Δ	+ .9	+ .5	+ .6
Commutativity			
70			
d	50.5	50.6	50.5
e	48.8	49.9	49.4
Δ	+1.7	+ .7	+1.1
90			
d	52.4	50.3	51.3
e	51.5	48.3	49.9
Δ	+ .9	+2.0	+1.4
Total			
d	51.4	50.4	50.9
e	50.2	49.1	49.6
Δ	+1.2	+1.3	+1.3

Note. Reference levels are 70 and 90 dB SPL; values for 90 dB have been corrected by -20 dB to ensure comparability with 70 dB condition. Values are means for c , d and d , e .

ilar form are discernible (but not significant because of the considerable amount of inter-individual variance components).

Effect of mean values. Table 4 shows the mean values for the c 's and d 's, and for the d 's and e 's under each condition. An ANOVA (randomized block design) yields significant results for all interactions with the exception of c 's. This result is due mainly to the condition 200 Hz/90 dB. As for the differences $c' - d'$, and $d' - e'$, an ANOVA does not show any significant main or interaction effects, even for interaction on $d' - e'$: $F(1, 33) = 3.07$ (all $\alpha = 5\%$).

Individual Differences

The same ANOVA can be used to test the effect of the "blocks" (i.e., our listeners). For

all the medians except the d' of commutativity, and for the differences $c' - d'$ and $d' - e'$, significant F values were obtained ($\alpha = 5\%$). This result is in accordance with the considerable interindividual variation in the z values reported previously.

Discussion

Binaural Additivity Rejected

In the majority of cases (68.75%), at least one of the two conditions necessary for additivity is violated. Only one subject in 12 fulfills the requirements of additivity under all four experimental conditions. This fact clearly refutes the additivity hypothesis as a general rule of binaural integration. Moreover, if a specific case shows both the property of cancellation and of commutativity, the existence of additivity is not guaranteed because cancellation as well as commutativity are necessary but not sufficient conditions for additivity (cf. the critique of Levelt et al.'s, 1972, study).

Our data contain considerable individual variation, not only in absolute values chosen in the matching task but presumably also in the rules of binaural loudness combination that may be inferred from the z values of Table 3. This finding parallels those of Krantz and Tversky (1975) and other investigators who collected evidence for individually differing combination rules in visual judgments (cf. Gigerenzer, 1981, for a general discussion).

A possible objection to our findings might be that the listeners in our experiment, despite their training, tended to make systematic errors in the matching task. The first possible type of matching bias would have occurred if the tone to be adjusted by the listener had always initially been too low or too high in intensity. The settings would then have been systematically too high or too low, respectively; but, because the initial intensity of the stimulus to be adjusted was randomized, this kind of systematic error can easily be ruled out.

A further source of errors might exist. Our listeners might have shown constant individual matching biases, together with random error variance. Although the random portion of such errors is somewhat accounted for by

taking medians from many trials, no such provision could be made for systematic biases. Considering that of the four medians relevant for testing additivity, three are based on trials in which one of the fixed stimuli actually was a median obtained in an earlier step of the experiment, errors could have been compounded from one experimental step to the next. This would enlarge overall variance between subjects in all steps of our experiment that rest on a previously determined median. Data from the test of cancellation provide an excellent means for checking this possibility. Overall variance from Step 2 of the test of cancellation (in which a previously established median, b' , is involved) was compared with the variance of d' from Step 3 (in which judgments are not based on a median but rest on stimuli that were fixed for all listeners). If systematic errors had been compounded, the former variance should have been markedly higher than the latter. This, however, holds true under only two of the four experimental conditions, $F(240, 240) = 1.426$ and 1.621 , $p < .01$, in both cases, whereas in both of the 200-Hz conditions, $F(240, 240) = 1.186$ and 1.079 , which are far from reaching significance. The possibility of compounding errors can thus be ruled out in the 200-Hz conditions. Nevertheless, more cases against the axiom of cancellation were found for the 200-Hz than for the 2-kHz conditions (12 against 6 significant values of z in Table 3). We concluded that the failure of strict binaural additivity certainly cannot be due to the design of our experiment.

Status of the LBA Model

First, it should be noted that in terms of severity (Popper, 1963), our test of the LBA model is stronger than the test that the additive model had to stand (and failed). Apart from postulating a general property of binaural loudness construction, the LBA model has been formulated in decibel units and is therefore much more readily refuted by empirical test. If the LBA model had been specified in subjective units, it would be more difficult to reject.

Second, the LBA model stands for a class of possible nonadditive models and is simplified in certain respects. It postulates partial

contralateral binaural inhibition as a central feature of binaural loudness construction. But it does not account for sensitivity differences between the left and right ears, and has made no provision for separate stages of loudness construction (as have been postulated, for instance, by Treisman & Irwin, 1967). In a more general formulation, allowing for different subjective scales f and g , and a threshold δ for dominance, binaural loudness $L(a, p)$ of a tone pair (a, p) could be given as the following:

$$L(a, p) = \begin{cases} f(a) & \text{if } f(a) - g(p) \geq 2, \\ g(p) & \text{if } g(p) - f(a) \geq 2, \\ (f(a) + g(p))/2 + \delta, & \text{otherwise.} \end{cases} \quad (11)$$

Third, the LBA model proves surprisingly successful even in its simplified form. Not only is it able to fit the data of Irwin (1965) and Keen (1972) reasonably well but it gives fairly good predictions for our experimental outcomes. Also, it is compatible with the results reported by Falmagne et al. (1979, Experiment 1). These authors assumed additivity as the basic characteristic and tested three specific versions of additive models: the logarithmic, linear, and exponential difference probabilistic additive conjoint measurement (PACOME) model. These three models make different predictions for the probability $P(ap, bp)$ that a tone pair (a, p) will be judged louder than a pair (b, p) . The logarithmic PACOME predicts that for fixed a and b , and increasing p , the value of $P(ap, bp)$ will decrease toward .5, whereas the linear PACOME postulates that $P(ap, bp)$ will not change. The exponential PACOME even predicts an increasing value of $P(ap, bp)$ under these circumstances. Falmagne et al.'s results show that $P(ap, bp)$ decreases toward a chance value of .5 with increasing p . Falmagne et al. (1979) are careful to take these data as confirmation for the logarithmic difference PACOME model only "if the general framework of the . . . model is retained" (p. 32).

Other models, additive or not, might of course be used to explain such data. Table 5 shows the behavior of the LBA model for Falmagne et al.'s experiment, which is a stepwise decrease of binaural difference towards

Table 5
Predictions of the Limited Binaural Additivity Model for Falmagne et al.'s (1979) Conjoint Weber's Law

a	Binaural pair (dB)		p	Predicted difference (dB)
	p	b		
49	34	47	34	+2
49	38	47	38	+1
49	42	47	42	+1
49	46	47	46	+1
49	50	47	50	+1
49	54	47	54	+1
49	58	47	58	+1
49	62	47	62	0
49	66	47	66	0
49	70	47	70	0

Note. The predictions refer to the experimental design of Falmagne et al. (1979, Experiment 1). Values for a , b , and p are taken from Falmagne et al.; the predictions of the LBA model are based on $\delta = 6$ dB. The difference between (a, p) and (b, p) decreases to zero when p increases. This corresponds to a decrease of $P(ap, bp)$ toward .5.

zero (equivalent, of course, to a decrease of $P[ap, bp]$ towards .5) with increasing p .

Matching and Forced-Choice Paradigms

The intensity-matching paradigm used by Falmagne (1976) and in the present study is not the only way to test for additivity of binaural loudness combination. Levelt et al. (1972) and Falmagne et al. (1979) used paired comparisons in a forced-choice paradigm. Can these two experimental tasks be considered as equivalent? Falmagne (1979) has resolved this issue by showing that under conditions normally fulfilled, "the comparison paradigm can be used to mimic the matching paradigm through a Stochastic Approximation technique" (p. 79). Therefore, the refutation of the model of strict binaural additivity of loudness (Equation 5) implies rejection of strict additivity in general and is not confined to a matching task.

Conversely, predictions may be derived from the PACOME models of Falmagne et al. (1979) that permit direct comparison with data from matching-type experiments. In contrast to the equiloudness contours derived from the LBA model (Figure 1), all strictly

additive models result in strictly monotone equi-loudness functions.

Although attempts have been made to "linearize" equi-loudness functions through numerous transformations (Treisman & Irwin, 1967), a fundamental difference remains between models that suggest limits to binaural additivity and those that do not. The present results indicate, in our opinion, that limits to binaural additivity of loudness exist and should be recognized as such. The quest for a more refined model of limited binaural additivity possibly constitutes a fruitful path for further study of sensory integration of binaural loudness.

References

- Anderson, N. H. *Foundations of information integration theory*. New York: Academic Press, 1981.
- Arbuckle, J., & Larimer, J. The number of two-way tables satisfying certain additivity axioms. *Journal of Mathematical Psychology*, 1976, 13, 89-100.
- Békésy, G. von. Zur Theorie des Hörens. Über die kaum merklichen Amplituden- und Frequenzänderungen eines Tones. *Physik*, 1929, 30, 721-745.
- Durlach, N. I., & Colburn, H. S. Binaural phenomena. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 4). New York: Academic Press, 1978.
- Falmagne, J. C. Random conjoint measurement and loudness summation. *Psychological Review*, 1976, 83, 65-79.
- Falmagne, J. C. On a class of probabilistic conjoint measurement models: Some diagnostic properties. *Journal of Mathematical Psychology*, 1979, 19, 73-88.
- Falmagne, J. C., Iverson, G., & Marcovici, S. Binaural loudness & summation: Probabilistic theory & data. *Psychological Review*, 1979, 86, 25-43.
- Fletcher, H., & Munson, W. A. Loudness: Its definition, measurement, and calculation. *Journal of the Acoustical Society of America*, 1933, 5, 82-108.
- Gigerenzer, G. *Messung und Modellbildung in der Psychologie*. Munich: Reinhardt, 1981.
- Hellman, R. P., & Zwislocki, J. Monaural loudness function at 1000 cps and interaural summation. *Journal of the Acoustical Society of America*, 1963, 35, 856-865.
- Irwin, R. J. Binaural summation of thermal noises of equal and unequal power in each ear. *American Journal of Psychology*, 1965, 78, 57-65.
- Keen, K. Preservation of constant loudness with interaural amplitude asymmetry. *Journal of the Acoustical Society of America*, 1972, 52, 1193-1196.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement* (Vol. 1). New York: Academic Press, 1971.
- Krantz, D. H., & Tversky, A. Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology*, 1975, 12, 4-34.
- Levelt, W. J. M., Riemersma, J. B., & Bunt, A. A. Binaural additivity of loudness. *British Journal of Mathematical and Statistical Psychology*, 1972, 25, 51-68.
- Lienert, G. A. *Verteilungsfreie Methoden in der Biostatistik* (Band 1). Meisenheim am Glan, West Germany: Hain, 1973.
- Marks, L. E. Binaural summation of the loudness of pure tones. *Journal of the Acoustical Society of America*, 1978, 64, 107-113.
- McClelland, G. H. A note on Arbuckle and Larimer, "The number of two-way tables satisfying certain additivity axioms." *Journal of Mathematical Psychology*, 1977, 15, 292-295.
- Popper, K. R. *Conjectures and refutations*. London: Routledge & Kegan Paul, 1963.
- Reynolds, G. S., & Stevens, S. S. Binaural summation of loudness. *Journal of the Acoustical Society of America*, 1960, 32, 1337-1344.
- Scharf, B. Dichotic summation of loudness. *Journal of the Acoustical Society of America*, 1969, 45, 1193-1205.
- Scharf, B., & Fishken, D. Binaural summation of loudness: Reconsidered. *Journal of Experimental Psychology*, 1970, 86, 374-379.
- Schönemann, P. H., Cafferty, T., & Rotton, J. A note on "additive functional measurement." *Psychological Review*, 1973, 80, 85-87.
- Stevens, S. S. *Psychophysics*. New York: Wiley, 1975.
- Strube, G. *Mikrocomputer in der Psychologie*. Stuttgart: Klett, 1981.
- Strube, G. Binaural processing of temporal and intensity cues: Suggestions for a one-system model. *Psychological Research*, 1982, 44, 45-49.
- Treisman, M., & Irwin, R. J. Auditory intensity discriminable scale. I: Evidence derived from binaural intensity summation. *Journal of the Acoustical Society of America*, 1967, 42, 586-592.
- Urbantschitsch, V. Über Wechselwirkungen beider Gehörorgane. *Archiv für Ohrenheilkunde*, 1893, 35, 1-2.
- Wilkening, F. W. Combining of stimulus dimensions in children's and adult's judgments of area: An information integration analysis. *Developmental Psychology*, 1979, 15, 25-33.

Received February 10, 1982

Revision received May 10, 1982 ■