BINAURAL ADDITIVITY OF LOUDNESS

By W. J. M. Levelt,† J. B. Riemersma‡ and A. A. Bunt‡

Groningen University

A definition of binaural additivity is given in terms of the theory of simultaneous conjoint measurement. Additivity is then tested and verified by a conjoint measurement procedure. Methods for deriving psychophysical scales from such procedures are discussed, and the experimental scales are compared with the usual ratio scales for loudness, derived from extensive measurement such as magnitude estimation. The functions are in good agreement and it is concluded that binaural additivity of loudness holds for non-zero stimulation of the ears.

The long history of experimentation on binaural loudness, which has been well reviewed by several authors (Hirsch, 1948; Reynolds & Stevens, 1960; Treisman & Irwin, 1967; Sharf, 1969), shows that the main procedure in such studies has been some comparison of monaural and binaural stimulation. Comparisons have been made, for example, between monaural and binaural thresholds (Urbantschitsch, 1893), monaural and binaural loudness of suprathreshold stimuli (Békésy, 1929; Fletcher & Munson, 1933; Reynolds & Stevens, 1960; Irwin, 1965, and several others), or monaural and binaural loudness functions (Fletcher & Munson, 1933; Reynolds & Stevens, 1960; Hellman & Zwislocki, 1963).

The monaural–binaural paradigm is essential for answering the question of whether left- and right-ear stimulation is summative to some extent, i.e. whether the effect of single stimulation of one ear adds (partly or entirely) to the effect of single stimulation of the other ear when the ears are stimulated simultaneously. A comparison with the related case of binocular brightness cautions against a too rapid conclusion that summation exists. A field viewed with one eye hardly differs in brightness from the same field viewed with both eyes: binocular summation of brightness is small or non-existent (Levelt, 1968).

Though there is no doubt from the literature that binaural summation of loudness exists to some extent, the strong initial statements about complete summation of either energies (Urbantschitsch, 1893; Békésy, 1929) or loudnesses (Fletcher & Munson, 1933; Stevens & Davis, 1938) have not been sufficiently substantiated. In fact, the energy-summation hypothesis was abandoned at an early stage (Fletcher & Munson, 1933; Stevens & Davis, 1938), whereas the loudness-summation hypothesis remained experimentally undecided (cf. Reynolds & Stevens, 1960; Hellman & Zwislocki, 1963) and led to a considerable theoretical

---

† Present address: Psychological Laboratory, Nijmegen University, Nijmegen, The Netherlands.

‡ Present address: Institute for Perception RVO-TNO, Soesterberg, The Netherlands.

controversy about the definition of loudness (Fletcher & Munson, 1933; Reynolds & Stevens, 1960). One solution to this controversy was offered by Treisman & Irwin (1967), who proposed a processing model for binaural loudness matching. The present paper proposes a different solution. First, it will be shown that the Fletcher–Reynolds controversy, which is discussed in the next section, can be studied without the use of a monaural–binaural paradigm, i.e. without a test of *summation*. Instead it suffices to test *additivity* of binaural stimulation. Second, using the results of such an experiment, the Treisman & Irwin model is simplified in such a way that the Fletcher–Reynolds controversy is in principle resolved, although the domain of the model is then restricted to binaural stimulation. Finally, arguments are put forward to show that this latter restriction should always be made.

1. THE FLETCHER–REYNOLDS CONTROVERSY

Fletcher & Munson (1933) simply *assumed* that a binaural stimulus sounds twice as loud as a monaural stimulus. Because it was obvious that this relation did not hold if loudness was measured in decibels, the authors consequently decided to rescale sound pressure in such a way that full summation would hold for subjective loudness as expressed on this psychological scale. So, by assuming full summation and by the use of a monaural–binaural experimental paradigm, a subjective loudness function could be derived.

Reynolds & Stevens (1960) reacted to this procedure as follows: 'That function, as it turned out, was not the simple power function derived from later experiment and recently recommended by the International Standards Organization.'

The first part of this statement indicates a disparity between the type of function measured by Fletcher's procedure and the usual power function obtained by other methods, such as magnitude estimation. If such a discrepancy holds in further experimentation, a theory is needed to explain the difference in results. In fact, Treisman & Irwin (1967) proposed such a theory, which will be discussed in the next section.

The last part of the statement seems to suggest that it is up to the International Standards Organization to decide what the 'real' loudness function is. This seems to be a dogmatic approach which cannot solve the empirical problems involved. The present authors will limit their argument to processing models, and leave the question of the definition of loudness aside.

It should be noticed that Reynolds & Stevens indeed gave important experimental evidence for their conclusion that if loudness is a power function of SPL, binaural loudness is not the sum of monaural loudnesses. The main arguments are:

1. From various direct scaling procedures it appeared that the exponent for binaural loudness differs from the exponent for monaural loudness. However, if simple binaural summation holds, this should not be the case. This can easily be seen in the following way. If the left-ear loudness is a power function of SPL, we can write it as $\psi = k_l \phi_l^n$, where ϕ stands for effective sound-pressure level at the left ear and k is a constant for that ear. Similarly, for the right ear

we have $\psi_r = k_r \phi_r{}^n$, on the assumption that the exponents are equal for both ears. In the situation of binaural presentation, it is assumed that effective SPL's are equal: $\phi_l = \phi_r = \phi$. If there is full loudness summation, the resulting binaural loudness can be written as $\psi_b = \psi_l + \psi_r = (k_l + k_r) \phi^n$. This differs from monaural loudness only in constant, not in exponent. Thus monaural and binaural loudness should have a constant ratio, i.e. loudness functions should be parallel on a log–log plot.

Hellman & Zwislocki (1963) showed that this difference in slope may well be less dramatic than Reynolds & Stevens concluded. They did not impose a power function on their data, and took into account differences in ear sensitivity in deriving a theoretical curve for binaural summation. The fit of this curve to actual binaural data was near to perfect.

2. Monaural–binaural loudness matching should give a straight line with slope 1·0 if monaural and binaural exponents are equal. If for the monaural situation we have $\psi_m = k_m \phi_m{}^n$, and similarly for the binaural loudness $\psi_b = k_b \phi_b{}^n$, then for the matching situation it should be the case that $\psi_m = \psi_b$, thus $k_m \phi_m{}^n = k_b \phi_b{}^n$. It follows that $\log k_m + n \log \phi_m = \log k_b + n \log \phi_b$, thus $\log \phi_m = \log \phi_b + n^{-1}(\log k_b - \log k_m)$, which implies a slope of 1·0. Data from several publications show smaller slopes, indicating that monaural and binaural exponents differ, i.e. that complete summation is not present. Again, Hellman & Zwislocki showed that this difference of slope, if it exists (especially for low sound-pressure levels), is not conclusive. From the same arguments, given above, they derive a non-linear matching curve if full summation is assumed. The actual data, including Fletcher & Munson's, are in close correspondence with the theoretical curve. It is not evident that later findings, especially Irwin's (1965) matching data, should qualify this conclusion.

2. THE TREISMAN & IRWIN MODEL

A processing model which can, in principle, take account of a possible empirical issue involved in the Fletcher–Reynolds controversy was proposed by Treisman & Irwin (1967). They distinguished between peripheral and central components of loudness processing. In the peripheral part, the physical input (i.e. stimulus intensity, I) causes a *discriminal effect*, E, $E = f(I)$. It is assumed that the discriminal effects, E_l and E_r, produced via left and right ear summate in producing a total or binaural discriminal effect, E_b. In the central part of the system some sort of 'metric processing' transforms the total discriminal effect into a total loudness, L_b, $L_b = g(E_b)$.

This model, however, allows for full summation, not of loudnesses, but of discriminal effects, $E_b = E_l + E_r$, and at the same time for power-type loudness function. In the loudness function, $L = g(f(I))$, neither f nor g need be simple power functions for their combination to have that form.

In their paper the authors derive a solution for f, the discriminal function. Briefly, their derivation is as follows. They start from the above-mentioned observation (2) that monaural–binaural loudness-matching data, especially Irwin's own observations, depart from slope 1·0. From this they conclude that

f cannot be a simple power function. For if the binaural pair (I_l, I_r) matches the monaural stimulus, I_m, i.e. $(I_m, 0)$ or $(0, I_m)$, then

$$E_b = E_l + E_r = f(I_l) + f(I_r) = f(I_m).$$

If $E = kI^n$, then $k(I_l^n + I_r^n) = k(I_m^n)$. In the monaural–binaural paradigm, $I_l = I_r = I_b$; therefore $2I_b^n = I_m^n$ or $\log 2 + n \log I_b = n \log I_m$. This implies the empirically rejected slope of $1 \cdot 0$ if I_b is plotted against I_m on decibel scales. The authors are therefore led to reject a simple power function for f. They then proceed to propose a power function for f, in which the exponent n decreases for increasing stimulus intensity. The problem then reduces to determining the function relating I to n.

At this point a method is introduced which differs from the usual monaural–binaural procedure. This is the basis of Irwin's equi-loudness data (Irwin, 1965). An equi-loudness curve is a set of stimulus pairs (I_l, I_r), which all produce the same binaural loudness. For such data it should be the case that $E_b = E_l + E_r = C$ for each given equi-loudness curve. Appropriate transformations of I_l and I_r into E_l and E_r should then give transformed equi-loudness curves that are straight lines with slope -1. They show that this is feasible if the exponent function is $n = \log 2 / \{(1 - b) \log I + \log a\}$, where a is the monaural–binaural matching ratio at threshold and b a free parameter.

3. Additivity and Summation

It should be clear from Treisman & Irwin's procedure that if binaural summation is assumed, data other than those from the usual monaural–binaural procedure can be used to investigate the type(s) of psychophysical function(s) needed in a model for binaural loudness processing. In the following we will in fact limit consideration exclusively to binaural–binaural data. Though it is obvious that such data cannot settle the summation issue, they can be used to test one of the necessary conditions for summation, namely additivity. If such a test is positive, then further conclusions can be drawn with respect to psychophysical functions in a model of binaural loudness processing.

The relation between additivity and summation can be explicated by the following definitions.

$L = \{a, b, c, ...\}$ is a finite set of m left-ear stimuli differing in intensity and $R = \{p, q, r, ...\}$ is a similar finite set of n right-ear stimuli. Let $L \times R$ denote the set of all stimulus pairs (i, j), where $i \in L$ and $j \in R$.

B is an $m \times n$ binaural loudness matrix, where b_{ij} is a measure of the binaural loudness of the stimulus pair (i, j).

Definition I. The binaural loudness matrix **B** is *additive* if and only if there are real-valued functions ψ_l, ψ_r and β defined over L, R and **B**, respectively, such that

(i) $\beta(a, p) = \psi_l(a) + \psi_r(p)$,

(ii) $\beta(a, p) \geqslant \beta(b, q)$ if and only if $b_{ap} \geqslant b_{bq}$,

for all a, $b \in L$ and all p, $q \in R$.

This means that we are able to rescale left and right stimulus intensities by functions such that the order of magnitudes of their joint binaural effects corresponds to the order of the sums of their scale values. In other words, if left and right stimulus intensities are correctly rescaled they have no *interaction* in their effect on binaural loudness.

Condition (ii) in Definition I implies that, in order to test binaural additivity, we do not need actual measures of binaural loudness, but only a rank order over binaural loudnesses.

What exactly is meant by full binaural summation? Treisman & Irwin take the meaning of summation to be 'that the resultant produced by summation of the effects arising in each ear is given by simple addition of those effects'. If we interpret the 'resultant' as the value of β and the 'effects' in each ear as the values of ψ_l and ψ_r, then this definition is essentially Definition I, i.e. additivity. But we have seen that the Fletcher–Reynolds controversy was based on a further interpretation of 'summation', namely that 'a binaural stimulus always sounds twice as loud as a monaural stimulus of the same SPL' (Reynolds & Stevens, 1960). Apart from the implicit assumption of equal effectiveness of both ears, this latter interpretation seems perfectly natural to us. In a slightly more generalized form, and in terms of ψ and β, this becomes

$$\beta(a, p) = \beta(a, \varnothing) + \beta(\varnothing, p),$$

where $\varnothing$ is the null stimulus. If, at the same time, we want to keep the additivity characteristic (see Definition I) in our definition of summation, then the latter equation follows from additivity if $\varnothing \in L, \varnothing \in R$. The equation can then be rewritten as

$$\psi_l(a) + \psi_r(p) = \psi_l(a) + \psi_r(\varnothing) + \psi_l(\varnothing) + \psi_r(p),$$

which is true if $\psi_r(\varnothing) + \psi_l(\varnothing) = 0$. If additivity holds, this latter condition can always be fulfilled, because any solution for ψ_l and ψ_r shows so-called 'Archimedian uniqueness': if ψ_l, ψ_r is a solution, then $a\psi_l + b, a\psi_r + c$ is also a solution (for $a > 0$). It should therefore be possible to find additive constants b and c, such that $\psi_r(\varnothing) + \psi_l(\varnothing) = 0$. This leads naturally to the following definition of summation:

Definition II. The binaural loudness matrix **B** is *summational* if and only if
   (i) **B** is additive,
   (ii) $\varnothing \in L, \varnothing \in R$, where $\varnothing$ is the null stimulus.
In other words, if additivity is preserved in the case of monaural–binaural comparison, we have binaural summation. Note, however, that the above definitions define neither binaural summation of *loudness* nor additivity of *loudness*. In order to use these terms one must agree that ψ_l, ψ_r and β are loudness functions. This can either be decided by definition (Fletcher & Munson) or by comparison with other measurements of loudness (Reynolds & Stevens in fact rejected the idea of loudness summation on this basis). Treisman & Irwin presented evidence for binaural summation, but rejected binaural summation of loudness. In their case the values of ψ_l, ψ_r and β are *discriminal effects*, not *loudnesses*.

Though one can define loudness in one way or another, the question of additivity and therefore summation is an empirical one. Fletcher & Munson *assumed* additivity without testing it. Reynolds & Stevens rejected binaural summation of *loudness*, but had no grounds for rejecting binaural summation in general.

The above definitions are inspired by the theory of conjoint measurement (Luce & Tukey, 1964; Krantz *et al.*, 1971). The essence of this approach is not to measure several variables independently in order to test whether some specified functional relation holds among them, but rather whether scales of measurement can be found for the variables (e.g. left, right and binaural intensity) which satisfy a certain composition rule (e.g. an additive rule). Only ordinal analysis is required to prove the existence of such scales. If they do exist one can further investigate their character (e.g. how they compare to loudness scales obtained by other procedures).

4. An Experimental Test of Binaural Additivity

4.1. *Procedure*

It appeared from Definition I that a test of binaural additivity requires a complete rank order of the set of left–right stimulus pairs in terms of loudness. The ideal procedure for deriving such a rank order would be paired comparisons among all binaural stimuli. Starting from six left-ear and six right-ear sound-pressure levels, i.e. from 36 binaural stimuli, such a test would require 36^2 paired comparisons. Moreover, it would be necessary for the control of order effects to present each pair in both successions. This would double the number of comparisons to 2592, which is not feasible for a single subject. It would, however, be a questionable procedure to use several subjects for the determination of one rank order of loudnesses. We therefore decided to reduce the number of paired comparisons per subject in two ways. (1) The 36 comparisons of the stimuli with themselves were left out. (2) We assumed an increasing monotonic relation between stimulus and binaural loudness as follows: If, for any pair of left-ear stimuli a and b, it is the case that $a < b$, then
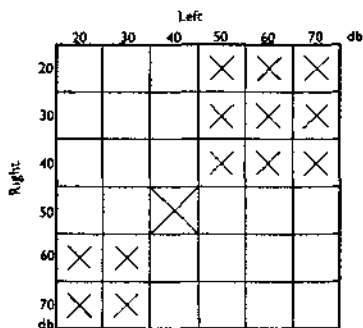

Fig. 1.—Paired comparisons for an exemplary stimulus pair (40 db, 50 db). It is compared with all binaural stimuli marked by × .

$b_{ap} < b_{bp}$, for all $p \in R$. Also, if $p < q$, for $p, q \in R$, then $b_{ap} < b_{aq}$, for all $a \in L$. So, for instance, we assumed that the pair (20 db, 40 db) would be less loud than the pair (30 db, 40 db).

Fig. 1 shows the type of comparisons made in the experiment. For the binaural stimulus (40, 50) taken as an example, only comparisons with the binaural stimuli marked by × were made. Together with the restriction under (1), this assumption enables the number of necessary pair comparisons made by a subject to be reduced to 450, including two orders of presentation for each pair.

It should be pointed out that this assumption does not imply additivity. A counter-example is sufficient to show this. For the **B** matrix in Table 1, where

Table 1.—A Data Matrix which is Monotonic but not Additive

		Right-ear stimuli	
B	p	q	r
a	1	3	4
b	2	5	8
c	6	7	9

Left-ear stimuli: a, b, c

$a < b < c$ and $p < q < r$, the assumed monotonic relation holds: $b_{ai} < b_{bi} < b_{ci}$ for $i = p, q, r$ and $b_{ip} < b_{iq} < b_{ir}$ for $i = a, b, c$. But there is no additivity. This can easily be shown by comparing the following inequalities:

   (i) $b_{aq} > b_{bp}$ (3 > 2),

   (ii) $b_{br} > b_{cq}$ (8 > 7),

   (iii) $b_{ar} < b_{cp}$ (4 < 6).

In the case of additivity (see Definition I): (i) implies that $\beta(a, q) > \beta(b, p)$, and therefore that $\psi_l(a) + \psi_r(q) > \psi_l(b) + \psi_r(p)$. Similarly, (ii) implies that

$$\psi_l(b) + \psi_r(r) > \psi_l(c) + \psi_r(q).$$

Addition of the latter two inequalities yields:

$$\psi_l(a) + \psi_l(b) + \psi_r(q) + \psi_r(r) > \psi_l(b) + \psi_l(c) + \psi_r(p) + \psi_r(q).$$

By *double cancellation*, i.e. cancellation of the two common left and right terms, this reduces to $\psi_l(a) + \psi_r(r) > \psi_l(c) + \psi_r(p)$. By Definition I this implies that $\beta(a, r) > \beta(c, p)$ and thus that $b_{ar} > b_{cp}$. This, however, is in contradiction to (iii).

More generally, and by a similar argument, it can be shown that if

   (i) $b_{aq} \geqslant b_{bp}$ and

   (ii) $b_{br} \geqslant b_{cq}$, then

   (iii) $b_{ar} \geqslant b_{cp}$.

This property of additive systems is called *double cancellation*. Krantz *et al.* (1971) show that, for all practical purposes, double cancellation is the key property for monotonic two-factor systems. If double cancellation is not

violated, the system is additive. A test of the double-cancellation property will be our main empirical check for the existence of binaural additivity.

The reduced test series took about $2\frac{1}{2}$ hr. The stimuli were 1000 cps pure tones. There were three stimulus conditions: 'soft', 'medium' and 'loud'. In the 'soft' condition the left and right sound-pressure levels used were 0, 10, 20, 30, 40 and 50 db. In the 'medium' condition, they were 20, 30, 40, 50, 60 and 70 db, whereas the 'loud' condition contained the levels 30, 40, 50, 60, 70 and 80 db. Each condition was studied in a separate test series.

Two subjects, J. R. and A. B., participated in the experiment. Both were tested on all three conditions. Subject A. B. did four replications of the whole experiment, i.e. 12 test series, and J. R. did the experiment twice, i.e. six series. The resulting 18 data matrices formed the basic material for all further analyses. Also, both subjects were tested for their absolute thresholds by a forced choice procedure. They had to choose in which of two $\frac{1}{2}$-sec. intervals (marked by flash lights) the 1000 cps test tone had been presented. The following sound-pressure levels led to 75 per cent correct response: subject A. B., left ear -8 db, right ear -1 db; subject J. R., left ear -6 db, right ear -4 db.

4.2. *Equipment*

The 450 stimulus pairs were recorded on the two tracks of a high-quality tape-recorder (type Revox). The temporal pattern of each comparison was: stimulus pair 1, 3 sec.; stimulus pair 2, 3 sec.; silence, 10 sec.

The stimuli were produced by an oscillator (Hewlett Packard), feeding into a system of two attenuators, two amplifiers and a time switch. The earphones (type Beyer) were calibrated on an artificial ear. Their input current was measured by means of a set of Audiovolt meters (type Bruehl & Kjaer). These meters were both used to set the stimulus levels at the tape-recording, and later to check the stimulus levels at the recorder output. For the 0–80 db stimulus range, recorder non-linearity turned out to be negligible.

The subject was seated in a sound-proof room. After each sequence of two stimulus pairs he noted down whether the first or the second pair was the louder one.

4.3. *Tests of Additivity*

For each of the 18 test series, the data were first transformed into a loudness rank order over the 36 binaural pairs, in the following way. For each binaural pair we counted the number of times the subject had actually judged it as 'louder'. Also, we computed the number of stimuli dominated by the binaural pair on the basis of the monotonicity assumption (see Section 4.1). Multiplied by 2,† this latter number was added to the former, yielding a loudness score for the particular binaural pair. The loudness scores of the 36 stimuli were then

---

† As noted in each series, each comparison was made twice: once for each of the two stimulus successions.

replaced by their rank orders. Now and then, ties of two occurred, and were handled in the usual way by giving the average rank number to both. In this way, 18 binaural loudness matrices resulted. These are 6×6 matrices, with rank numbers as entries. If rows and columns are ordered according to left and right sound-pressure level, respectively, these **B** matrices are necessarily monotonic, i.e. the rank numbers increase with sound-pressure level both along rows and columns.

We have noted above that, for a monotonic matrix, *double cancellation* is the key property of additivity. From the definition of double cancellation it should be clear that for one test of this property, three left-ear and three right-ear sound-pressure levels are involved (respectively a, b, c and p, q, r). However, it is not the case that each pair of left and right stimulus triples yields a test. The cancellation property consists of two antecedents ($b_{aq} \geqslant (\leqslant) b_{bp}$ and $b_{br} \geqslant (\leqslant) b_{cq}$) and one conclusion ($b_{ar} \geqslant (\leqslant) b_{cp}$). A test is possible only if both antecedents hold. If not, we have a NO TEST pair of triples. There is no *a priori* means of knowing which pairs of triples allow for a test and which do not. Therefore all pairs of triples should be checked. There are $\binom{6}{3} = 20$ left triples, and similarly 20 right triples. For each **B** matrix we have therefore 400 possible tests of double cancellation. These tests are not independent, since the conclusion of one may be the antecedent of another. For the pairs of triples where the antecedents hold, there are *two possible* outcomes: (1) ACCEPTANCE (the conclusion holds), (2) REJECTION (the conclusion does not hold). In this latter case we should make one further distinction. Because of ties in the ordering of pairs, the case of equality of rank numbers occurs now and then. This may lead to a very weak

Table 2.—Tests of Double Cancellation

Conditions	Type of outcome	Subject J. R. replications		% of tests	Subject A. B. replications				% of tests
		1	2		1	2	3	4	
Soft	ACCEPTANCE	241	247	99·8	228	219	238	229	97·5
(0–50 db)	REJECTION	1	0	0·2	1	3	9	8	2·5
	STRONG	0	0	0·0	0	0	5	5	1·1
	WEAK	1	0	0·2	1	3	4	3	1·4
	NO TEST	158	153		171	178	153	163	
Medium	ACCEPTANCE	263	235	97·4	247	238	250	269	98·0
(20–70 db)	REJECTION	8	5	2·6	6	6	6	2	2·0
	STRONG	4	4	1·6	5	0	5	1	1·1
	WEAK	4	1	1·0	1	6	1	1	0·9
	NO TEST	129	160		147	156	144	129	
Loud	ACCEPTANCE	290	290	99·8	235	224	236	226	98·5
(30–80 db)	REJECTION	0	1	0·2	4	1	5	4	1·5
	STRONG	0	0	0·2	0	1	0	0	0·1
	WEAK	0	1	0·0	4	0	5	4	1·4
	NO TEST	110	109		161	175	159	170	

form of rejection, namely antecedents: $b_{aq} = b_{bp}$, $b_{br} > b_{aq}$; conclusion: $b_{ar} \geqslant b_{cp}$; actual data: $b_{ar} < b_{cp}$. With more extensive data, the chance for ties would be smaller, so that the equality in the antecedents would really turn out to be an inequality. In one case $(b_{aq} < b_{bp})$ this would result in a NO TEST situation, and only in the other case $(b_{aq} > b_{bp})$ in a rejection. In what follows we will therefore take special note of WEAK rejections due to an equality in the antecedents.

A special computer program was written† to check the 400 possible tests for each of the 18 data matrices.

Table 2 summarizes the results. Of all TEST cases, only 1·4 per cent are cases of rejection. Most of these are weak: only 0·5 per cent are cases of strong rejection. The 18 data matrices strongly show the double cancellation property, which leads to the conclusion that, for our subjects and conditions, *binaural additivity exists.*

5. THE CONSTRUCTION OF PSYCHOPHYSICAL SCALES FROM CONJOINT LOUDNESS MEASUREMENTS

5.1. *Derivation of Scales from the Paired Comparison Data*

The existence of additivity implies that there are scales of measurement for left, right and binaural loudness (ψ_l, ψ_r, β), which satisfy an additive composition rule: $\beta(a, p) = \psi_l(a) + \psi_r(p)$—see Definition I.

It was decided to compute the best fitting scales for each condition (soft, medium, loud) and each subject separately. We assumed that the scales for a subject would be constant over replications within a condition.

The computer program ADDIT (Roskam‡ & van Gilst, 1967) was employed for this purpose. The input was the 6×6 data matrices obtained from a subject for a given condition, and the output took the form of interval scale values for the six left-ear and the six right-ear stimuli. These scale values have the property that their pairwise sums are optimally rank ordered with respect to the experimentally obtained rank order of the 36 binaural pairs. This is realized by minimizing stress (S) in Kruskal's (1964) sense.§

Table 3 shows the stresses for the six solutions (2 subjects × 3 conditions). In view of the good results of the double cancellation test, it is not surprising to find that the average stress is lower than 3 per cent.

---

† The authors are grateful to Mr H. Camstra for writing this program.

‡ Professor Roskam's contribution to this article greatly exceeds the actual computation of the ψ scales. The main ideas of the present experiment emerged from various discussions the authors had with him in the different stages of experimentation.

§ $S = \sqrt{\left[\dfrac{\sum \{\beta(l, r) - \hat{\beta}(l, r)\}^2}{\sum \{\beta(l, r) - \bar{\beta}\}^2} \right]}$,

where $\beta(l, r) = \psi_l(l) + \psi_r(r)$. $\hat{\beta}(l, r)$ is the average scale value for all stimuli in the same 'block' with (l, r), 'blocks' being defined as minimal sets of stimuli for which the average scale value is non-decreasing with respect to the data. $\bar{\beta}$ is the average scale value.

The output interval scales ψ_l and ψ_r are unique in the sense of Archimedian uniqueness, i.e. if ψ_l, ψ_r is a solution, then $a\psi_l + b$, $a\psi_r + c$ is also a solution (for $a > 0$).

Table 3.—*Stress Values for Additive Solutions*

Subject ...	J. R.	A. B.
Conditions		
Soft	0·01734	0·03784
Medium	0·02774	0·07880
Loud	0·00004	0·01440

Before we report the scale solutions, one further consideration, for which Archimedian uniqueness is essential, should be made. We tried to 'align' the scale solutions for the three conditions (soft, medium, loud) by comparing the scale values in their domains of overlap: the soft and medium conditions overlap for the values 20, 30, 40 and 50 db, the medium and loud conditions for the values 30, 40, 50, 60 and 70 db. By a least squares procedure,† values a, b and c were computed for the transformation $(a\psi_l + b, a\psi_r + c)$ of the scale values of the 'soft' condition in order to get an optimal match with the scale values of the 'medium' condition. Similarly the scale values of the 'loud' condition were transformed in order to match the 'medium' condition.

The transformed scale solutions for the two subjects and three conditions are presented in Table 4, which shows that the scales for the three conditions connect extremely well. It was therefore decided to compute the average scale values over the three conditions as the psychophysical scale values of the respective sound-pressure levels for the two ears. These scales still did not satisfy the convention of non-negativity. In the next section this convention will be adopted in a very natural way. Before we study the character of these scales in relation to models for binaural loudness, we shall describe two other conjoint measurement procedures by which scales can be produced in a more direct way.

5.2. *The Construction of a Dual-standard Sequence*

Luce & Tukey (1964) show that if the variables of an additive system are sufficiently continuous—i.e. satisfy their solution axiom—one can freely construct so-called 'dual-standard sequences'. This notion is easily explained by means of Fig. 2. Fig. 2a shows a theoretical set of equi-loudness curves, each of which is a set of left, right stimulus pairs which all produce the same binaural

---

† If ψ_l and ψ_r are the left- and right-ear scales for the soft condition, $\psi_l{}'$ and $\psi_r{}'$ those for the medium condition, the least squares solution consists in minimizing

$$F = \sum_i \{a\psi_l(i) + b - \psi_l{}'(i)\}^2 + \sum_i \{a\psi_r(i) + c - \psi_r{}'(i)\}^2,$$

where the summation is over the i overlapping values. Setting the partial derivatives of F with respect to a, b and c equal to zero gave the appropriate solution for a, b and c.

loudness. A dual-standard sequence (DSS) is a set of stimulus pairs (l_i, r_j) where the stimuli l and r are chosen and rank ordered in such a way that the pair (l_m, r_n) matches (l_p, r_q), whenever $m+n = p+q$. It is clear from Fig. 2a that

Table 4.—*Separate and Average Left- and Right-ear Scale Values for the Three Conditions*

| Sound-pressure level (db)	Left-ear condition				Right-ear condition			
	Soft	Medium	Loud	ψ (mean)	Soft	Medium	Loud	ψ (mean)
				SUBJECT J. R.				
0	−0·921			−0·921	−1·088			−1·088
10	−0·889			−0·889	−1·048			−1·048
20	−0·805	−0·851		−0·828	−0·973	−0·954		−0·963
30	−0·698	−0·665	−0·602	−0·655	−0·866	−0·881	−0·976	−0·908
40	−0·535	−0·544	−0·333	−0·471	−0·624	−0·604	−0·662	−0·630
50	−0·200	−0·179	−0·020	−0·133	−0·158	−0·181	−0·287	−0·208
60		0·453	0·356	0·404		0·416	0·494	0·455
70		1·786	1·450	1·618		2·203	2·384	2·294
80			4·496	4·496			6·900	6·900
				SUBJECT A. B.				
0	−1·104			−1·104	−1·326			−1·326
10	−1·024			−1·024	−1·202			−1·202
20	−0·891	−0·967		−0·929	−1·087	−1·119		−1·103
30	−0·731	−0·744	−0·843	−0·773	−0·918	−0·862	−0·922	−0·900
40	−0·490	−0·432	−0·466	−0·456	−0·648	−0·588	−0·529	−0·588
50	−0·053	−0·022	−0·034	−0·036	−0·089	−0·172	−0·112	−0·124
60		0·482	0·612	0·547		0·617	0·660	0·638
70		1·682	1·677	1·680		2·124	2·021	2·073
80			4·505	4·505			4·717	4·717
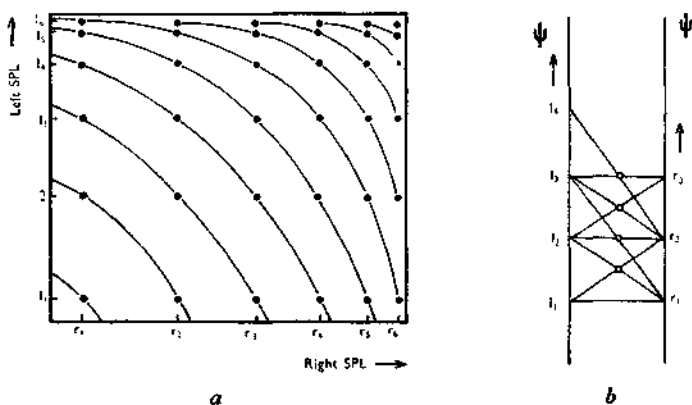
FIG. 2.—*a*, A theoretical set of equi-loudness curves and a 'dual-standard sequence'.
b, The staircase method of constructing a dual-standard sequence.

the 36 possible binaural pairs that can be constructed from the ranks of the left and right stimuli (the black dots) form a DSS. For example, the pair (r_4, l_1) matches (r_3, l_2), (r_2, l_3) and (r_1, l_4), because these are all on the same equi-loudness curve.

A quick means of constructing such a DSS is by the staircase method, depicted in Fig. 2b. Starting from an arbitrary pair (l_1, r_2), and an arbitrary stimulus $r_1 < r_2$, the subject is given the task of finding an l_2 such that (l_1, r_2) matches (l_2, r_1) in loudness. Then the pair (l_2, r_2) and the stimulus r_1 are presented to the subject, and his task is to find an l_3 such that (l_2, r_2) matches (l_3, r_1). This procedure can be continued—'climbing the staircase' so to speak.† The
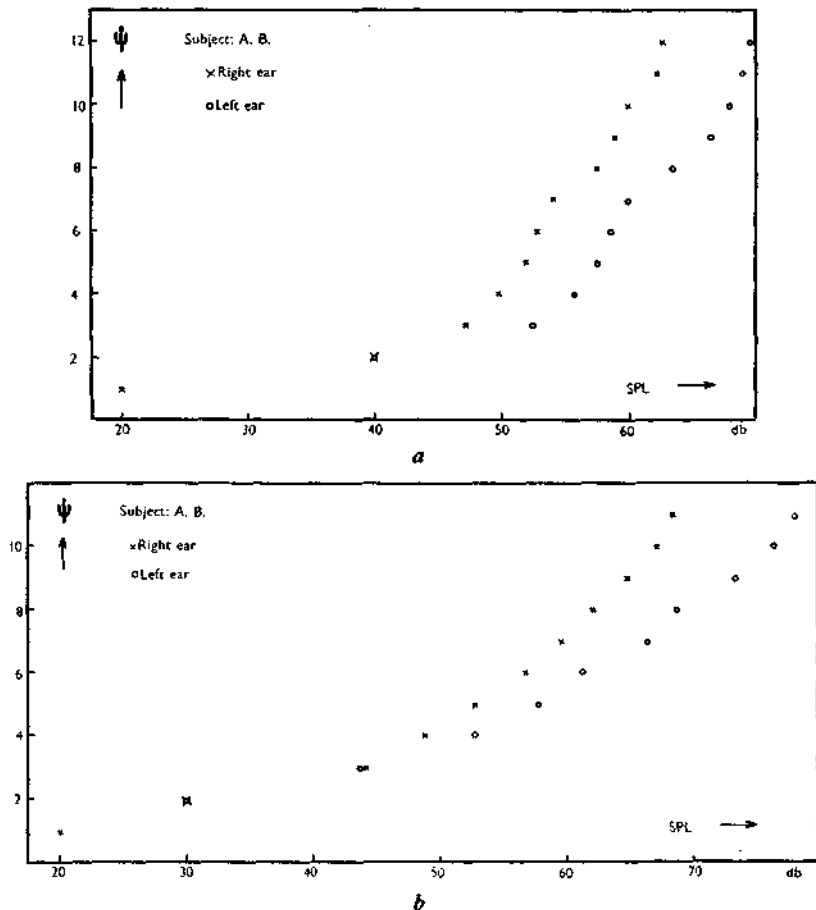

FIG. 3.—Worst (*a*) and best (*b*) result of binaural staircase measurement of loudness for subject A. B.

† To the best knowledge of the authors this psychometric method was first used by Davidson *et al.* (1957) for interval measurement of utility.

Cartesian product of the obtained sets of left and right stimulus values forms a dual-standard sequence.

It should be noted that this staircase method is, in a way, the reverse of the paired comparison procedure, by which we tried to find scale values for a set of given stimulus intensities. By the staircase method we try to find stimuli corresponding to equal-interval scale steps.

The staircase method is a quick way of measuring psychophysical interval scales simultaneously for left and right ears. In order to give some impression of what can be expected from the use of this procedure, we present in Fig. 3 the worst and the best results of eight such measurements for subject A. B.

Fig. 3a shows the worst result. Here the initial pair $(l_1, r_2) = (40 \text{ db}, 40 \text{ db})$, $r_1 = 20$ db. The subject's first match was $l_2 = 52 \cdot 5$ db, i.e. (20 db, 52·5 db), etc. This result shows the typical disadvantage of the method: errors of successive measurements are dependent. If one bad match is made the curves become discontinuous. Davidson *et al.* (1957) develop a procedure by which this effect can be largely reduced, whereas at the same time yielding confidence intervals for the measurements. However, their procedure—essentially checking measurements back and forth—is rather time-consuming. We did not apply it and will not discuss it. Fig. 3b shows the best results from the same subject. The series started with the pair (30 db, 30 db), and $r_1 = 20$ db. The resulting curves are smooth, and it is inviting to make a comparison between the scale thus obtained and the scale resulting from the paired comparison procedure. In Fig. 4 the
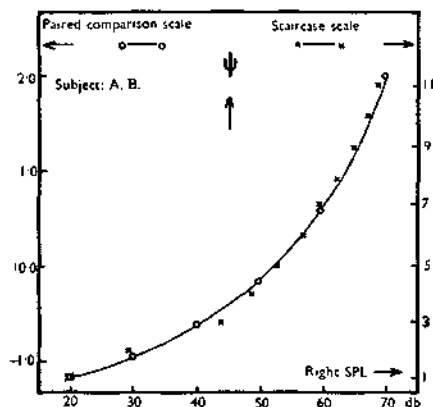

Fig. 4.—A comparison of psychophysical curves derived from paired comparison data and from the staircase procedure.

right-ear psychophysical curves (subject A. B.) from the paired comparison data (Table 4, bottom right column) and the staircase procedure (Fig. 3b) are brought into alignment. It is obvious that we have the same type of psychophysical relation in the two cases. Before we discuss this relation further, we want to mention one more procedure of the conjoint measurement variety, which we will call 'standard interval matching'.

5.3. *The Standard Interval Matching Procedure*

For this procedure a standard interval is chosen for one of the variables, e.g. in the right sound-pressure domain. Let us denote the pair by (r_s, r_s'), where $r_s < r_s'$. For each of a set of left-ear stimuli, l_i, the subject is required to find an l_i', such that (r_s', l_i) matches (r_s, l_i'). In this way intervals $\Delta l_i = l_i' - l_i$ are constructed, which are equal sensation steps for all i. Thus standard interval matching is the conjoint measurement counterpart of the method of equal appearing intervals. If one chooses l_{i+1} to be l_i', the equal appearing intervals are conjoint, and there is no accumulation of error, as with the previous procedure.

6. Binaural versus Monaural Stimulation

It was noted earlier that Treisman & Irwin adopted the conclusion of Reynolds & Stevens, that scales derived from the assumption of binaural summation (i.e. Fletcher's procedure) are different from the power scales obtained by direct judging methods. We have seen that this assumption is based solely on the character of monaural–binaural data, and that the assumption leads to a two-stage model: an additive stage, followed by a rescaling stage.

Next, we considered the feasibility of deriving scales from binaural–binaural data only. This removes the necessity of adopting the assumption of binaural summation. For binaural–binaural data the assumption of full binaural summation reduces to the additivity hypothesis, which is an empirical issue. Further reasons for restricting the model to binaural–binaural data will be given presently. The additivity hypothesis was tested and verified in Section 4, which opens the way to reconsidering the issue of whether for binaural–binaural data the loudness functions derived from this assumption are similar or dissimilar to the 'usual' loudness functions. If they turn out to be different, a two-stage model of the Treisman & Irwin type will be necessary to describe the data. But if they are similar, the rescaling stage of the model can be eliminated.

In the following we will take 'usual' monaural loudness functions to be loudness functions that are power functions of sound pressure for values from 20 db on, with exponent n around 0·54. This interpretation is, we think, a faithful reflection of the position taken by Reynolds & Stevens in discussing their experiments. Clearly, the power relation does not extend below 20 db SPL. Because of experimental variability, the character of the psychophysical function is less clear for these low sound-pressure levels, and we will therefore not use them as evidence, but limit the following considerations to values greater than 20 db.

The question, then, is whether the data in Table 4 (i.e. curves as in Fig. 4) can be reduced to a power relation, $\psi' = k\phi^n$, where ψ' is a psychological magnitude and ϕ a physical magnitude. Of course, this requires ψ' to be a ratio scale, whereas the ψ measures in Table 4 are only of the interval type. Our present problem, then, is whether values b can be found such that $\psi' = \psi + b$ and $\psi' = k\phi^n$. The value b is an additive constant, which defines the origin of a psychological scale. We determined the four optimal b values for the left- and

right-ear scales of the two subjects.† Figs. 5 and 6 give the log–log plots of ψ' against sound pressure, for subjects J. R. and A. B., respectively. It is obvious that one could hardly have expected better linearities from experimental data.
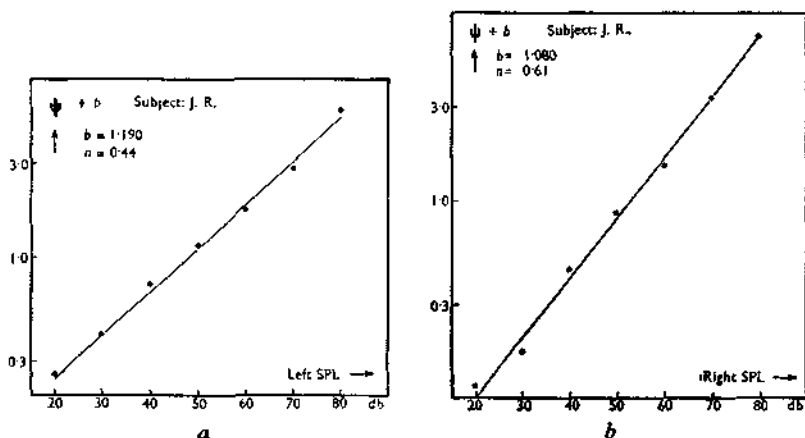

FIG. 5.—A log–log plot of ψ' and sound pressure for left ear (*a*) and right ear (*b*) of subject J. R. Exponents are $n = 0.44$ and $n = 0.61$, respectively.
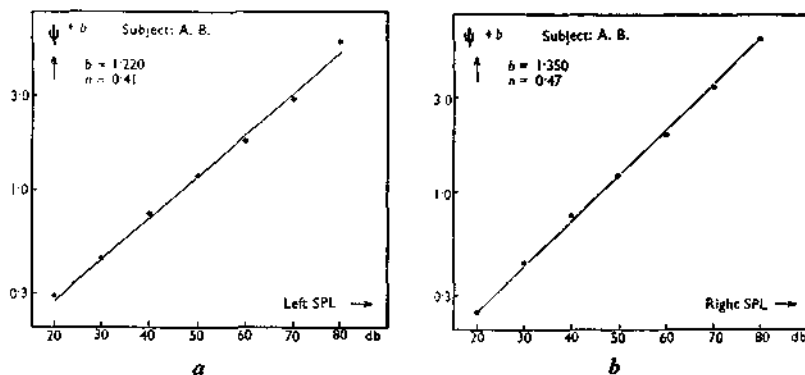

FIG. 6.—A log–log plot of ψ' and sound pressure for left ear (*a*) and right ear (*b*) of subject A. B. Exponents are $n = 0.41$ and $n = 0.47$, respectively.

† The b values can be computed via a least squares solution of

$$\sum_i \{\log (\psi_i + b) - (n/20) S_i - c\}^2,$$

where ψ_i and S_i are the respective scale values and decibel values in Table 4, n the exponent of sound pressure and c a constant. By setting the three partial derivatives with respect to b, n and c equal to zero, one obtains the three equations from which b, n and c can be solved. This, however, is very laborious. In fact, we computed b values by graphical iteration. It consisted of choosing a b value, drawing a log–log plot, checking departure from linearity and changing b slightly in the appropriate direction, etc. Computation would at best have given even better linearities than the ones obtained.

The exponents are given in the figures: they are 0·44 and 0·61 for the left and the right ears of subject J. R. and 0·41 and 0·47 for the left and the right ears of subject A. B. They are, clearly, of the order of magnitude of the usual exponents for loudness.

Moreover, it is worth noticing that the obtained b values, which are additive constants, not only remove negativity for the values $\geqslant 20$ db, but also for 0 and 10 db. By extrapolation, zero-points would lie around -10 db, which is just about threshold. The transformations therefore yield, in a natural way, non-negativity for the ψ' scales.

These results lead to the conclusion that a one-stage model suffices for binaural-binaural data: binaural loudness is an additive function of monaural loudnesses.

For the data at hand, no 'controversial' loudness functions are obtained. The complicated functions proposed by Treisman & Irwin (power functions with decreasing power) can be dispensed with, but the restriction is that the model only applies to simultaneous non-zero stimulation of the two ears.

Should such a restriction be made? Of course, we do need a model for the description of monaural–binaural matching data, and our additive loudness model is clearly insufficient for this purpose. Nevertheless, the restriction made is rather natural. In several experiments, which will be reported in a later paper, we have been able to show that the loudness of a 1000 cps tone in one ear is strongly influenced by the quality of a 1000 cps tone in the other ear. The contribution of a continuous 1000 cps tone to binaural loudness is much less if the tone in the other ear is regularly interrupted (three times per second for 30 msec.) than if it is also continuous. It appears that attention shifts to the interrupted tone. Obviously, if no tone is given in one ear, full attention will be given to the monaural tone of the other ear, which may lead similarly to an increased loudness effect for that tone. What is suggested here for binaural loudness has been proved for binocular vision (Levelt, 1968): for monocular vision the brightness parameters are different than for binocular vision.

We suspect that a similar change in parameters (i.e. of k and/or n in the functional relation $\psi' = k^n$) occurs if, due to monaural stimulation, an attention shift takes place. For this reason it seems unlikely that an additive model with $fixed$ parameters, such as that of Treisman & Irwin, can take account of both monaural–binaural and binaural–binaural matching. A fixed-parameter model, for the latter type of data only, is therefore a natural restriction. Extension of the model requires the study of attentional factors. Or, as Carl Stumpf (1890) remarked about observers reporting on changing from monaural to binaural listening: 'what they really heard as an increase in clarity, fullness, and volume, they thought they heard as an increase in loudness' (cited by Hirsch, 1948). An important step towards unravelling this knot of tone quality, attention and loudness in binaural hearing is to be found in Sharf's recent (1969) article.

<div align="center">REFERENCES</div>

Békésy, G. von (1929). Zur Theorie des Hören. Über die eben merkbare Amplituden- und Frequenzänderung eines Tones. Die Theorie der Schwebungen. *Physik* 30, 721–745.

DAVIDSON, D., SUPPES, P. & SIEGEL, S. (1957). *Decision Making*. Stanford: Stanford University Press.

FLETCHER, H. & MUNSON, W. A. (1933). Loudness: its definition, measurement and calculation. *J. acoust. Soc. Am.* **5**, 82–108.

HELLMAN, R. P. & ZWISLOCKI, J. (1963). Monaural loudness function at 1000 cps and interaural summation. *J. acoust. Soc. Am.* **35**, 856–865.

HIRSCH, I. J. (1948). Binaural summation: a century of investigation. *Psychol. Bull.* **45**, 193–206.

IRWIN, R. J. (1965). Binaural summation of thermal noises of equal and unequal power in each ear. *Am. J. Psychol.* **78**, 57–65.

KRANTZ, D., LUCE, R. D., SUPPES, P. & TVERSKY, A. (1971). *Foundations of Measurement*. New York: Academic Press.

KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a non-metric hypothesis. *Psychometrika* **29**, 1–27.

LEVELT, W. J. M. (1968). *On Binocular Rivalry*. The Hague: Mouton.

LUCE, R. D. & TUKEY, J. W. (1964). Simultaneous conjoint measurement: a new type of fundamental measurement. *J. math. Psychol.* **1**, 1–27.

REYNOLDS, G. S. & STEVENS, S. S. (1960). Binaural summation of loudness. *J. acoust. Soc. Am.* **32**, 1337–1344.

ROSKAM, E. E. C. I. (1968). *Metric Analysis of Ordinal Data in Psychology: Models and Numerical Methods for Metric Analysis of Conjoint Ordinal Data in Psychology*. Voorschoten: VAM.

ROSKAM, E. E. C. I. & VAN GILST, W. (1967). 'Addit' 007/02. (Department of Psychology Computer Bulletin, Nijmegen University.)

SHARF, B. (1969). Dichotic summation of loudness. *J. acoust. Soc. Am.* **45**, 1193–1205.

STEVENS, S. S. & DAVIS, H. (1938). *Hearing: Its Psychology and Physiology*. New York: Wiley.

STUMPF, C. (1890). *Tonpsychologie*. Leipzig: Hirgel.

TREISMAN, M. & IRWIN, R. J. (1967). Auditory intensity discriminal scale. I. Evidence derived from binaural intensity summation. *J. acoust. Soc. Am.* **42**, 586–592.

URBANTSCHITSCH, V. (1893). Über Wechselbeziehungen zwischen beider Gehörorgane. *Arch. Ohrenhlk.* **35**, 1f.