Individual Differences in Loudness Processing and Loudness Scales

Daniel Algom and Lawrence E. Marks
John B. Pierce Foundation Laboratory, Yale University

Parameters of the psychophysical function for loudness (a 1000-Hz tone) were assessed for individual subjects in three experiments: (a) binaural loudness summation, (b) temporal loudness summation, and (c) judgments of loudness intervals. The loudness scales that underlay the additive binaural summation closely approximated S. S. Stevens's (1956) sone scale but were nonlinearly related to the scales that underlay the subtractive interval judgments, the latter approximating Garner's (1954) lambda scale. Interindividual differences in temporal summation were unrelated to differences in scaling performance or in binaural summation. Although the exponents of magnitude-estimation functions and the exponents underlying interval judgments varied considerably from subject to subject, exponents computed on the basis of underlying binaural summation varied less. The results suggest that interindividual variation in the exponent of magnitude-estimation functions largely reflects differences in the ways that subjects use numbers to describe loudness and that the sensory representations of loudness are fairly uniform, though probably not wholly uniform, among people with normal hearing. The magnitude of individual variation in at least one measure of auditory intensity processing, namely, temporal summation, seems at least as great as the magnitude of the variation in the underlying loudness scale.

The construction of a valid scale to measure loudness has long been a major concern of sensory investigators for both theoretical and practical reasons. To this end, S. S. Stevens (1956) developed direct estimation procedures, notably magnitude estimation, through which he derived the modern sone scale of loudness. The sone function states that the loudness (L) of a 1000-Hz tone increases as 0.6 power of sound pressure (SPLs above 40 dB). Thus L in sones is defined as follows:

\[ L = k P^{0.6} \]

where L is 1 sone when P, the sound pressure of a 1000-Hz tone heard binaurally, has a value of 0.002 N/m², which is equivalent to 40 dB (SPL); k is a constant that defines the scale unit. The scale is mainly based on results of direct numerical judgments of loudness, especially on results of magnitude estimation. Although the sone scale has been accepted by the International Standards Organization as the scale of loudness, its validity has been repeatedly questioned (e.g., Beck & Shaw, 1967; Birnbaum & Elmasian, 1977; Schneider, 1980).

Response Biases and Individual Differences

Is the sone scale a valid sensory scale? How well does it represent loudness functions of individual subjects? The questions are interrelated. The prima facie validity of direct scaling techniques such as magnitude estimation relies on the assumption that putative numerical ratios represent actual ratios of loudness; yet as early as 1954 — barely a year after S. S. Stevens first suggested the method of magnitude estimation and the resulting power law (see S. S. Stevens, 1975) — Garner questioned this underlying assumption. Ever since, many investigators have pointed to the possible existence of nonlinear judgmental factors influencing numerical estimates (e.g., Anderson, 1972; Birnbaum & Elmasian,
1977; Curtis, Atteave, & Harrington, 1968; Torgerson, 1961). The continuing controversy over the sone scale has been triggered in part by demonstrations of a nonlinear relation between judgments based on loudness magnitudes and judgments based on loudness intervals or differences (Beck & Shaw, 1967; Marks, 1974a; S.S. Stevens, 1971; S.S. Stevens & Galanter, 1957). If there is a single, valid scale of loudness, it should not depend on method. Though the method of cross-modality matching provides a test of internal consistency (S. S. Stevens, 1975), it does not provide a test of validity (Treisman, 1964; Zinnes, 1969) and may not serve to differentiate magnitude scales from interval scales (Marks, 1974a).

The biases involved in the numerical responses obtained by direct estimation methods seem especially relevant to another characteristic of the data that these methods generate, namely, individual differences. Even under well-controlled conditions, loudness scales derived from methods such as magnitude estimation vary considerably from subject to subject. Even after removing systematic effects such as those of stimulus range or standard stimulus (Poulton, 1968; S.S. Stevens, 1956; Teghtsoonian, 1973), this variability remains. The exponent of the power function for loudness can vary by 2:1 or more within a single experiment (J. C. Stevens & Guirao, 1964; Marks, 1974a). It seems unlikely that these differences are attributable mainly to variation in sensory processes. As S. S. Stevens (1971) and Marks (1982) pointed out, to take the different exponents at their face values would mean to accept differences of 100- or even 1,000-fold among normal subjects in the perception of the loudnesses of ordinary sounds. When exponents range from below unity to above unity (which is occasionally the case), this would mean substantial variation in the operating characteristics of the subjects' auditory systems. Still, some of the variation may have a sensory basis, so we should entertain the possibility that individual functions differ considerably (cf. Luce, 1972).

However, the alternative hypothesis—that the auditory system does not change materially from subject to subject (or from task to task or from condition to condition)—seems more probable (Marks, 1982; cf. Teghtsoonian & Teghtsoonian, 1983). The variations in the responses presumably result largely from differences in the ways that people use numbers. Following Teghtsoonian (1973), it is possible to reduce variation in the exponent to variation in the dynamic response range because the response range is one expression of the idiosyncratic ways that people use numbers. However, in this case the average scale, the sone scale, itself becomes suspect. Again, we face the problem of validity. S. S. Stevens's (1971, p. 429; 1975, p. 32) approach—taking a “balanced average” of results of estimation and production—is irrelevant to a decision between the alternatives: If people differ notably in their use of numbers, then averaging their data may amount to no more than providing a summary statistic of little value in assessing sensation.

Is there a way to identify the source or sources of the intersubject variability and thus simultaneously to validate the sone scale or some other scale of loudness?

Validation of Loudness Scales and Models of Sensory Processes

The lack of a satisfactory criterion for validation has for a long time adversely affected the work on magnitude estimation and has cast doubt on the power law. Without such a criterion, any choice among methods is essentially arbitrary (Anderson, 1970; Garner & Creelman, 1967). The basic problem may not even be so much the absence of proper validation techniques as the absence of a theoretical basis to derive such techniques. Growing recognition of this lack has recently led to the construction of substantive models of sensory processes that contain scales as their natural derivatives. Thus scaling becomes an integral part of the psychophysics of sensory processes, the validation of the theory constituting at once a validation of the scale (Anderson, 1970; Marks, 1979b). An effort has been made to erect models, including their judgmental elements, whose formal properties are testable (see Krantz, Luce, Suppes, & Tversky, 1971).

The approach (1970, 1972) for judgments (or other psychological quantities) in conjunction with a combining scale (or other psychological quantity) as an additive structure (or other psychological quantity) while the numerical judgment itself (such as the magnitude estimation design) are limited to the scale values. The basic problem is determining the transformation that complicate the relations.

Marks's (1975) audiometric loudness judgments and loudness scales within the auditory processing. The introduction of loudness—its auditory process is divided into several stages. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

1. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

   a. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

      i. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

         1. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

            a. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

               i. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

                  1. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

                     a. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

                        i. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

                           1. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:

                              a. The transformation of auditory phenomena to the world of loudness and the auditory process is divided into several stages:
The approach advocated by Anderson (1970, 1972) makes use of judgments of averages (or other operations of mental arithmetic) in conjunction with a linear model for combining scale values as the basis for scaling psychological variables such as loudness. If an additive structure is directly evident in the numerical judgments, then the responses (such as the marginal means in a factorial design) are linearly related to the underlying scale values. This solution to the scaling problem is designated functional measurement. Note, however, that superimposing complex cognitive operations (such as judgments of averages) on sense perception, a common technique of functional measurement, may actually add more nonlinear transformations onto the data and thus may complicate the solution (Marks, 1978b).

Marks's (1979b) theory of loudness and loudness judgments tries to integrate loudness scales with a substantive theory of auditory processing. According to this theory, information about sound intensity—about loudness—is processed in a hierarchy of stages. The theory accounts for several auditory phenomena, exemplifying the intricate, indeed critical, relations that exist between loudness and many well-documented auditory processes. Just how we describe certain auditory phenomena can depend on our definition of loudness, that is, on its relation to sound intensity. On the other hand, this very quantification of loudness should be reasonable in terms of our immediate auditory experiences because one aspect of what we experience is loudness.

This twofold relation expresses itself clearly in the phenomenon of binaural summation of pure tones (Marks, 1978a). The summation of loudnesses across the ears is linear only when loudness is computed in sones. According to the equation for sones, the experience of loudness doubles with a 10-dB increase in sound pressure. Now, the loudness of a binaurally presented tone equals the loudness of a corresponding monaural tone augmented by 10 dB, and this, in turn, means linear addition in sones. Thus the validation of the sone scale reduces to a set of invariances: Sounds that are equal in total number of sones are judged equal in loudness. The measurement of loudness as a psychological quantity comes naturally, therefore, from data and theory of auditory psychophysics.

The scheme just described provides a general framework for resolving the problem of individual differences in loudness functions. The main question is this: Are differences in the exponents of the individual loudness functions accompanied by corresponding differences in other indices of sensory experience related to their loudness functions, say in binaural summation? If so, then these differences in exponents could justifiably be attributed to real differences in the way individuals process auditory intensity. If, however, inter-subject variation is greatly reduced by employing other indices of auditory function, indices that are derived independently of subjects' numerical judgments, then we would be more confident in concluding that the variation of individual scaling functions results largely from differences in the ways subjects employ numerical concepts.

The purpose of the following experiments was to relate parameters of individual loudness functions (in particular, power-function exponents) to other indices of auditory processing. We obtained the following measures for each subject: (a) psychophysical (magnitude-estimation) functions, (b) indices of binaural summation, (c) indices of temporal summation, and (d) psychophysical functions for judgments of loudness intervals between sounds. The last measure had the additional purpose of showing that all auditory experiences do not necessarily converge on a single loudness scale. Similar sets of converging operations can be used repeatedly and hierarchically to validate different scales of loudness (Marks, 1979b, 1982).

Experiment 1: Binaural Summation and Individual Loudness Functions

The purposes of this experiment were as follows: (a) to extract parameters of each subject's loudness function, (b) to uncover the rules of binaural summation for each subject, (c) to derive a loudness scale based on each subject's binaural summation, and (d) to compare this scale to the magnitude-estimation scale.
Marks (1978a, 1980) used the method of magnitude estimation in conjunction with a factorial design to explore the rules of binaural loudness summation. In all cases, the results were consistent with a model of linear summation for pure tones and narrow-band noises. The estimates of loudness of a binaurally presented tone approximated the linear sum of the loudness estimates of the left-ear and right-ear components: A binaural sound was twice as loud as a monaural sound of the same sound pressure level (SPL). To equal the loudness of a binaural tone, a monaural tone has to be 10 dB greater in intensity (except at low SPLs). This difference in intensity—the summation index—implies the operation of the same scale because 10 dB represents a doubling of loudness in sones (S. S. Stevens, 1956). Thus Marks concluded that the underlying loudness scale could be described by the same scale (by a power function with an exponent of about 0.6). Comparable results were obtained by Jankovic and Cross (1977), using a similar design, by Hellman and Zwillocki (1963), and by Levelt, Riemersma, and Bunt (1972), using other experimental designs (but see also Scharf, 1969, and Scharf & Fishken, 1970, for results suggesting less-than-complete summation).

Recent scaling studies of binaural summation (e.g., Marks, 1978a, 1979a, 1980) have reported group data (pooled over subjects). The question posed by this study, using a factorial design similar to that employed in the above experiments, is as follows: How uniform from subject to subject is the loudness scale underlying binaural summation?

By using the method of magnitude estimation to examine binaural summation, the present study makes it possible both to examine the rules underlying summation and to assess over psychophysical functions directly from numerical judgments. This enables us to compare the numerical scales with implicit loudness scales underlying binaural summation, the latter being derived independently of the metric of the subjects' numerical judgments.

Method

Apparatus. The 1000-Hz signal from a Heath tone generator was gated (Grason-Stadler 920C electronic switch), timed (Grason-Stadler 471 timer), and then split into channels for the left and right ears. Stimuli lasted 1 s, with 10-ms rise and fall times. The signal in each channel could be attenuated independently before being fed to the TDH-39 headphones.

Procedure. The experimental design was that used by Marks (1978a). The subject sat in a sound-shielded booth. Seven different levels of sound pressure (six SPLs from 20 dB to 50 dB, plus a level well below threshold—0) to the left ear were combined factorially with the same seven levels to the right ear, making 49 different stimuli in all. Stimuli were presented in a fixating stimulus; the subject for the subject received three replicates of the stimulus in a session and in five sessions, thus giving 15 judgments per stimulus in all. The order of presentation of the stimulus was randomized and was different for each subject ( orderly sequences were intentionally avoided).

The method was magnitude estimation. Subjects were instructed to assign to the first stimulus whatever number seemed most appropriate to represent its loudness; then subjects were instructed to assign numbers, in proportion, to succeeding stimuli. If no sound was heard, the subjects were to assign the number 0. Subjects were told that they could use whole numbers, decimals, and fractions as needed. In addition, they were instructed to disregard the apparent location of the stimulus and to judge only their loudness.

Subjects. Nine young men and women, all paid volunteers from the Yale community, served as subjects. Five of them had previous experience with the method of magnitude estimation, though not necessarily in judging loudness. Because these same subjects were to serve in the subsequent experiments, they were not informed individually of their results before all testing was completed.

Results and Discussion

Pooled data. We present the group results first. The magnitude estimates assigned to each stimulus were brought to a common modulus (Lane, Cariatu, & Stevens, 1961) for all subjects and then were averaged geometrically. These means are plotted as a function of the SPL delivered to the right ear (see Figure 1). Each contour represents a different SPL delivered to the left ear. Because the loudness estimates are plotted on a linear scale, the hypothesis of linear loudness summation may be assessed by visual inspection. A salient characteristic of this family of curves is the roughly equal spacing in the vertical dimension (though a slight trend toward convergence at the upper right is evident). The parallel spacing implies linear additivity of the numerical responses (Anderson, 1970). To assess the Left Ear X Right Ear interaction, the matrix of judgments was subjected to an analysis of variance (ANOVA). Because the present data had reasonably unform variance, the critical additivity was tested using a two-factor ANOVA (Anderson, 1970) (nonsignificance of the interaction, F 392 < 1, the results of the ANOVA are given).

Plotted in Figure 1 is a subset of the data from the same experiment, the present data. The responses are for simple binaural summation.

The dashed line was drawn by a computer program depicting the stimuli that fall along the monaural response curve. As shown in Figure 1, the binaural responses as well as the monaural responses and the average responses to the stimuli are all equal, loudness estimates.

Another analysis was performed on the data, which are given in Table 1. The data are divided into columns and a computer program was used to perform the analysis, and then by...
form variances (not always the case with magnitude estimates), the application of an ANOVA seemed legitimate. The interaction is the critical term to assess because failure of additivity will appear as a significant interaction (Anderson, 1970). The data showed a nonsignificant overall interaction, $F(36, 392) < 1$, thus confirming statistically the results of the graphical test of parallelism: linear addition of loudness across the two ears.

Plotted in the upper portion of Figure 2 is a subset of the present results, namely, data for the presentations that were monaural and simple binaural (equal SPLs to the two ears). The dashed line represents the prediction made by a model of linear summation, depicting the sum of the magnitude estimates of the monaural (left ear and right ear) stimuli; as shown in the figure, this prediction agrees well with the actual magnitude estimates of the binaural stimuli. In fact, the (geometric) average ratio of the binaural:monaural loudness estimates was 1.99.

Another indicator of the underlying scale values comes from taking marginal means, which are generated by summing down columns and across rows of the response matrix and then by setting to zero the scale values at zero stimulus intensity. Because they are derived from the entire response matrix, the marginal means have a fuller basis than do the monaural judgments (which, by definition, are derived from only a subset of the data). The marginal means are plotted in the lower part of Figure 2. Here again, the dashed line gives the prediction of the model of linear summation, and again the line passes close to the actual magnitude estimates of the binaural stimuli.

The monaural and binaural loudness functions—both the magnitude estimates and the marginal means—approximate power functions (with the proviso that the function includes a subtractive constant as a “threshold” correction). The exponent of the function—determined by an iterative method of least squares, operating on log response and log (stimulus minus threshold) (see J. C. Stevens & Marks, 1967) —is 0.58 in both cases. S. S. Stevens (1956) proposed 0.6 as the value governing the sone scale. Given a 0.6 power function and a linear rule of summation, it follows that a monaural stimulus must be 10 dB greater than an equally loud binaural stimulus. The average horizontal separation between the monaural curves and
the binaural curve in Figure 2 is close to 10 dB, implying practically complete binaural summation in sones. The results, taken as a whole, can therefore be characterized by a simple rule: Binaural loudness equals the linear sum of loudnesses of the components presented to the right and left ears, when loudness is counted in sones.

These pooled results serve two purposes: First, they substantiate the findings of a line-rule binaural summation for pure tones and of the underlying zero scale. Second, the same types of analyses with the same kinds of interpretations may be applied to the data of each individual subject. Thus the group data serve as a common frame of reference to evaluate individual cases.

Individual psychophysical functions. Analyses of the data on individual subjects, the main concern of the present study, appear in Table 1. Two aspects of each subject's results are of special interest: the psychophysical function for loudness and the rule of binaural summation.

First, consider the loudness functions. As is typical, subjects differed considerably in the ways they assigned numbers to the various sound stimuli. Although each subject's data could be described by a power function, the exponents \( \beta \) of the magnitude-estimation functions for monaural and equal-SPL binaural stimuli varied notably, from 0.37 to 0.86 with a mean of 0.60 (second column in Table 1). This range of loudness exponent, which is somewhat more than 2:1, is typical

<table>
<thead>
<tr>
<th>Subject</th>
<th>( \beta_{m} )</th>
<th>( \beta_{n} )</th>
<th>( BR )</th>
<th>( \beta_{m1} )</th>
<th>( \beta_{n1} )</th>
<th>( \beta_{m2} )</th>
<th>( \beta_{n2} )</th>
<th>( \beta_{run} )</th>
<th>( \beta_{runm} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.72</td>
<td>2.42</td>
<td>0.56</td>
<td>2.09</td>
<td>0.66</td>
<td>2.23</td>
<td>0.61</td>
<td>0.55</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>0.43</td>
<td>1.71</td>
<td>0.55</td>
<td>2.11</td>
<td>0.60</td>
<td>2.12</td>
<td>0.41</td>
<td>0.45</td>
<td>0.54</td>
</tr>
<tr>
<td>3</td>
<td>0.48</td>
<td>1.61</td>
<td>0.70</td>
<td>1.81</td>
<td>0.79</td>
<td>2.19</td>
<td>0.51</td>
<td>0.73</td>
<td>0.66</td>
</tr>
<tr>
<td>4</td>
<td>0.86</td>
<td>2.69</td>
<td>0.50</td>
<td>1.99</td>
<td>0.87</td>
<td>2.74</td>
<td>0.71</td>
<td>0.72</td>
<td>0.60</td>
</tr>
<tr>
<td>5</td>
<td>0.66</td>
<td>2.10</td>
<td>0.61</td>
<td>1.96</td>
<td>0.60</td>
<td>1.96</td>
<td>0.53</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>6</td>
<td>0.37</td>
<td>1.34</td>
<td>0.87</td>
<td>1.61</td>
<td>0.72*</td>
<td>1.76</td>
<td>0.86</td>
<td>0.80</td>
<td>0.77</td>
</tr>
<tr>
<td>7</td>
<td>0.71</td>
<td>1.98</td>
<td>0.72</td>
<td>1.78</td>
<td>1.19*</td>
<td>3.13</td>
<td>0.67</td>
<td>0.73</td>
<td>0.75</td>
</tr>
<tr>
<td>8</td>
<td>0.46</td>
<td>2.50</td>
<td>0.78</td>
<td>1.70</td>
<td>0.88*</td>
<td>2.20</td>
<td>0.50</td>
<td>0.68</td>
<td>0.68</td>
</tr>
<tr>
<td>9</td>
<td>0.68</td>
<td>2.55</td>
<td>0.50</td>
<td>2.28</td>
<td>0.70</td>
<td>2.67</td>
<td>0.50</td>
<td>0.52</td>
<td>0.68</td>
</tr>
<tr>
<td>M</td>
<td>0.60</td>
<td>1.98</td>
<td>0.65</td>
<td>1.93</td>
<td>0.78</td>
<td>2.33</td>
<td>0.62</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>SD</td>
<td>0.16</td>
<td>0.47</td>
<td>0.11</td>
<td>0.20</td>
<td>0.17</td>
<td>0.40</td>
<td>0.12</td>
<td>0.08</td>
<td>0.08</td>
</tr>
</tbody>
</table>

Note: \( \beta_{m} \) = exponent of the loudness function and \( BR \) = average binaural:monaural ratio; both are from original magnitude-estimation data. \( \beta_{run} \) = exponent derived from magnitude-estimation data, corrected by binaural ratio; \( BR_{1} \) = rescaled binaural ratio, corrected by loudness exponent, \( \beta_{run1} \) = rescaled loudness exponent, binaural additivity maximized; \( \beta_{runm} \) = exponent derived from marginal means, original data; \( \beta_{runm} \) = exponent derived from marginal means, rescaled data; \( \beta_{run} \) = exponent derived from binaural summation index.

*Derived from data that required rescaling, that is, from data that had a significant interaction term.

Individual binaural loudness values, measured under monaural:monaural conditions, were converted into different thresholds of loudness, and the linear judgments were fitted into a power function.
(J. C. Stevens & Guirao, 1964; also see Marks, 1974a). Although interindividual differences could, in theory, represent real differences in the input–output function of the auditory apparatus, it seems more likely that much of the variation results from differences in the use of numbers. In this regard, it is useful to compare the subjects' number behavior with their degree of binaural summation.

Binaural: monaural loudness. The third column in Table 1 lists, for each subject, the (geometric) average of the ratios of binaural loudness to monaural (average left-ear-plus-right-ear) loudness. Recall that when the data were pooled over subjects, the average ratio was nearly 2:1. However, individual subjects show considerable fluctuation, as indicated in Table 1; the mean ratio is 1.98, suggesting virtually perfect loudness summation on the average, but the standard deviation is nearly 25% of the mean. Inspection of columns 2 and 3 of Table 1 indicates a close connection between the exponent and the summation ratio: Where the exponent is great (or small), so is the ratio (for $\beta_{me}$ and log $BR$, r(7) = .93, p < .01, where $\beta_{me}$ is the exponent of the loudness function from the original magnitude-estimation data and $BR$ is the average binaural:monaural ratio). This is precisely what one would predict if the variations in the exponent and the binaural:monaural ratio were due largely to a single source, namely, variation in the use of numbers.

Consider a simple case, where every person’s underlying, sensory, psychophysical function for loudness (barring hearing loss) is the same:

$$L_m = k_1 P^n_m$$

for monaural listening,

$$L_b = k_2 P^n_b$$

for binaural listening,

and, moreover, that binaural summation of loudness is perfect, binaural loudness being exactly twice monaural, so that we have the following:

$$k_2 = 2k_1.$$

Individual differences arise in the size of the measured exponent and in the size of the binaural:monaural loudness ratio because of different transformations imposed by nonlinear judgmental processes. Assuming a power function relates the numerical magnitude estimates ($ME$) to the underlying loudness ($L$),

$$ME = L^\beta,$$

then the measured exponents $\beta_{me}$ will equal $\beta \cdot a$, and the measured binaural ratios will equal $2^a$.

This simple model implies a proportionality between the measured exponent and the logarithm of the measured binaural ratio $BR$,

$$\beta_{me} = \beta \log BR \over \log 2.$$

Plotted in Figure 3 is $\beta_{me}$ against log $BR$; the line gives a proportional relation, one based on the assumption that the underlying value of $\beta$ is 0.6 (the sone scale). The slope of this line is 2.0 (it would be 1.0 if the acoustic stimulus were measured in terms of power rather than pressure, making the exponent $\beta$ equal to 0.3 instead of 0.6). The slope of the best-fitting line (averaging regressions in both directions) is 2.04, virtually identical with the predicted value.

This outcome implies that at least some of the variation in both measures is derived from a single source: differences in the use of
numbers. (It is conceivable, of course, but unlikely, that people actually differ both in loudness scale and in binaural summation and that these differences are fortuitously correlated.) In fact, the proportionality (not just linearity) between $B$ and $\log BR$ accounts for 56% of the variance in the individual exponents.

If we assume that binaural summation does not differ across individuals, then it is possible to "correct" the observed loudness exponents in accordance with the observed binaural ratios. Column 4 of Table 1 contains revised exponents, computed by multiplying each subject's magnitude-estimation exponent of column 2 by $1/(\log BR/\log 2)$. We raised each subject's exponent to whatever power was needed to make that subject's binaural:monaural ratio equal to 2.0. The average of these revised loudness exponents is 0.65, a bit higher than the original 0.60. The standard deviation, though, is smaller: 17.4% of the mean rather than 26.2%. If binaural summation is truly constant (and complete), the remaining variability should represent actual variation in the underlying loudness exponent (and noise in the measurements).

At the other extreme, it is conceivable that the underlying loudness function is constant across individuals. Assuming this to be so and that the value of the exponent is 0.6, it is also possible to revise the estimates of the individual binaural loudness ratios (see column 5 of Table 1). The mean ratio across subjects remains nearly 2:1; but the standard deviation is considerably reduced, to 10% of the mean instead of 24%. An argument could be made about variation in summation along lines similar to the argument made for variation in exponent: If it is the exponent that is invariant, then the revised estimates of the summation ratio reflect real intersubject variation (plus noise).

Binaural additivity. The question of summation may yield to an analysis of the additivity evident within each subject's entire response matrix. As shown in Figure 1, the pooled data conform well to a simple additive model. So, too, do most (but not all) of the individual data. Visual inspection of each subject's loudness contours—from graphical displays analogous to the one shown in Figure 1 (see Figure 4 for two examples)—shows that despite greater noise in the individual data, in most cases the curves are about equally displaced in the vertical plane. ANOVAs confirmed that the data from 6 of the 9 subjects had nonsignificant interaction terms, implying additivity of their numerical responses. For the 6 subjects, $F(36, 686) \leq 1.47, p \geq .05$. For the other 3, the interaction terms were significant, $F(36, 686) = 4.10, 4.51$, and $4.65, p \leq .01$; however, nearly all of the interaction resided in each case in the binaural component, $F(1, 686) = 63.8, 77.5$, and $101.9, p \leq .001$. Graphically, a significant binaural term appears as a tendency for the family of functions to diverge or converge at the upper right. Divergence or convergence is the hallmark of many nonlinear numerical scales when the underlying metric is additive (see Marks, 1979b). This nonlinearity can be eliminated by rescaling the numerical values to produce a set of curves that neither converge nor diverge. Such rescaling captures the spirit of two-stage models of magnitude estimation, as espoused by Curtis et al. (1968). (Rescaling involved an iterative procedure in which the original data $R$ are transformed by the equation $R' = R^p$, with $p$ varying from iteration to iteration to find the value that eliminates the divergence or convergence.)

The exponents $b_{\text{med}}$ were determined from the rescaled data, again by the iterative method of least squares, and these exponents appear in the sixth column of Table 1. Note that only those exponents marked with a superscript $a$ are derived from data that required rescaling, that is, from data that had a significant interaction term. The average of the rescaled exponents is 0.78, a value somewhat higher than the 0.6 -- 0.67 expected for loudness. Note, too, that the variability in the exponents is still sizable compared with that in the exponents derived from the original data ($b_{\text{med}}$); however, the variability in the rescaled data is inflated by the extreme value of 1 subject (No. 7), with which the mean exponent would drop to 0.73 and the standard deviation, to 0.10. Curiously, in this regard, the raw data of Subject 7 produced a binaural:monaural loudness ratio of 1.98, suggesting virtually complete loudness summation; yet the full set of data implies significant nonadditivity.

Figure 4. The data from Subject 4, from 0 to 30 dB.

It is important to take care in interpreting these results. First, the binaural information is a subset of the overall parallelism in the response matrix assumed by the additivity doctrine. A more complete increase, some at the high level and action at a lower level.

One possible equal or nonadditive (additive) model would be more discrepant from a model with Strube (1959) loudnesses, which suggests that the loudnesses when differ
nonadditivity. Ideally, of course, the two measures (summation and additivity) should coincide. One possibility, suggested earlier, is that data for some of the subjects may not obey a rule of exact addition.

The seventh column of Table 1 revises the individual binaural:monaural ratios in terms of the power functions used to rescale the individual data matrices. Rescaling tends to make the loudness ratios exceed 2.0 and, in particular, produces an exceptionally high ratio (3.13) for Subject 7. Again, although this discrepancy between additivity and summation may be fortuitous, it does pose the possibility of a distinct nonadditive element in Subject 7’s data.

It is important to indicate two reasons to take care in drawing too firm a conclusion. First, the binaural ratio depends on only a subset of the data. Second, the assessment of parallelism (additivity) in an entire data matrix assumes uniform variance; yet the variability does tend to be larger as the means increase, so a larger absolute deviation at a high level is less reliable than the same deviation at a low level.

One possible resolution is that stimuli of equal or near-equal intensities do summate (add totally and linearly), but stimuli with more discrepant intensities do not. Just such a model was proposed by Gigerenzer and Strube (1983), who suggested that when loudnesses are equal they sum linearly, but when different enough, the louder component may inhibit the softer one. In such a case, binaural:monaural ratios of 2.0 could coexist with overall deviations from additivity.

On the assumption that binaural additivity exists at least as a first-order approximation, we can derive scale values from each subject’s data matrix by calculating marginal means down columns and across rows, adjusting to zero the scale value of the zero-intensity stimulus. Columns 8 and 9 in Table 1 show, respectively, for original data (with nonsignificant interactions) and for rescaled data, the mean left-ear-and-right-ear exponents of power functions fitted by iterative least squares to the marginal means. The average exponent over subjects is much like the average obtained from the raw monaural and binaural estimations (column 2). However, the variability found after rescaling is smaller; perhaps by rescaling and calculating marginal means, we have determined parameters of an underlying loudness function that is more similar across individuals.

**Binaural summation index.** Yet another summary of performance is provided by the binaural summation index. This index ($I'$) is the difference in decibels between equally loud monaural and binaural stimuli, obtained by averaging the separation in decibels between monaural and binaural magnitude-estimation functions over the range 20–50 dB. Given the scale with linear binaural summation of loudness, this index should...
equal 10 dB. In fact, the index \((N)\) ranged from 7.5 dB to 11.1 dB, with a mean of 9.2 dB.

The magnitude of the binaural gain in decibels is closely connected to the form and parameters of the psychophysical function for loudness. In fact, these indices of binaural summation can be converted to equivalent exponents of the underlying power function for loudness. The sone function (exponent of 0.6) signifies that every 10-dB increase in SPL entails a doubling of loudness. Complete summation also entails a doubling of loudness. Hence complete summation corresponds to a 10-dB increase in SPL. If complete loudness summation corresponded to a 20-dB increase in SPL, then loudness would double every 20 dB, and the exponent would equal 0.3.

Because linear summation means that binaural loudness is twice monaural loudness, each binaural gain \(N\) entails a corresponding power function exponent \(\beta_{\text{numm}}\), which equals \(\log(2/(N/20))\). Values of \(\beta_{\text{numm}}\) are listed in the 10th column of Table 1. These turned out to range from 0.54 to 0.80, a smaller range than the original one derived directly from the magnitude estimations. The standard deviation, accordingly, is much smaller than that derived from the original data, suggesting considerable, though not complete, uniformity in the scale values underlying the summation process. Moreover, these exponents are much like those \(\beta_{\text{num}}\) derived from the binaural summation ratio \(BR\), \(r(7) = .75, p < .05\), and are much less like the original magnitude-estimation exponents \(\beta_{\text{mest}}\), \(r(7) = .31, n.s\).

The last point is notable. It is perfectly possible for a person to have a relatively high (or low) exponent in magnitude estimation, but a low (or high) exponent based on the summation index. Everything else remaining the same, changing the range of numerical responses will change the magnitude-estimation exponent but not the summation-index exponent: The decibel difference between equally loud monaural and binaural tones remains invariant when the response scale changes. If a monaural tone of 70 dB matches in loudness a binaural tone of 60 dB, the binaural gain of 10 dB corresponds to an exponent of 0.6, and this value is wholly independent of the person's numerical responses. As Marks (1979b) argued, it is the invariance properties of loudness, the equalities and inequalities, not the vagaries of numerical response, that characterize the nature and form of the underlying representation.

### Conclusions

Results reported by several investigators (Fletcher & Munson, 1933; Heilman & Zwillocki, 1963; Jankovic & Cross, 1977; Levelt et al., 1972; Marks, 1978a, 1979b, 1980) agree well with the overall results of the present experiment: linearly additive binaural summation of the loudness of pure tones. The average magnitude-estimation exponent \(\beta_{\text{numm}}\) of 0.6, characterizing the present results, is typical of values found in the literature relating loudness judgments to the sound pressure of a 1000-Hz tone (see Marks, 1974a).

The large spread of the individual estimation exponents \(\beta_{\text{num}}\) found in the present study is also typical (e.g., J. C. Stevens & Guirao, 1964). However, when individual judgments are rescaled to be consistent with complete summation, the variability of the corresponding exponents \(\beta_{\text{mest}}\) is reduced. Still, the most striking feature of the present findings is that the subjects were significantly more alike in terms of their rule of binaural summation than in terms of their magnitude-estimation exponents. The dispersion of exponents \(\beta_{\text{numm}}\) derived from the binaural summation indices is remarkably small; certainly the large variability in the magnitude-estimation functions is not reflected in this more direct index of sensory processing. Marks (1979a) used the binaural index to find the implied loudness exponents for 15 subjects, in a study employing graphical ratings of the loudness of a 1000-Hz tone. Implicit exponents ranged from 0.50 to 0.77, similar to the 0.54–0.80 range in the present experiment. Scharf and Fishkin (1970), however, found an average binaural gain of 7 dB to 8 dB in their measurements. Given complete binaural summation (not their interpretation), this would imply exponents from 0.86 to 0.75. Although such values are compatible with some of the rescaled exponents of the present experiment, they are inconsistent with the individual data, as well as with the conclusion of the study. The nature of the underlying loudness processes is unclear.

In sum, the results of the present study suggest that sensory processes are probably very different both across persons with normal hearing and across the underlying loudness process. Hence S. S. Stevens's principle that individual differences are the way sounds are perceived, that is, the ways that people perceive different physical magnitudes.

### Experiment 2

#### Individual Differences: Magnitude Estimation

The results of Experiment 1 indicate that people vary considerably in how they are represented by the numbers to represent loudness. Therefore, it is important to examine the individual differences in the sensory variables. The individual variables of loudness are derived from the loudness estimation exponents derived from the loudness judgments to the original monaural ratios. These exponents are derived from the original loudness judgments. Moreover, they are derived from the original loudness judgments to the original monaural ratios. The loudness estimation exponents are derived from the individual sensations to the original monaural ratios.

The results of Experiment 1 indicate that people vary considerably in how they are represented by the numbers to represent loudness. Therefore, it is important to examine the individual differences in the sensory variables. The individual variables of loudness are derived from the loudness estimation exponents derived from the loudness judgments to the original monaural ratios. These exponents are derived from the original loudness judgments. Moreover, they are derived from the original loudness judgments to the original monaural ratios. The loudness estimation exponents are derived from the individual sensations to the original monaural ratios.

Using Zwicker's (1955) psychoacoustic model, one might infer that the loudness estimation exponents derived from the loudness function would also be derived from the loudness summation. Therefore, it is important to examine the individual differences in the loudness summation exponents derived from the loudness judgments to the original monaural ratios. These exponents are derived from the individual sensations to the original monaural ratios. Moreover, they are derived from the individual sensations to the original monaural ratios. The loudness estimation exponents are derived from the individual sensations to the original monaural ratios.
consistent with present results taken as a whole, as well as with other studies mentioned earlier. The reason for the discrepancy is unclear.

In sum, the results of Experiment 1 suggest that sensory representations of loudness probably vary somewhat from person to person with normal hearing, but that the typical underlying loudness scale agrees well with S. S. Stevens's (1956) tone scale and that individual differences in numerical judgment scales largely represent differences in the ways that people assign numbers to sensation magnitudes.

Experiment 2: Temporal Integration and Individual Loudness Functions

The results of Experiment 1 suggest that people vary considerably in the ways they use numbers to represent sensations but seem to vary less in the magnitudes of the sensations themselves, at least as we inferred these magnitudes from binaural summation: Interindividual variability was smaller in the exponents derived from summation (binaural: monaural ratio; summation index) than in the original magnitude-estimation exponents. Moreover, subjects were much more alike in the exponent derived from the binaural: monaural ratio and from the decibel gain in binaural summation than in the exponent of the magnitude-estimation function for loudness. The question arises whether a similarly greater uniformity holds with respect to other sorts of auditory processes, such as temporal integration.

Using Zwischen's (1969) theory of temporal summation as a starting point, one might infer that if individual differences in loudness functions do exist, these differences would also express themselves in temporal summation. In a nutshell, Zwischen's theory states that temporal summation results from a central neural process and hence that the summating mechanism operates by integrating values of loudness. The fact that temporal integration is nearly linear with respect to sound intensity and not with respect to loudness implies the presence of a neural, temporal decay process in the auditory system that is nonlinear; the nonlinearity is inversely related to the nonlinearity of the psychophysical function for loudness. If there is sizable individual variation in the loudness transformation and if this variation is independent of any variation in the decay process, then individual differences in the loudness function should be correlated at least in part with individual differences in temporal summation.

The present experiment sought to compare in several subjects the parameters of the psychophysical function (the exponent of the magnitude-estimation function) with indices of temporal summation (the slope of the trading function and the critical duration).

Psychophysical studies of auditory temporal integration demonstrated an inverse (often reciprocal) relation between stimulus intensity and duration necessary to obtain a constant response. Two indices commonly used to describe such data are the slope of the reciprocity function (the function relating log intensity to log duration) and the critical duration (the longest duration producing reciprocity). In general, the slope of the reciprocity function falls close to −1, implying a simple trade-off between intensity and duration (Port, 1963; Zwicker, 1966; also see, Scharf, 1978), though some studies report greater summation (Small, Brandt, & Cox, 1962; J. C. Stevens & Hall, 1966) and others, less (Garner, 1949; Miller, 1948). At threshold, the ear integrates the acoustic energy of a sinusoidal signal linearly up to a critical duration of about 250 ms (Algom & Babbott, 1978; Algom, Babbott, & Ben-Uriah, 1980; Garner & Miller, 1947; Watson & Gengel, 1969). Suprathreshold criteria give somewhat smaller critical durations (e.g., J. C. Stevens & Hall, 1966); integration is evident only up to about 150 ms. Despite such summary values, no universal agreement has yet been found on the exact size of the critical duration, especially with suprathreshold stimuli, nor on the exact nature and slope of the trading function (Scharf, 1978).

The purpose of the present experiment was to measure the course of psychoacoustic temporal integration at suprathreshold levels of stimulation in several subjects. It was not our intention here to go into any of the details of the many uncertainties that still mark the precise form of the time-intensity trading function or the proper interpretation of the critical duration (e.g., whether it should be...
thought of as a real discontinuity). However, these issues may be settled; the slope of the time-intensity trading function and the size of the critical duration do provide reasonable and simple summary descriptions of the empirical findings. Such summary statistics were extracted individually for each of the 8 subjects who took part in the experiment. Because all of these subjects had participated in the previous experiment, parameters of their psychophysical functions for loudness already exist.

The questions posed by the present experiment are as follows: How uniform from subject to subject are the nature of the time-intensity trading relation and the size of the critical duration? How do these measures compare with the corresponding magnitude-estimation functions? In particular, does the exponent of the psychophysical function for loudness relate to the time-intensity trading relation?

Method

Apparatus and procedure. The apparatus was the same as that of Experiment 1 except that the channel to the left ear was disconnected. Thus, all stimuli were presented monaurally to each subject’s right ear. Rise and decay times were 2.5 ms; waveforms were monitored on an oscilloscope.

The set of stimuli comprised all 42 combinations of seven durations (10–40 ms) and six SPLs (20–50 dB) of a 1000-Hz tone. These were presented 15 times at a time to the subject for judgment. Each subject was presented three replicates of the matrix in a given session (in pseudorandom order) and served in five sessions. Thus, each subject made 15 judgments per stimulus in all. Again, the method was free magnitude estimation: The subject tried to assign numbers in proportion to loudness.

Subjects. Eighty young men and women still available from the previous pool of 9 served as subjects.

Results and Discussion

Individual psychophysical functions. Geometric means of the 15 magnitude estimates were computed for each stimulus for each subject. At each duration, the geometric means were plotted in log-log coordinates against the sound pressure level in decibels, thereby producing seven magnitude-estimation functions. We took advantage of the evidence that the form of the function is the same at every duration (e.g., Small et al., 1962; J. C. Stevens & Hall, 1966; as well as Scharf, 1978) in order to arrive at a single function for each subject. The data for the seven durations were collapsed as follows. First, consecutive points on the magnitude-estimation function for the longest duration (400 ms) were connected by straight-line segments. Then the function for each shorter duration in turn was shifted (by a constant factor, in decibels) to make the function coincide with the function at 400 ms. To the complete set of adjusted data points, a threshold-corrected power function was determined by the method of least squares. (For an example, see Figure 5.)

Individual exponents appear in the second column of Table 2. These values are considerably smaller than the corresponding exponents from the magnitude estimations produced by the same subjects to 1-s stimuli in the previous experiment. The average here is 0.46, compared with the previous value of 0.6. The smaller absolute size may reflect the operation of a numerical response bias, resulting from greater difficulty in judging stimuli that vary in duration as well as in intensity (J. C. Stevens & Hall, 1966; S. S. Stevens & Greenbaum, 1966). Interestingly, the intersubject variability is also smaller; the standard deviation is only 14% of the mean.

What seems to be more important with respect to the current study is the significant correlation between the exponents in Experiment 1 (βcrit) and the exponents in the present experiment, r(6) = .77, p < .05. This result is consistent with evidence that individual differences in an exponent can remain stable over time (Barbenza, Bryan, & Tempest, 1970; Hellman, 1981; Wanschura & Dawson, 1974), suggesting that subjects tend to use numbers and numerical concepts in a rather consistent, albeit individualized, way.

Time-intensity trading. The course of psychoacoustic integration was evaluated for each subject in terms of the multiplicative factors used to collapse data for various durations onto a single function (see previous section, Individual psychophysical functions). These decibel factors themselves provide a measure of temporal summation. Shown in Figure 6 are the trading functions obtained for all 8 subjects. For all subjects, as stimulus duration increases, the intensity necessary to produce an equal-loudness judgment decreases. This trading relation does not, how-
ever, obey a rule of simple reciprocity because the linear region of each subject's function does not necessarily have a slope of $-1.0$. The slopes, based on the four shortest durations, are listed in the third column of Table 2; they range from $-0.51$ to $-1.85$, with a mean of $-1.30$. On the average, then, intensity had to be decreased by $13\, \text{dB}$ for each $10$-fold increase in duration. The average trading function is very much like those obtained by Small et al. (1962) and J. C. Stevens and Hall (1966) in their studies of suprathreshold integration. Note, though, that those studies used a wide-band noise as a stimulus, in contrast to the 1000-Hz tone used here. Typically, tones yield temporal summation with a slope closer to $-1.0$ (see Scharf, 1978).

The main point with respect to the current experiment, however, is not the absolute size of the (reciprocity function) slopes but rather the individual differences among them. The more-than-threefold variation from $-0.51$ to $-1.85$ is comparable to—indeed, greater than—the variation in the magnitude-estimation exponents of the same subjects in Experiment 1 and is considerably larger than the variation of magnitude-estimation exponents in the present experiment. Moreover, there is no correlation between the slopes of the reciprocity functions and either the present exponents ($r(5) = .04$, ns) or the exponents derived from binaural summation in Experiment 1 ($r = -.137$ to +.370).

**Critical duration.** In order to assess values of the critical duration, a straight-line segment was fitted to the first four data points and a horizontal line was fitted to the last two data points of each subject's time–intensity trading function. **Critical duration** was defined as the duration at the intersection of the line segments. (Given the arbitrariness of such a procedure, we also assessed critical duration by omitting the fourth point and/or

![Figure 5. Temporal summation: The magnitude estimates of loudness plotted as a function of relative sound pressure for 1 subject of the 8. (Each symbol represents a constant duration: $\times = 10\, \text{ms}$, $\blacktriangle = 20\, \text{ms}$, $\bigcirc = 40\, \text{ms}$, $\triangledown = 80\, \text{ms}$, $\square = 120\, \text{ms}$, $\bigtriangledown = 200\, \text{ms}$, $\bigtriangledown = 400\, \text{ms}$. The sets of points for the various durations are adjusted in relative sound pressure to maximize the fit to a single function.)](image-url)
the penultimate one. Though the absolute values were then altered, none of the correlations described later were appreciably. Critical durations appear in the fourth column of Table 2. Given the indirect determination here of equal loudness, these values should be treated cautiously. Nevertheless, a change in the slope of the respective trading function does appear in the vicinity of the values listed. The critical duration, like the slope of the reciprocity function, appears quite variable, even more variable than the loudness exponents for the same subjects. Critical durations range from 59 ms to 180 ms, with a mean of 112 ms. As Scharf (1978) pointed out, because of the large variation across data from different experiments, no firm, general conclusion is yet possible about either the slope of the trading function or the size of the critical duration. The critical duration does not correlate significantly with the exponent of the loudness function \( r = .186 \) or with various exponents derived from binaural summation in Experiment 1 \( r = -.022 \) to -.413.

Conclusions

What loudness scale underlies the temporal integration of pure tones? To begin with, there seems to be compelling evidence that the summation process in hearing takes place at a high level of the auditory system and not at the periphery (Algol & Baboff, 1984; Zwislocki, 1960). The implication is that what is summed in temporal summation is loudness rather than incoming acoustic energy. This makes apparent the critical relation between the quantification of loudness and the description of dynamic auditory phenomena, such as temporal summation. The analysis of this relation constitutes the core of several theories of auditory temporal summation (e.g., Zwislocki, 1960, 1969).

Zwislocki's (1969) analysis of suprathreshold summation is a major effort to show how the auditory system can produce linear (or close to linear) energy integration despite the nonlinear relation between sound intensity and loudness (specifically, the sone function). Zwislocki argued for the existence of a temporal decay process in neural excitation—one that counters the nonlinear transformation of sound intensity to loudness—prior to the entry of the neural signal to the loudness decider. The problem of the loudness integrator can also be considered an independent process or process in the auditory system. In this case, it turns out, a process of an energy-based sound stimulus which in a sense is needed for parameters of the system characterized by Zwislocki (1969). Indeed, modern sonometers are constructed. Thus, a process of neural firing within the auditory system that is a power function of intensity with an exponent of about 0.3 (Zwislocki, 1960).

A linear time function of intensity implies that loudness grows with a duration of 0.3, given the same sound pressure intensity. Thus, (b) there is evidence of a nonlinearity of loudness, (c) a linear loudness function, and (d) review points out (Zwislocki, 1969) that the relation for the loudness is linear, despite the powerful effect of the greater summation of the differences measured between the summation exponent and the underlying function of sound intensity. If the exponents in the same experiment are in the same range, however, no evidence in the loudness functions for each of the 8 subjects (X = Subjects 1, 3, 5, and 7; O = Subjects 2, 4, 6, and 8, respectively, from top curve to bottom curve.)
signal to the linear temporal integrator. The problem of reconciling these nonlinear neural processes with a seemingly linear integrator can also be looked at in reverse, as was historically the case: What loudness scale must we assume in order to counteract the independently demonstrable nonlinear process or processes of neural decay (e.g., Galambos, 1952; Galambos & Davis, 1943)? As it turns out, a nonlinear compression of the sound stimulus—close to S. S. Stevens's (1956) sone scale—is the transformation needed for psychoacoustic data to show the characteristics of a linear temporal integrator (cf. Penner, 1978). When Munson (1947) developed his theory of temporal summation and fixed his stimulus and weighting functions, which are almost identical to those of Zwilich (see Algol & Babb, 1984), the sone scale had not yet been constructed. Thus, (a) the nonlinear decay of neural firing with (b) the linear temporal integration of loudness imply a nonlinear, compressive psychoacoustic scale of loudness. The specific time constants involved in Processes a and b point to a loudness scale that is a power function of sound pressure with an exponent on the order of 0.6 (see Zwilich, 1969, for details).

A linear time–intensity trading relation implies that loudness must be a power function of duration with an exponent of about 0.3, given the following: (a) the sone scale for loudness, (b) a temporal decay process whose nonlinearity offsets the sone function, and (c) a linear loudness integrator. Scharf's (1978) review points to a linear time–intensity trading relation for tones as the typical result, despite the present finding of somewhat greater summation. If the individual differences measured here in the magnitude-estimation exponents represent differences in the underlying sensory transformation from sound intensity to loudness, then the differences in the exponents should relate to differences in the slopes of the reciprocity function. However, no positive relation emerged, nor did the slope of the reciprocity function relate positively to any of the loudness exponents derived, in Experiment 1, from binaural summation. (There are some modest correlations between the exponents derived from binaural summation and the critical duration of temporal integration. However, these correlations account for at most only about 25% of the variance.) One simple explanation is that the measured differences in the exponent of the magnitude-estimation function depend largely on judgmental or random factors, that individual differences in the underlying loudness function are relatively small, and that the measured differences in the slope of the time–intensity trading function largely reflect variation in the nonlinear decay process independent of the transformation from sound intensity to loudness.

Experiment 3: Judgments of Intervals and Individual Loudness Functions

Despite the large intersubject variability of the individual magnitude-estimation functions, especially in Experiment 1, the average exponent was quite consistent with the sone scale in one case (0.6), though less consistent in the other (0.46). (The somewhat lower exponent in the second experiment perhaps is attributable to the experimental task being relatively difficult.) Moreover, intrasubject consistency between the two experiments was high. In addition, the use of more direct sensory indices (e.g., the binaural summation index) revealed a more uniform input–output characteristic of the auditory system among subjects; these indices are not only consistent with the sone scale but actually entail it. However, there may be more than one perceptual relation underlying scales of sensory intensity. This possibility comes from the commonly reported nonlinear relation between judgments of loudness intervals and judgments of loudness magnitudes (S. S. Stevens, 1971; S. S. Stevens & Galanter, 1957).

That different sets of psychological values (different loudness scales) apply (a) to loudness as a magnitude and (b) to loudness relations such as differences is formalized in Marks's (1979b) theory. Several studies over the recent years demonstrated that judgments of loudness intervals or differences cannot be based on linear differences in sones (e.g., Beck & Shaw, 1967; Birnbaum & Elmasian, 1977; Schneider, Parker, Valenti, Farrell, & Kanow, 1978). These experiments usually showed well-defined metric struc-
tures in conjunction with an underlying representation for loudness that is consistent with Garner's (1954) lambda scale. Lambda is defined as follows:

\[ \lambda = k^{0.25} \]

As was the case with the sone scale, the validation of the lambda scale can be reduced to a set of invariances: Pairs of sounds that are separated by equal intervals in lambda (but not in sones) are judged to be equally different or dissimilar in loudness (Beck & Shaw, 1967; Schneider et al., 1978; see also Carterette & Anderson, 1979). This means that the sone scale and the lambda scale can be distinguished from each other solely on the basis of the principle of invariance; their simultaneous and independent coexistence was demonstrated by Marks (1979b). Thus, according to the theory, judgments based on differences (often operating in category rating methods) usually tap the lambda scale, whereas judgments of magnitudes usually tap the sone scale. Plotting the lambda scale against the sone scale results in a negatively accelerated function typical of the concave downward curve obtained when plotting rates against magnitude estimates (S. S. Stevens & Galanter, 1957). According to Mark's theory, the difference between the scales does not simply reflect different judgmental biases; the two scales entail different psychological representations of loudness.

Many studies aimed at exploring the relation between judgments of magnitudes and judgments of differences, including Mark's (1979b) attempt at combined validation, rested on group data. The questions raised by the present experiment are as follows: Does the nonlinear relation between the sone and the lambda scales underlie each subject's data? How uniform from subject to subject is this nonlinear transformation?

Method

Apparatus and procedure. Sounds were generated in a manner identical with that of Experiments 1 and 2 except for the addition of a timing device that made it possible to present tones in pairs, separating the onsets of the two tones in a pair by 2 s. Each stimulus comprised the following sequence: tone for 1 s, 1-s pause, tone for 1 s. Rise and decay times were 10 ms. As in Experiment 2, presentation was monaural (the right ear of each subject). The output of the electronic switch to the channel for the right ear was gated in sequence through Attenuators 1 and 2, serving the first and second members of a tone pair, respectively.

The stimulus pairs were generated from a 7 × 7 (First Tone × Second Tone) factorial design. This gave 49 different stimuli (pairs of tones) in all. The seven levels were 20, 27, 34, 40, 48, 54, and 60 dB. Pairs were presented at a time to subjects for judgment. Each session consisted of four replications of one half of the stimulus matrix; either the more intense tone of a pair always appeared first (or both members had the same intensity) or always appeared second. Thus each session contained four replications of 28 different stimuli (combinations on the diagonal — equally intense stimulus pairs — were always included). The order of presentation of the two halves of the matrix alternated between sessions for each subject. Within a session, the order of presentation was pseudorandom and was different for each subject. Each subject served in four sessions, thus giving 16 judgments per stimulus pair in all.

The method used was what S. S. Stevens (1958) called "interval estimation" or what is otherwise known as magnitude estimation of intervals or differences. The subjects were told that every stimulus would contain a sequence of two tones and that the task was to try to assign numbers in proportion to the size of the interval or difference in loudness between the tones.

Subjects. Five subjects still available from the original pool of 9 took part in this experiment.

Results and Discussion

For each stimulus for each subject, the geometric mean was calculated across the numerical judgments. Differences of 6 dB or 8 dB occasionally yielded estimates of 0. To avoid geometric means of zero in such cases, we resorted to a rule of discarding, for every zero, the highest estimation given to that stimulus pair and calculating the geometric mean from the truncated distribution. Presented in Figure 7 are the geometric means computed in this manner from data normalized and pooled over subjects. The mean estimates are plotted against the level of the second stimulus; each curve represents a different SPL of the first stimulus. We use the convention of assigning a negative value when the first tone was, on the average, judged louder. For example, the difference in loudness of the 40–20 dB pair was judged as -2.8, on the average, when presented in that order; but the difference was judged as -1.7 when presented in the reverse order (20–40 dB). Three of the 5 subjects gave slightly higher estimates to the presentation "loud-soft" than to the presentation "soft-loud" of the same pair of tones. For these 3 subjects, the mean of the judgments differed significantly from zero. However, the number of subjects was too small to ensure the presence of a significant interaction effect between tone order and SPL. The mean judgments for tone order at a constant SPL were all positive, indicating that the difference between loudness judgments for the same pair of tones of tones was always 20 dB or more. This effect was small in comparison with the variance in loudness judgments of the individual subjects.

If the subject is aware of the instructions, the subtraction structure of judgments represents a psychological representation of the subtractive model. This is illustrated in Figure 7 such that the curves do appear to be parallel to one another by a common variability (mainly due to the lack of variability in tone order effect results in the judgments of the individual subjects appearing in different orders)

If the subject is unaware of the instructions, the subtraction structure of judgments represents a psychological representation of the additive model. This is illustrated in Figure 7 such that the curves do appear to be parallel to one another by a common variability (mainly due to the lack of variability in tone order effect results in the judgments of the individual subjects appearing in different orders).
the mean of the log \( [(\text{loud} - \text{soft})/\text{soft} - \text{loud}] \) judgments differed significantly from zero. However, the magnitude of this order effect, which is opposite to the usual direction of the time-order error, is small; the mean difference between the two orders never exceeds the mean judgment given to the least different pair of tones. That the average order effect was small is evident from visual inspection of the curves in Figure 7. Possibly the order effect resulted from small differences in judgments of the half-matrices, as these were presented in different sessions.

If the subjects behaved in compliance with the instructions, the data should conform to a subtractive structure operating on the psychological representations of loudness. A subtractive model predicts that the curves in Figure 7 should be parallel. Indeed, the curves do appear parallel; neighboring pairs of functions appear to be separated from one another by a constant amount. Despite some variability (mainly because small differences too often were judged as zero), the overall form of the functions suggests that a subtractive structure holds for the pooled data, at least to a first-order approximation. An ANOVA confirmed this conclusion; the interaction term was nonsignificant, \( F(36, 196) = 1.19 \).

What loudness scale underlies these judgments of loudness intervals? To answer this question, we computed marginal means of the judgment matrix for the first stimulus and for the second stimulus and then averaged the two sets. If the difference judgments are based on sones, then the marginal means should relate linearly both to the corresponding marginal means from the matrix of judgments of the binaural summation study (Experiment 1) and to the monaural and binaural psychophysical functions derived from those judgments. This should be so because (a) marginal means are estimates of the underlying scale values and (b) the loudness scale underlying the judgments in binaural summation has already been shown to be consistent with the sone scale. Because an additive model was shown to hold for the judgments in Experiment 1, psychophysical functions, though constituting only a subset of those data, should parallel the corresponding marginal means. They do not. For all subjects, th
means of the difference judgments are nearly a linear function of SPL, whereas the marginal means and the psychophysical functions from binaural summation are more accelerated functions of SPL. Because the magnitude estimates and the rules of binaural summation were shown to be consistent with S. S. Stevens's scale, it follows that the loudness scale underlying the interval judgments is not the same scale but is a nonlinear function of it. Equal differences in tones did not yield equal judgments of loudness intervals; clearly it is a violation of the principle of invariance to assume that the same scale underlies these judgments.

The present marginal means define an underlying scale unique only up to multiplication by a positive constant and addition of a constant (an equal-interval scale). It is possible to extract psychophysical power functions from such data by assuming a power function and, using an iterative least squares procedure that operates on log (scale value minus constant) and log stimulus, estimating the true zero point of the scale for each subject. Individual scales are plotted with respect to sound pressure in Figure 8, and the values of the slopes (exponents of power function) are listed in Table 3. The fits to the power functions are excellent. The most striking feature of the results, however, is their reasonable agreement with the lambda scale (whose exponent = 0.26). The average exponent is 0.21 but with considerable variation among these 5 subjects. The magnitude-estimation exponents, marginal-mean exponents, and summation-index exponents of the same subjects from Experiment 1 are presented for comparison. The propriety of calculating marginal means here, one should recall, rests on the additive structure of the data matrix. In fact, as indicated in Table 3, there was a small but distinct and reliable nonadditivity in the data of 2 subjects. Although the nature of the deviations from additivity did not appear to be consequential to these analyses, we still deemed it appropriate to examine the results without relying on the data's metric properties. Thus for comparison, power functions were also fitted to scales derived by nonmetric analysis (TOSCA-9, a routine that generates interval scales from rank orders among pairs, applied to the rank order of each subject's estimates). For all subjects, the results of the nonmetric analysis strongly resemble those of the metric analysis (see Table 3).

It is interesting that the exponents derived from the application of judgments of loudness differences should yield about as much variability (about 25% of the mean) as the exponents derived from the numerical judgments of loudness magnitudes. Only the exponents derived from the binaural summation index show notably smaller variation (as do the magnitude-estimation exponents of Experiment 2), a result that suggests the possibility that the sensory representation for loudness is more uniform than is the representation underlying loudness-difference judgments. Hence there may be significant nonlinear components of the processes involved in generating loudness differences. The findings suggest that in processing, binaural summation precedes the generation of

---

**Figure 8.** Loudness intervals: The loudness functions derived from the marginal means of the magnitude estimates of loudness intervals plotted against sound pressure level (SPL) for each of the 5 subjects. (O = Subjects 1 and 4, Δ = Subjects 2 and 5, and □ = Subject 3, respectively, from bottom curve to top curve.)

---

**Table 3**

<table>
<thead>
<tr>
<th>Subject</th>
<th>(\lambda)</th>
<th>(\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.23</td>
<td>0.25</td>
</tr>
<tr>
<td>2</td>
<td>0.22</td>
<td>0.24</td>
</tr>
<tr>
<td>3</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>4</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>5</td>
<td>0.26</td>
<td>0.26</td>
</tr>
<tr>
<td>(M)</td>
<td>0.24</td>
<td>0.25</td>
</tr>
<tr>
<td>(SD)</td>
<td>0.02</td>
<td>0.02</td>
</tr>
</tbody>
</table>

Note. --- = data not available. \(\lambda\) = exponent of marginal means, \(\beta\) = exponent of Lambda scale derived by nonmetric analysis. * \(p < .05\).
the lambda representation. If so, this would provide evidence contrary to the models of Garner (1954) and Birnbaum (1982), who proposed that summation was subsequent to the representation of lambda.

Other experiments dealing with interval judgments have also yielded representations of loudness like the lambda scale (Beck & Shaw, 1967; Parker & Schneider, 1974; Schneider, Parker, & Stein, 1974; Schneider et al., 1978). (One exception was Popper’s, 1983, study showing that under a certain condition—where subjects matched the loudness of a tone to the size of loudness intervals—the underlying scale resembled some more than it resembled lambda.) However, considerable intersubject variability was evident in some of these studies. In Garner’s (1954) original study, in which the lambda scale was derived, subjects differed considerably in parameters of their loudness function. Schneider (1980) used a method of direct comparisons of intervals, thereby avoiding numerical judgments altogether, and found that the loudness scales differed markedly from subject to subject: a range of exponents over 2:1, comparable to the present results.

Some investigators have interpreted the scales based on the usually well-defined metric structures resulting from interval judgments as the perceptual scale, rejecting the validity of other representations (e.g., Birnbaum & Elmasian, 1977), though the possibility of multiple representation is acknowledged (Anderson, 1974). Considerations based on mathematical convenience (e.g., Birnbaum & Elmasian, 1977) cannot by themselves form the basis for deciding which values are loudnesses. Criteria are needed beyond mere metric structure. Measures of substantive sensory performance can serve as such criteria, providing a framework for models of sensory—perceptual processing and pointing toward the possibility that metric structures can operate at different levels of the auditory system, acting on different sets of psychological values (Marks, 1979b).

In sum, the present results suggest that individual sensory representations for intervals of loudness vary from person to person with normal hearing and that on the average, the underlying loudness scale agrees fairly well with Garner’s lambda scale, even though the same individuals use a scale closer to sone to represent loudness as magnitude.

General Discussion

Scales of Loudness

The results of these experiments imply that one can speak of at least two different scales of loudness, both of which can display clear metric structures. One, called the L scale, operates in binaural summation (Experiment 1; Marks, 1978a, 1979b), presumably operates in temporal summation (Experiment 2), and also presumably operates in the summation of multicomponent, narrow-band sounds where the components...
are widely separated in frequency (Marks, 1979b). The \( L \) scale closely resembles S. S. Stevens's sone scale, approximating a 0.6 power function of sound pressure. The other scale, the \( D \) scale, operates when subjects judge loudness differences or loudness intervals (Experiment 3; Marks, 1979b). The \( D \) scale closely resembles Garner's (1954) lambda scale, which is approximately a 0.26 power function of sound pressure. Hence, on the average, \( D \) is about the square root of \( L \).

**Problems of Validation**

Are there two (or more) distinct representations for loudness? The present experiments suggest there are. On the other hand, because judgments of differences and judgments of ratios typically have the same rank order, some investigators have suggested that both of these tasks induce subjects to base their judgments on a single perceptual relation, specifically, the interval-scale representation (e.g., Anderson, 1972; Birnbaum & Veit, 1974; Torgerson, 1961). This outcome perhaps reflects most directly on the method of ratio estimation (e.g., Rule, Curtis, & Mullin, 1981). In fact, ratios of loudness (\( L \)), as derived from binaural summation in the present study, are not ordinarily identical with differences in \( D \) (see Marks, 1979b). The pattern of individual differences implies a transformation to the \( L \) scale, followed by summation, and then followed by a transformation to the \( D \) scale.

The fundamental issue is validation. What is needed to serve as a validational base is sensory theory together with sensory data. As the present experiments demonstrate, the \( L \) and the \( D \) scales have their own internally consistent structures (one additive, the other subtractive), and these structures are independent of the metric properties of the numerical estimates used by the subjects. Thus the difference between the scales rests on the nature of sensory reality, not on the vagaries of numerical responses. Moreover, individual variations differ in the two scales in a fashion consistent with the view that the \( L \) scale precedes the \( D \) scale.

The reluctance of some researchers to give up their claim for a unirepresentational scheme may stem, at least in part, from the difficulty in accepting the possibility that the task affects the fundamental sensory response. Such is not the only—probably not the correct—interpretation. Certainly a change in task does not influence the sensory activity in primary auditory receptors on the basilar membrane; task is unlikely to modify initial transduction processes or auditory-nervus function. Still, the \( L \) scale and the \( D \) scale presumably do exist simultaneously and independently: as organizational somewhere in the auditory system. The activation of one scale or the other depends on the task.

**Individual Differences**

The data presented in this article with previous ones (Marks, 1979b) were collected over several years. For the sake of brevity, scales that were used in the experiments that are not relevant to the sensory properties of the \( L \) and \( D \) scales are left out. However, the data presented here lend support to the view that the sensory properties of the \( L \) and \( D \) scales are different. The question arises, then, what might account for these differences? One possibility is that the differences are caused by differences in individual difference scores, or, more precisely, why do individual differences differ? Another possibility is that the differences are caused by differences in the nature of the task presentations.
imposed on the subject (Marks, 1979b). The D scale may operate in loudness discrimination (Parker & Schneider, 1980) and in simple reaction time (Kohfeld, Santee, & Wallace, 1981; Marks, 1974a; McGill, 1961), as well as in loudness—difference or dissimilarity; the L scale, in additive masking (Lutf, 1983), in combination tones and in two-tone suppression (Duijshuis, 1976), as well as in loudness summation.

The idea that the same set of stimuli assumes different sets of psychological values depending on the task is neither especially startling nor really new. It is an accepted assumption in the related domain of the psychology of similarity (e.g., Krumhansl, 1978; Tversky, 1977). Thus, subjects judge North Korea to be much more similar to the Soviet Union than the Soviet Union is to North Korea; the mere wording of a statement changes the relevant psychological values. Germany and France seem much more alike when Brazil is included in the matrix of countries to be judged for similarity than when only European countries are included. The rank order of the same stimuli can be changed when judged in conjunction with different sets of other stimuli. The point is clear: The psychological representation, the scale, may in many circumstances depend on so-called task variables (see also Eisler, 1963).

Individual Differences and Sensory Processing

The theoretical stance taken throughout this article and urged by the second author over several years (Marks, 1972, 1974b, 1979b, 1982) argues that psychophysical scales take on their meaning and their value in the elaboration and the elucidation of sensory processes. It is instructive, if only from a heuristic vantage, to inspect individual differences in this respect. When we ask, "Do individual psychophysical functions differ?" or, more appropriately, "How much do they differ?" we may well be seeking two different kinds of answers: a quantitative one (that, for instance, the standard deviation is so much a percentage of the mean) and a qualitative one (that functions vary "very little" or "somewhat" or "much"). The latter, relativistic, assessment demands a yardstick. What better comparison than individual variation in other measures of sensory processing, especially in parameters that describe the quantitative characteristics of intensity processing in the sense modality in question? To the question, "What is a 'large' variation in psychophysical function?" we may parry the question, "What is the variation in binaural summation? in critical duration? in time-intensity reciprocity?"

The point is, our scientific judgment as to the magnitude of individual variation in a psychophysical scale should consider variation in other sensory capacities. That the quantitative character of temporal integration, for example, can vary so widely from person to person may give one pause. Variation in such properties could, in principle, bear on differences in a loudness scale. A greater degree of reciprocity or a larger period of summation could, ceteris paribus, in theory yield a greater range of loudness, hence a larger loudness exponent. Even more significant, the very fact that a fundamental sensory process such as summation can fluctuate so much from one person to another puts into perspective individual differences in loudness scales: Compared with the variation in temporal summation, as well as with the variation in magnitude-estimation scales, the scale underlying binaural summation seems remarkably stable over individuals.

In general, much of the variability in overt judgment scales (scales produced by methods like magnitude estimation) probably results from differences in the mode of judgment (see also Teghtsoonian & Teghtsoonian, 1983). This view receives support from the present series of experiments, which suggests that individual representations of loudness are reasonably, though probably not wholly, uniform from person to person. This implies that magnitude estimation and production may on the average yield linear representations of the L loudness scale, though any given overt judgment scale is likely to exhibit bias (see also, Hellman, 1981). To be sure, magnitude-estimation scales of loudness show diversity, individuality, and idiosyncrasy; but beneath lies a common core of uniformity in sensory-perceptual processing of sound intensity and, at one stage of processing at least, in the underlying scale for loudness.


