

Additivity of loudness across critical bands: A critical test

RONALD HÜBNER and WOLFGANG ELLERMEIER
Universität Regensburg, Regensburg, Germany

The use of magnitude estimation as well as axiomatic measurement theory has led to the suggestion that loudness adds across critical bands. In the present paper, we challenge this postulate by applying a more sensitive methodology, based on Falmagne's (1976) random conjoint measurement procedure. A necessary condition for additivity of loudness was investigated in tone complexes consisting of 2-kHz and 5-kHz (resp. 3-kHz) components; the results showed systematic deviations from additivity. We argue that these deviations are due to asymmetric masking of the higher component by the lower one, and we propose a tentative quantitative model to account for the data. Such a model is in line with results from tone-on-tone masking, which show masking to be effective over a range of several critical bands.

A problem central to models of loudness and noise evaluation is to determine how different frequency components combine to produce what is perceived as the overall loudness of a sound. In the most elementary situation conceivable, that of a two-tone complex, classical investigations (Fletcher & Munson, 1933; Scharf, 1961a, 1961b, 1970; Zwicker, Flottorp, & Stevens, 1957) showed that loudness remained constant as long as two sinusoidal components (or noise bands) fell within a *critical band*—a narrow range of frequencies, the width of which is roughly 1/8 of the center frequency. When the components were farther apart, however, overall loudness increased. On the basis of these results, it was postulated that although energy sums within the critical band, the loudness of the frequency components sums when the components fall into different critical bands or auditory filters.

Closer inspection of the relevant literature, however, reveals that this generally accepted postulate of loudness summation across critical bands rests on relatively weak methodological grounds. Most of the data that strongly suggest additivity were obtained with the use of direct scaling methods such as *magnitude estimation* (see, e.g., Marks, 1978, 1979). But this approach does not permit one to disentangle the effects of the psychophysical functions involved or of the way the auditory channels are combined. If the overall loudness, L , is produced by two components with intensities a and x —that is,

$$L(a, x) = F[l_1(a), l_2(x)],$$

with F denoting the combination rule (additive or other),

and l_1, l_2 denoting psychophysical functions for the two frequency components—then it must remain indeterminate whether an outcome suggesting nonadditivity, for example, is due to a nonadditive combination rule, or to inadequately chosen psychophysical functions. Marks (1979), for example, drew the latter conclusion, and, not taking his subject's numerical responses at face value, transformed them to produce additivity. The complementary strategy, however, of assuming a specific psychophysical function (such as the sone scale) and then trying to determine the combination rule is equally conceivable. The problem remains that with the use of direct scaling methods, the combination rule cannot be determined independently of assumptions about the psychophysical functions, or vice versa.

One way out of this dilemma is formulated in axiomatic measurement theory, which permits one to test for additivity without having to make assumptions about the psychophysical functions. All that is required is to test a number of qualitative conditions implying additivity. These are formulated in Luce and Tukey's (1964) theory of *conjoint measurement*.

In this framework, additivity of loudness across critical bands may be formulated as follows: If a, b , and c are intensities of a 2-kHz tone, for example, and if x, y , and z are intensities of a 5-kHz tone, the total loudness of a two-tone complex may be written as

$$L(a, x) = l_2(a) + l_5(x).$$

Or, in a paired comparison task, a two-tone complex with intensities a and x at 2 and 5 kHz is judged as equally loud as one with intensities b and y at these frequencies if and only if the sums of the component loudnesses are equal:

$$(a, x) \sim (b, y) \text{ iff } l_2(a) + l_5(x) = l_2(b) + l_5(y). \quad (1)$$

For additivity to hold, a number of conditions must be fulfilled, one of which is a cancellation law known as the

Portions of this research were presented at the 33rd Meeting of Experimental Psychologists (33. TeaP) held at Regensburg, Germany, April 9–12, 1990. Correspondence can be addressed to R. Hübner, Psychologisches Institut, Technische Universität Braunschweig, Spielmannstr. 19, D-3000 Braunschweig, Germany.

Thomsen condition (Krantz, Luce, Suppes, & Tversky, 1971):

If

$$(a,y) \sim (b,z), \quad (2)$$

and

$$(b,x) \sim (c,y), \quad (3)$$

then

$$(a,x) \sim (c,z). \quad (4)$$

It is easy to show that this must hold, by substituting the component loudnesses according to Equation 1 (see Krantz et al., 1971).

Schneider (1988) tested the preceding cancellation condition for paired comparisons of two-tone complexes consisting of 2- and 5-kHz components. He found that these conditions were violated in less than 5% of all tests, and he concluded that loudness is additive across widely separated frequencies. The problem with the original conjoint measurement theory, however, is that it is formulated deterministically. Thus, it remains unclear how many violations are tolerable, since no criterion for the rejection of additivity is given.

This is why Falmagne (1976) proposed *random conjoint measurement*, a probabilistic version of the original approach that treats subjects' judgments or adjustments as random variables, thus permitting statistical tests of whether violations of additivity are significant. Details of the theoretical derivations can be found in Falmagne; it will suffice here to delineate the procedural steps necessary for an investigation of loudness summation across critical bands.

To test the cancellation condition, the equivalence relations postulated in Equations 2-4 are tested in a very straightforward manner. The subject matches a standard stimulus with fixed intensities at the two frequency components (left sides of Equations 2-4) to a comparison stimulus, for which one of the components (b or c in Equations 2-4) is made variable. Thus, in Step 1, the subject is to find a loudness match between a two-tone complex with intensities (a,y) and another complex with the same component frequencies, but with intensities (b,z) . Since a , y , and z are fixed by the experimenter, and b is under the subject's control, repeating Step 1 a number of times yields a distribution of b (see Equation 2) with median b_{med} .

In Step 2, the subject must find a match between (b_{med}, x) and (c,y) , with c variable and all other intensities fixed (see Equation 3). This yields a distribution of c with median c_{med} .

In Step 3, the subject must find a match between (a,x) and (c,z) with a , x , and z fixed, and with c variable (see Equation 4). This yields another distribution of c , with median c'_{med} .

If additivity holds, there should be no significant difference between the two distributions of c with medians c_{med} and c'_{med} .

Although Falmagne's (1976) approach constitutes a valuable theoretical and methodological advance, it has seldom been applied. Falmagne reported largely illustrative data on binaural loudness summation obtained from a single subject, and the only extensive study done with the approach was reported by Gigerenzer and Strube (1983), who also investigated binaural additivity of loudness. Both studies have in common that a standard method of adjustment was used to test the crucial equivalence relations corresponding to Steps 1-3; this method is known to be less sensitive, and at greater risk for a number of biases than are the preferable adaptive procedures (see Jesteadt, 1980; Levitt, 1971).

Thus, in the present study, we employed an adaptive procedure to test the cancellation axiom. Furthermore, this study constitutes the first application of Falmagne's (1976) random conjoint measurement approach to the issue of loudness summation across critical bands.

EXPERIMENT 1

Method

Subjects. Seven normal listeners served as subjects. All had hearing sensitivity within 10 dB of the ANSI (1970) standard at the frequencies used in the experiment.

Apparatus. The stimuli were generated by means of a 12-bit D/A converter (32-kHz sampling rate), which was controlled by a personal computer. The signals were attenuated by a two-channel attenuator (two AD 7111), which was also under computer control. Subsequently, the signals were added, lowpass filtered at 15 kHz, amplified by a standard amplifier, and presented to the subject's right ear through headphones (Beyer DT-48A). An artificial ear (Brüel & Kjaer Type 4153) and a measurement amplifier (Brüel & Kjaer Type 2607) served for calibration. The subjects were seated in a double-walled soundproof room. A control panel with two LEDs indicated which signal corresponded to each of the two buttons.

Stimuli and Procedure. Two-tone complexes consisting of a 2-kHz and a 5-kHz component set at different intensities served as the stimuli. They had a duration of 400 msec, with 10-msec rise/fall times. Each trial consisted of the presentation of the standard and the comparison stimulus (containing the adjustable component) separated by an interstimulus interval of 400 msec. The order of presentation of the standard and comparison stimuli was randomized. The subjects had to indicate which of the two complexes sounded louder, by pressing one of two buttons. There was no time limit for the decision. After an intertrial interval of at least 1 sec, the next trial began.

A two-interval, adaptive forced-choice procedure was used to find stimuli of equal loudness. The procedure followed a one-up-one-down rule, and the step size was changed according to the accelerated decision rule proposed by Falmagne (1985). The starting step size of 4 dB was reduced after each turning point (a change from an increase to a decrease in intensity or vice versa) until a minimum step size of 1 dB was reached. The following six turning points were averaged to obtain an estimate of the point of subjective equality (PSE). In one block of trials, lasting approximately 5 min, four such "tracks" providing individual estimates of the PSE were randomly interleaved. In one session, subjects usually completed three trial blocks, which were separated by short rest breaks.

The selection of the fixed intensities followed the steps outlined in the introduction, with a , b , and c denoting the intensities of the 2-kHz component, and with x , y , and z , the intensities of the 5-kHz component; a was set at 50 dB; x , at 66 dB; y , at 60 dB; and z , at 50 dB (SPL). The order of Steps 1-3 was counterbalanced across

Table 1
Test of the Cancellation Axiom, Using 2-kHz and 5-kHz
Components With the 2-kHz Component Variable

Subject	Medians			Step 3 - Step 2 ($c'_{med} - c_{med}$)
	Step 1 b_{med}	Step 2 c_{med}	Step 3 c'_{med}	
R.H.	54.5	58.59	59.75	+1.16
W.E.	56	61.47	62.78	+1.31*
P.B.	53.3	57.37	60.75	+3.38†
P.F.	55.59	59.34	60.03	+0.63
N.N.	57.88	62.78	64.84	+2.06*
R.O.	59.75	66.97	67.09	+0.12
Mean deviation ($c'_{med} - c_{med}$)				+1.44

Note—All values are given in decibels (SPL). * $p < .05$. † $p < .001$.

subjects as much as possible within the constraints of the procedure. Thirty-six estimates of the PSE were obtained from each subject in Step 1, and 24 estimates in Steps 2 and 3, requiring a total of seven sessions per subject to complete a test of the cancellation axiom.

Results

Table 1 shows the median levels of the variable stimulus components producing a match in the adaptive procedure in Steps 1-3 of the cancellation condition. The crucial information to be gained from this table is a comparison between the matches produced in Step 2 and Step 3 (c_{med} and c'_{med} , respectively). For additivity to hold, these matches should essentially coincide, as for Subject R.O. Most subjects, however, show a small, but systematic difference averaging about 1.44 dB (see the last data column). To determine whether these differences were statistically significant, for each individual subject the matches obtained in Step 2 were compared with those obtained in Step 3 by means of a median test, as in Falmagne (1976). This test indicated a significant difference, and thus a violation of additivity, for 4 of the 6 subjects in the condition involving 2-kHz and 5-kHz components.

Furthermore, the observed departures from additive behavior appear to be systematic; deviating in the same direction, they suggest an interaction between the components due to some perceptual or judgmental process.

EXPERIMENT 2

Although Experiment 1 produced evidence for systematic violations of additivity, one might suspect these to be due to experimental peculiarities. We were particularly worried about two sources of bias. First, we suspected that subjects might be able to disregard or discount the very high-pitched 5-kHz component. Furthermore, the fact that it was always the low-frequency component that was variable in these tests might conceivably produce some sort of bias. Thus, in a second experiment, we tested the cancellation condition for component frequencies of 2 and 3 kHz, and in two sets of conditions made either the high-frequency or the low-frequency component variable.

Method

The apparatus and general procedure were identical to those in Experiment 1. When the 2-kHz component was variable, the stimulus levels and all details of the procedure were as described in Experiment 1, except that the frequency of the higher component was changed from 5 to 3 kHz. In a second set of conditions, the 3-kHz component was made variable simply by exchanging the frequencies assigned to the intensity components of the stimuli. Thus, in these conditions, a , b , and c refer to intensities of the (variable) 3-kHz component, and x , y , and z denote intensities of the (fixed) 2-kHz component. Other than this, the levels chosen and the steps of the procedure were identical to those in Experiment 1.

Results

The results of the 2-kHz-variable condition are given in Table 2. They essentially mimic those of the first condition. Again, 2 of the 4 subjects show significant deviations from additivity, and the direction and magnitude of the departure are essentially identical to what was observed with the 2-kHz and 5-kHz components.

With the 3-kHz component variable, the results given in Table 3 were obtained. Again, the matches made by 1 of the subjects show evidence of violations of additivity, and the matches made by 2 of the remaining subjects deviate in the same direction. Note, however, that a simple bias due to the adjustable component should have produced identical patterns of matches in the two stimulus configurations, whereas the reversal of the difference between c_{med} and c'_{med} suggests a more substantial effect.

DISCUSSION

The results of the present investigation suggest that additivity of loudness across critical bands should be re-

Table 2
Test of the Cancellation Axiom, Using 2-kHz and 3-kHz
Components With the 2-kHz Component Variable

Subject	Medians			Step 3 - Step 2 ($c'_{med} - c_{med}$)
	Step 1 b_{med}	Step 2 c_{med}	Step 3 c'_{med}	
R.H.	58.35	64.28	65.94	+1.66*
W.E.	57.3	60.48	62.78	+2.3†
P.B.	57.53	63.97	64.28	+0.31
S.K.	61.31	67.69	68.19	+0.50
Mean deviation ($c'_{med} - c_{med}$)				+1.19

Note—All values are in decibels (SPL). * $p < .01$. † $p < .001$.

Table 3
Test of the Cancellation Axiom, Using 2-kHz and 3-kHz
Components With the 3-kHz Component Variable

Subject	Medians			Step 3 - Step 2 ($c_{med} - c'_{med}$)
	Step 1 b_{med}	Step 2 c_{med}	Step 3 c'_{med}	
W.E.	63.69	71.09	69.72	-1.37*
P.P.	57.42	62.84	62.31	-0.53
S.K.	58.43	63.39	61.77	-1.62
C.P.	56.78	62.09	62.09	0
Mean deviation ($c_{med} - c'_{med}$)				-0.88

Note—All values are in decibels (SPL). * $p < .01$.

jected, since a necessary condition for additivity was not met in a significant number of tests. This outcome disagrees with the findings of Schneider (1988), who did not reject additivity. Note, however, that while Schneider applied the (deterministic) concept of *conjoint measurement*, in the present investigation we used Falmagne's (1976) *random conjoint measurement* methodology, which focuses on a few crucial tests to be performed. Combined with the criterion-free adaptive forced-choice procedure used in the present experiments, this methodology may be more sensitive than alternative approaches in detecting violations of additivity. In this respect, the present situation is analogous to that encountered in binaural loudness summation, which the classical conjoint measurement approach suggested to be additive (cf. Levelt, Riemersma, & Bunt, 1972), whereas its probabilistic version found systematic violations of additivity (Falmagne, 1976; Gigerenzer & Strube, 1983).

Since the observed deviations are systematic, we cannot merely reject additivity; we are also encouraged to speculate about processes that might explain the nonadditive outcome. A closer look at the literature on masking reveals that there is ample evidence for tone-on-tone masking to extend across several critical bands (e.g., Egan & Hake, 1950). Even at the suprathreshold levels with which we were concerned in the present experiment, there is partial masking, with one component affecting the component loudness of the other (cf. Zwicker & Fastl, 1990)—an effect that was taken into account in early attempts to model the loudness of multicomponent tones (Fletcher & Munson, 1933; Garner, 1959). Hence, we suspect that some sort of masking effect may have caused the nonadditivity evident in our data.

To model this effect, several possible interaction models were considered. The model that accounted for our data best is given by

$$L(a, x) = f_1(a) + f_h(x)f_1(a)^m,$$

where $f_1(a) = na^r$ and $f_h(x) = nx^s$, with $n, r, s > 0$, are power functions transforming the intensities of the lower and higher frequency components, respectively. As one referee pointed out, this model is meaningful only if n is assumed to be a dimensional constant, the value of which depends on the exponents r, s , and m .

Our model implies that the lower component masks the higher, but not vice versa, which is consistent with the "asymmetry of masking" generally found in the studies cited above. Since Subject W.E. participated in both conditions, he produced data for the 2-/3-kHz configuration (Experiment 2) with both the lower and the higher component variable. We therefore fitted the model to his data and obtained the following parameters: $n = 4.49$, $r = 0.30$, $s = 0.52$, and $m = -1$. If we use these parameters to calculate data entries for the tables, we get 57.32, 62.02, and 62.59 dB (SPL) for b, c , and c' , respectively, with a difference of 0.57 ($c' - c$) dB for the 2-kHz-variable condition, and 63.21, 70.80, and 69.71 dB (SPL), with a difference of -1.09 dB for the 3-kHz-variable condi-

tion. Although the predicted deviations from perfect additivity are in the same direction as the empirical ones, the differences between Step 2 and Step 3 are underestimations somewhat, particularly for the 2-kHz-variable condition, in which the prediction is only 0.57 dB when in fact a difference of 2.3 dB was obtained. However, this is mainly due to a large deviation in Step 2.

If asymmetric masking was indeed responsible for the present outcome, one would expect there to have been more masking between the 2-kHz and 3-kHz components than between the 2-kHz and 5-kHz components. This was the case for 2 of the 3 subjects, who participated in both Experiments 1 and 2 (cf. Tables 1 and 2). Interestingly, although they did not permit strict tests of additivity, some of the early studies of loudness summation (e.g., Scharf, 1961b, 1970) revealed a possibly related dependence on the frequency separation, in that the total loudness of two tones continued to increase as they were separated by more and more critical bands (see Scharf, 1970, Figure 3), with not even a tendency to asymptote evident in the data. Clearly, this behavior is consistent with the across-critical-band masking explanation suggested by the present experiment. Furthermore, it suggests that it may simply be difficult to find conditions of spectral additivity unspoiled by masking, even if the component sounds are several thousand hertz apart.

It should be noted that the suggested model is only a tentative step in constructing a general model of loudness summation, and that it is obviously only appropriate for two-tone masking situations, like those realized in the present experiment. It would be preferable to have a more general model that would make reasonable predictions for stimuli comprising one component, or more than two components, as well. The model suggested here has the unreasonable characteristic that if the lower component has no energy, the predicted loudness is not defined.

To conclude, it seems that although the additivity of loudness across critical bands is an acceptable first approximation, the use of a more sensitive methodology reveals systematic violations of additivity. These can most likely be attributed to well-known masking effects, which should be incorporated into a general model of loudness summation.

REFERENCES

- ANSI (1970). *Specifications for audiometers* (ANSI S3.6-1969). New York: American National Standards Institute.
- EGAN, J. P., & HAKE, H. W. (1950). On the masking pattern of a simple auditory stimulus. *Journal of the Acoustical Society of America*, *22*, 622-630.
- FALMAGNE, J.-C. (1976). Random conjoint measurement and loudness summation. *Psychological Review*, *83*, 65-79.
- FALMAGNE, J.-C. (1985). *Elements of psychophysical theory*. New York: Oxford University Press.
- FLETCHER, H., & MUNSON, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, *5*, 82-108.
- GARNER, W. R. (1959). On the lambda loudness function, masking, and the loudness of multicomponent tones. *Journal of the Acoustical Society of America*, *31*, 602-607.

- GIGERENZER, G., & STRUBE, G. (1983). Are there limits to binaural additivity of loudness? *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 126-136.
- JESTEADT, W. (1980). An adaptive procedure for subjective judgments. *Perception & Psychophysics*, **28**, 85-88.
- KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement: Vol. 1*. New York: Academic Press.
- LEVELT, W. J. M., RIEMERSMA, J. B., & BUNT, A. A. (1972). Binaural additivity of loudness. *British Journal of Mathematical & Statistical Psychology*, **25**, 51-68.
- LEVITT, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, **49**, 467-477.
- LUCE, R. D., & TUKEY, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, **1**, 1-27.
- MARKS, L. E. (1978). Phonion: Translation and annotations concerning loudness scales and the processing of auditory intensity. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3, pp. 7-31). Hillsdale, NJ: Erlbaum.
- MARKS, L. E. (1979). A theory of loudness and loudness judgments. *Psychological Review*, **86**, 256-285.
- SCHARF, B. (1961a). Complex sounds and critical bands. *Psychological Bulletin*, **58**, 205-217.
- SCHARF, B. (1961b). Loudness summation under masking. *Journal of the Acoustical Society of America*, **33**, 503-511.
- SCHARF, B. (1970). Critical bands. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, pp. 159-202). New York: Academic Press.
- SCHNEIDER, B. (1988). The additivity of loudness across critical bands: A conjoint measurement approach. *Perception & Psychophysics*, **43**, 211-222.
- ZWICKER, E., & FASTL, H. (1990). *Psychoacoustics: Facts and models*. Berlin: Springer-Verlag.
- ZWICKER, E., FLOTTORP, G., & STEVENS, S. S. (1957). Critical bands in loudness summation. *Journal of the Acoustical Society of America*, **29**, 548-557.