TESTING A NEW THEORY OF PSYCHOPHYSICAL SCALING: TEMPORAL LOUDNESS INTEGRATION

Karin Zimmer, R. Duncan Luce and Wolfgang Ellermeier
Institut für Kognitionsforschung der Universität Oldenburg, Germany
Institute for Mathematical Behavioral Sciences, University of California, Irvine, USA
Institut für Psychologie der Universität Regensburg, Germany
e-mail: karin.zimmer@uni-oldenburg.de

Abstract

A theory recently proposed by Luce (2001) details qualitative assumptions underlying methods of sensory integration, magnitude scaling, and cross-modality matching. Apart from judgments of ordering and proportion, the theory invokes a psychological concatenation operation, something that does not often appear in psychological applications. One such operation is the focus of the present investigation in which concatenation is realized as the sequential presentation of two noise samples which the auditory system is thought to integrate temporally with respect to loudness. To that end, two 50 ms-bursts of white noise in immediate succession but differing in intensity were presented diotically to each participant. The task was to adjust the loudness of a third 50 ms-noise burst so that its loudness matched that of the composite noise. The theory leads to two distinct pairs of properties for the operation: either commutative and associative or non-commutative and bisymmetric. Commutativity means that the temporal order of a stimulus pair does not alter the match; associativity means that the two groupings of three stimuli in fixed order have the same match; and bisymmetry means that in a grouping of four ordered stimuli into two pairs, an interchange of the two middle stimuli leads to the same match.

Results of Experiment I showed that, except for one participant, loudness adjustments tended to be influenced more by the intensity of the first component than by that of the second. Using a 7-dB range of noise levels and 6 participants, commutativity was not directly rejected for anyone; however its theoretical consequence, associativity, was rejected for 5 of 6 participants (Mann-Whitney U-tests, $\alpha=0.10$). The magnitude of the effect depends upon the range of stimulus intensities used. With 3, 7, 15, and 20 dB-ranges in Experiment II, clearer violations of both commutativity and assocativity were found as the range of levels increased. Experiment III served to establish that this finding does not change when the matching stimulus is increased from 50 ms to 100 ms, the overall duration of the composite noise to be matched. Preliminary results of a fourth experiment currently in progress indicate that the bisymmetric property better describes participants' loudness adjustments than does commutativity and associativity.

Scaling constitutes the assignment of numbers to magnitudes of sensation that are evoked by physical stimuli. Following an approach advocated by Stevens (1956, 1975), subjects either produce their numerical estimates of sensation magnitudes evoked by stimuli (magnitude estimation) or they are presented with numerals and are asked to adjust stimulus intensity so that its sensation magnitude matches the given numeral (magnitude production). In Stevens' approach, these numerals are taken at face value in the sense of assuming the numerals to exhibit all the mathematical properties of numbers, and are found to form approximately power functions of signal intensity. Different representational systems have been proposed that specify fundamental conditions that have to be met in order to justify this procedure of treating numerals as numbers (Krantz, 1972, Luce, 1959, 1990, Shepard, 1981, Narens, 1996). All of these structures have been restricted in applicability to one of several scaling methods frequently used.

Recently, however, Luce (2001) has proposed a comprehensive theoretical framework of psychophysical scaling that treats, and seeks to unify, its most important methods, namely sensory integration of stimuli, magnitude (or proportion) scaling, and cross-modality matching. Moreover, this theory not only makes it possible to test whether these methods can be validly employed at all, but provides for the construction of a scale. This is done by identifying conditions that allow for a specification of the parameters both of the psychophysical function relating stimulus intensity to sensation strength, and of the weighting function relating overt numerical judgments to the underlying (mathematical) numbers.

The present investigation focuses on the psychological concatenation operation employed in the theory because, in general, such an operation is not often studied in psychological applications. Here, concatenation is realized as the sequential presentation of two noise samples to the auditory system, which is thought to perform some sort of temporal integration with respect to loudness. It is an operation in the sense that if $a \oplus b \sim u$, $(a \oplus b) \oplus c \sim u \oplus c \sim v$ is an intensity, where ~ denotes a subjective match.

Luce' s sensory-integration structure involves three fundamental properties:

(1) Commutativity: the temporal order of a stimulus pair does not matter, e.g., presenting a 70dB stimulus, *a*, followed by a 60 dB stimulus, *b*, should lead to the same loudness estimate as presenting the 60 dB stimulus, *b*, first, and the 70 dB stimulus, *a*, second, i.e., $a \oplus b \sim b \oplus a$.

(2) Associativity: two groupings of three stimuli *a*, *b*, *c*, in fixed order - either concatenating *a* and *b* first, then concatenating the resultant with *c*, or concatenating *b* and *c* first, and then concatenating the resultant with *a* - are matched by the same stimulus, i.e., $(a \oplus b) \oplus c \sim a \oplus (b \oplus c)$.

(3) Bisymmetry: in grouping four ordered stimuli *a*, *b*, *c*, *d*, into two pairs, interchanging the two middle stimuli leads to the same match, i.e., $(a \oplus b) \oplus (c \oplus d) \sim (a \oplus c) \oplus (b \oplus d)$.

The theory arrives at two types of sensory-integration structures: either commutative and associative or non-commutative and bisymmetric. In the presence of the other assumptions of the theory, commutativity implies associativity, which establish order-invariance of stimuli, whereas the assumption of non-commutativity implies bisymmetry, which allows for a differential weighting of stimuli with respect to the order in which they are presented.

In four experiments, the empirical validity of these three properties was investigated, using as stimulus material 50-ms noise samples differing in sound pressure level. Experiment I served to test commutativity and associativity in 7 and 15 dB ranges of stimulus levels. In Experiments II and III, the influence of stimulus range, and of the duration of the noise samples on axiom validity was investigated, using 3, 7, 15, and 20 dB intensity ranges. In Experiment IV, in addition to commutativity and associativity, the bisymmetry condition was tested in 15 and 20 dB-ranges of stimulus intensities. Data collection for this experiment has not been completed yet, therefore the present report focuses on the first three experiments only.

Method

Subjects

A total of 12 subjects between 21 and 43 years of age participated in the experiments. Six participated in Experiment I, and six participated in both Experiment II and III.
Using Békésy tracking, all participants were within 25 dB of the hearing norm in a range of 500 to 6000 Hz. Except for the first and third authors, who took part in the first experiment, all subjects were naive with respect to the goals of the investigation.

Stimuli and Apparatus

Stimuli consisted of 50 ms-bursts of white noise (including 2.5 ms rise and decay ramps) which were generated digitally by a TDT AP2 signal processor, and converted with a 50 Hz-sampling-rate to analogue signals by a 16-bit TDT DD1 converter. Signals were then passed through a low-pass filter set at 20 kHz (TDT FT6), and were set to appropriate sound pressure levels by a programmable attenuator (TDT PA4), before being delivered diotically via headphone amplifier (TDT HB6) to AKG-K501 headphones. All levels, as specified in table I, for example, are overall sound pressure levels of the noise, as measured at the earphones.

Procedure

Subjects, who were seated in a sound-proof chamber, were presented diotically with two 50 ms-bursts of white noise that were different in level and succeeded each other without temporal separation. The participants' task was to adjust the loudness of a third 50 ms-noise burst, the comparison sound, presented 400 ms later, so that its loudness matched that of the composite noise.
Participants indicated via a button press whether the comparison should be made louder or softer. A modified method of adjustments, in which step-sizes successively decreased from 4 dB to 0.5 dB, was used to obtain loudness matches.
Experiments I and II tested both commutativity and associativity with two and four stimulus sets, respectively. Subjects had to produce all adjustments in every block of trials. The trial order was random. After one practice block, participants completed 15 blocks per experimental condition. The stimulus levels employed in the experiments are given in table I.

Table I. Overall sound pressure levels [dB SPL] used in Experiments I and II.

Experiment	stimulus a	stimulus b	stimulus c	stimulus range
I	67	71	74	7
	62	69	77	15
II	64	66	67	3
	67	71	74	7
	62	69	77	15
	60	70	80	20

Note. The (a,b) stimulus pairs were used in testing commutativity.

Experiment III evaluated the commutative property again using the same four stimulus pairs as in Experiment II (s. Table I), but with the comparison sound lengthened to 100 ms, the duration of the composite noise. The order in which sessions for Experiment III, and the 7 dB/15 dB-, and 3 dB/20 dB-conditions of Experiment II were run followed a latin-square design across participants.

Results

Data were evaluated on an individual level. For all subjects and all SPL combinations, the comparison sound level was higher than that of the lower level in the composite.

Experiment I

In testing the commutative property, median differences between loudness matches for $a \oplus b$ and $b \oplus a$ ranged from –1 to +1 dB and from –1 to +2 dB, when noise bursts were 2 dB and 7 dB apart, respectively. With the exception of one participant, median differences were rather small, and, if present at all, they were negative for both ranges. This means that, presenting the less intense composite of a stimulus pair first resulted in a lower-level match, than when the more intense signal was first.
Likewise, in testing associativity median differences between matches for $(a \oplus b) \oplus c$ vs. matches for $a \oplus (b \oplus c)$ ranged between 0.5 and 1.5 dB, when stimulus intensities spanned a 7 dB-range, and between 1.5 and 3 dB when stimuli were in the 15 dB-range of stimulus intensities. For all subjects and for both ranges, median differences were positive, again showing a trend for lower loudness matches when the first composite of the stimulus was lower in level than the second.
In using a 7-dB overall range of noise levels, commutativity was not violated by anyone and associativity was violated by 5 of the 6 participants in a statistically significant way (Mann-Whitney U-tests; overall α=0.10, Bonferroni-corrected for multiple testing of participants to test-wise α=0.026). In the broader, 15 dB-range of stimulus intensities, commutativity and associativity failed in 1 of 6 and in all of 6 participants, respectively.
Thus, with the exception of one participant, loudness adjustments tended to be influenced more by the intensity of the first than by the intensity of the second component. This trend shows in testing both properties, and is rather clear-cut for the associative axiom, whereas, for commutativity, the effect is not sufficiently strong to produce statistically significant results. Note, however, that in the theory rejecting associativity indirectly rejects commutativity.

Experiment II

Experiment II served to investigate whether violations of the axioms depended on the range of stimulus intensities. Accordingly, 3, 7, 15, and 20 dB-ranges of stimulus intensities were employed to evaluate commutativity and associativity of loudness matches.
Testing commutativity, three of six participants for all stimulus ranges exhibited negative median differences reaching up to –1.5 dB. For two other participants, both positive and negative median differences (between –1 and +1 dB) were found. Only one participant showed positive differences, reaching a maximum of +1 dB. Median differences reached a statistically significant level for three of (4 ranges x 6 participants =) 24 tests (Mann-Whitney U-tests, α=0.10, Bonferroni-correction for multiple testing of participants: test-wise α=0.013). In all three cases, violations occurred in the two largest stimulus ranges, and median differences were negative.
As was already observed in Experiment I, results on the associative property were more clear-cut (s. Table II) in that, with one exception out of 24 cases, all median differences were

positive. Moreover, they tended for all subjects to be larger for larger ranges of stimulus intensities. Except for one subject, differences reached statistical significance for the two largest stimulus ranges, 15 and 20 dB for all participants. In addition to that, three subjects already showed statistically significant violations of associativity in the 7 dB-range of stimulus intensities.

Thus, the larger the range of stimulus levels used, the clearer the violations of associativity. The results for commutativity are more equivocal than that. Since median differences are rather small, any possible effect of stimulus range on axiom validity may have been blurred altogether.

Table II. Experiment II: Results of testing the associative property in four stimulus ranges.

	3 dB range		**7 dB range**		**15 dB range**		**20 dB range**	
Subject	Med diff.	z-score	Med diff.	z-score	Med diff.	z-score	Med diff.	z-score
BG	0.5	0.04	-1.0	0.10	2.0	1.93	1.5	1.77
CB	0.5	1.76	1.5	**3.48**	3.0	**4.68**	5.5	**4.68**
DO	0	0.17	1.0	**2.29**	3.0	**4.24**	3.5	**2.95**
HE	0	0.04	1.0	1.10	2.5	**3.90**	2.5	**3.73**
PI	1.0	0.62	1.5	**4.46**	3.5	**4.10**	2.0	**3.52**
VE	1.0	1.84	1.0	2.31	3.0	**4.13**	3.0	**4.67**

Note. For every subject, median differences between $(a \oplus b) \oplus c$ and $a \oplus (b \oplus c)$, and corresponding z-scores, are given for 3, 7, 15, and 20-dB stimulus ranges. Median differences are based on 15 adjustments each. Statistically significant z-scores are high-lighted (Mann-Whitney U-tests, α=0.10; Bonferroni-correction for multiple testing of participants: test-wise α=0.013).

Experiment III

Using the same stimuli as in Experiment II, commutativity was again evaluated, but with the comparison sound lengthened from 50 ms to 100 ms, the overall duration of the composite noise. Median differences ranged from –1.5 dB to +1 dB, that is, they were in the same general range as in Experiment II. With one exception in the smallest stimulus range, the results of those three participants who had shown negative differences in all ranges of stimulus intensities in Experiment II were replicated. For the remaining three participants, either no, or exclusively positive differences were found. Commutativity was violated in a statistically significant way for three of 24 cases (Mann-Whitney U-tests, α=0.10, Bonferroni-correction for multiple testing of participants: test-wise α=0.026). Again, in all statistically significant cases, median differences were negative, and violations occurred only in the two largest stimulus ranges.

Because the results for commutativity of Experiments II and III are similar, we conclude that the duration of the comparison sound does not have a crucial influence on estimating the validity of the commutative property.

Discussion

The present investigation shows that the associative property, and by implication also commutativity, is not valid in describing the concatenation of noise-bursts presented immediately after each other with respect to an overall loudness impression. Furthermore, axiom violations tend to increase with increasing sound-level differences between the stimulus components.

Apparently, the sound level of the first noise-component is more influential in determining the loudness judgment. Thus, it may well be that a structure which is able to account for weighting the presentation-order of stimuli will form a better representation. The theoretical alternative, under some smoothness conditions described in Luce (2001), is the bisymmetric property. If it is empirically correct, an invariance argument implies that the psychophysical function should be a power function. Indeed, results for the first three out of six subjects who have been run in an experiment evaluating the bisymmetric property point in this direction. If this finding holds for the other participants as well, one of the fundamental structures of Luce’s theory, namely the sensory-ingegration representation, can be thought to hold in this domain.

The next step in evaluating other aspects of this psychophysical theory is to test the properties of numerical judgments (or adjustments), which have not been at issue yet in the present sensory-integration experiments. To that end, experiments should be performed, which investigate the internal consistency of these judgments based on Luce’s generalization of Narens’ (1996) axiomatization. Furthermore, conditions which involve both sensory integration and numerical judgments („binary segregation“ and „simple joint-presentation decomposability“) should also be investigated.

If this programme can be successfully completed, a global theory may be established, which specifies the (perceived) loudness of noise-bursts, no matter which of several psychophysical scaling methods is applied. If any of the fundamental conditions assumed in the theory fail, then an appropriate modification of the theory will be called for as occurred, for example, when the original commutative theory of Luce (2001) was rejected both for a temporal-integration (present experiments) and for a binaural-summation interpretation (Steingrimsson, personal communication) of concatenation.

References

Ellermeier, W. & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens’s ratio-scaling approach: I. Loudness production. *Perception & Psychophysics*, **62(8)**, 1505-1511.

Krantz, D. H. (1972). A theory of magnitude estimation and cross-modality matching. *Journal of Mathematical Psychology*, **9**, 168-199.

Luce, R. D. (1959). On the possible psychophysical laws. *Psychological Review*, **66**, 81-95.

Luce, R. D. (1990). “ On the possible psychophysical laws“ revisited: Remarks on cross-modal matching. *Psychological Review*, **97**, 66-77.

Luce, R.D. (2001). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*. In press.

Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, **40**, 109-129.

Shepard, R. N. (1981). Psychological relations and psychophysical scales: On the status of “direct“ psychophysical measurement. *Journal of Mathematical Psychology*, **24**, 21-57.

Stevens, S.S. (1956). The direct estimation of sensory magnitude – loudness. *American Journal of Psychology*, **69**, 1-15.

Stevens, S.S. (1975). *Psychophysics: Introduction to ist perceptual, neural, and social prospects*. New York: Wiley.