

# A behavioristic analysis of Stevens' magnitude estimation method

TADASU OYAMA

CHIBA UNIVERSITY, JAPAN

A procedure independently proposed by Oyama and Lewis for treating the data of magnitude estimation is discussed in relation to the Graham-Stevens controversy on the psychophysical methods. In this procedure, the critical value of the stimulus needed to make the subject report numbers equal to or greater than each integral number is sought, instead of the median or geometrical mean of numbers reported for each level of stimulus intensity. Stevens' magnitude estimation method has no special advantage over the other cross-modality matching methods, and the subjective magnitude of number should be estimated if one wants to know real relations between stimulus intensities and sensory magnitudes in various modalities.

The studies carried out by Stevens and his associates in recent years have helped to accumulate scientific data concerning subjective judgments of sensory magnitudes, and to develop hypotheses on the relation of such data to the objective intensity of stimuli. Stevens' contributions should be acknowledged also for their influence in directing the attention of experimental psychologists to problems which had been sidetracked since Fechner's time.

However, the methods and procedures used in Stevens' laboratory seem incompatible with Graham's (1950, 1965) behavioristic point of view on the psychophysical methods. According to the latter view, psychophysical methods are techniques to seek stimulus values or stimulus-stimulus relations which produce responses having constant critical characteristics or with predetermined probabilities of occurrence. In Stevens' magnitude estimation method, contrary to Graham's view, medians or geometrical means of numbers reported by Ss are directly related to stimulus intensities.

Oyama and Takekawa (1962) and Oyama (1962) presented a new procedure to analyze their data on loudness estimation. They gathered data under a Stevens-type instruction and treated them through a Graham-type analysis to find critical values of stimuli for various number-responses. More recently, Lewis (1965) independently developed a very similar procedure in order to treat his data on brightness estimation of flashes in a study on intensity x time relations.<sup>1</sup> Stevens (1966b) criticized Lewis' procedure and ascribed his failure to find the Broca-Sulzer effect to the inappropriateness of his procedure in data treatment. However, some parts of Stevens' criticism seem to be based on a misunderstanding of Lewis' procedure and other parts cannot be applied to the study of Oyama and Takekawa.

An attempt will be made here to clarify the basic character of the Oyama-Lewis procedure and its implications for the Graham-Stevens controversy on psychophysical methods.

## The Oyama-Takekawa Study

It will be worthwhile to summarize briefly the study of Oyama and Takekawa (1962) which might be unfamiliar to American psychologists, and perhaps more suitable than Lewis' study for illustrating the Oyama-Lewis treatment of estimation data.

In the study of Oyama and Takekawa, a standard stimulus was employed. It was a 1000 cycle tone whose intensity was set at 74 dB sound pressure level. The comparison stimulus was also a 1000 cycle tone, but its intensity varied from 53 to 92 dB in 3 dB steps. Tone pairs were presented 50 times

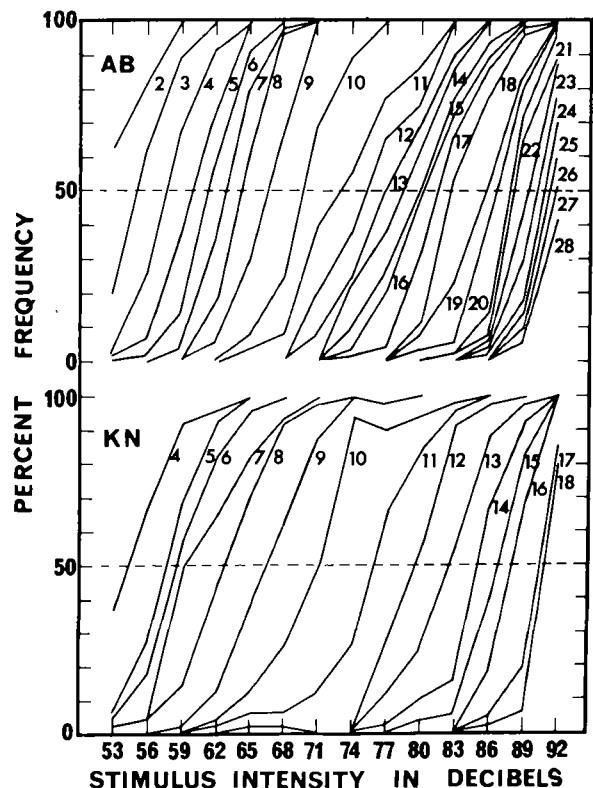


Fig. 1. Distributions of number-responses in two Ss. Each curve shows the percent frequency of reporting numbers equal to or greater than the number attached to the curve.

in random sequences covering each intensity level of the comparison stimulus. Five Ss individually served in six sessions and were instructed to report a number which appeared appropriate to represent the subjective magnitude of the second (comparison) sound of each pair when that of the first (standard) sound was regarded as "10." The Ss were free to use any number, whole numbers, fractions, or decimals. The procedure of Oyama and Takekawa does not differ in its essentials from Stevens' methods in the parts described so far, but the treatment of the data is different.

Figure 1 shows two examples of the results obtained by Oyama and Takekawa. Each curve indicates the percentage of cases in which the number reported by the S was equal to or greater than an integral number (criterion number) as a function of the stimulus intensity (in decibels) of the comparison sound. Twenty-seven curves are presented for Subject AB as the criterion number varied from 2 to 28, while 15 curves for Subject KN correspond to criterion numbers ranging from 4 to 18. Nearly all curves show the familiar S-shape seen in psychometric functions obtained via traditional constant methods. The latter were classified into the category of stimulus-response functions by Graham (1965).

Suppose that the instructions of this experiment were changes so that the S responded "yes" if the perceived ratio of the two sounds was greater than 10:6 and said "no" otherwise. Then, the method would result in curves similar to those labeled "6" in Fig. 1. Other curves in Fig. 1 would then be obtained if the ratio in such instructions were varied in turn from 10:2 to 10:28. In this sense, the curves in Fig. 1 may be regarded as psychometric functions, and the ordinary procedures for estimating thresholds can be applied to them. The intersections between the psychometric curves and the horizontal lines indicating the 50% level of response frequency will represent such thresholds.

Figure 2 shows threshold values in this sense for five Ss along a decibel scale of stimulus intensity. These values were estimated by linear interpolation. The thresholds indicate the stimulus values at which the probabilities of occurrence of number-responses equal to or greater than various criterion numbers would be 0.5. Each space between two adjacent threshold-lines represents the stimulus range in which a certain number-response was predominant. Figures 1 and 2 both indicate heterogeneity in frequency of occurrence and in width of stimulus range among various number-responses. In most of the Ss, numbers 9, 10, 11, etc. occur frequently over wide ranges of stimulus intensity, but 5, 15, 17, etc. occur much less frequently and in narrower ranges. This may partly reflect individual preferences for numbers or perhaps other biases in their subjective number systems.

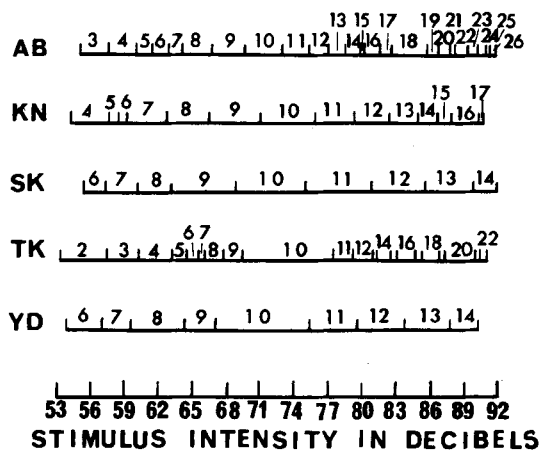


Fig. 2. Number scales for five Ss and the decibel scale of stimulus sound.

### Stevens' Criticism of the Lewis Study

Stevens' (1966b) criticism of Lewis' (1965) study consisted of two main parts. First, Stevens pointed out that the responses of Lewis' Ss were restricted to a limited number of categories and that such restrictions had been demonstrated by many experiments to result in scales nonlinearly related to the ratio scale obtained under free use of the number continuum. This part of Stevens' criticism may be effective as far as Lewis' study is concerned. In fact, Lewis restricted Ss' responses to five or 11 integral numbers. But Oyama and Takekawa (1962) did not impose such restrictions. Their Ss were as free to use numbers as Stevens would wish. Lewis' treatment of estimation data can be used without any restriction of this kind.

Lewis used the term "category judgments" for his study and suggested the similarity between his procedure and the method of successive categories. These facts, however, have nothing to do with the basic character of the Oyama-Lewis procedure we are discussing now. The procedure can be used in any study in which the stimulus varies along some quantitative scale and the responses emitted by Ss can be classified by the investigator into several categories arranged in order. In the study of Oyama and Takekawa, the Ss did not make any category judgments, but their responses were classified and it was assumed that these response classes were ordered.<sup>2</sup>

The second part of Stevens' criticism was concerned with the regression method used by Lewis. Stevens maintained that Lewis averaged the stimulus values for each number-response and tried to find the regression of the stimulus on the response, although the appropriate regression was that of the response on the stimulus. Actually, Lewis *did not* average the stimulus values. Instead, he found criti-

cal values of stimuli at each of which the dominance of a response class was overtaken by the dominance of another. In other words, he obtained stimulus boundaries of various response classes.

This type of regression of stimulus upon response was also used by Oyama and Takekawa. Which kind of regression is better, response upon stimulus or stimulus upon response, is a difficult problem for which we cannot offer a general conclusion. It might be said that the regression of response upon stimulus is better, as in Stevens' conclusion, only under such limited conditions that the measure of response constituted a quantified scale on a level approximating that of the measure of the stimulus. Our main concern in such circumstances is the prediction of response values from stimulus values. However, in many cases, including the case in question, the measure of the response is poorer in quantification than that of the stimulus, and the prediction of responses from a given stimulus condition is not our sole concern. We equally need to know the most probable value of the stimulus that will produce a critical point in the distribution of responses. Traditional psychophysical methods generally used to find thresholds, differential limens, and points of subjective equality, as well as Stevens' (1957) magnitude or ratio production method, essentially originated from the need of psychologists to know critical values of the stimulus for some constant response.

Figure 3 shows two kinds of regression applied to the data of the five Ss of Oyama and Takekawa. The filled squares represent the individual medians of reported numbers for each level of stimulus intensity, and the open circles indicate the threshold values of stimulus intensity obtained with the procedure discussed above for each integral number. The positions of the circles plotted in this figure are lowered by 0.5 as compared with those of the squares by reason of the fact that these thresholds represent the stimulus boundaries of the successive response classes rather than the central tendencies of individual response classes, which correspond to integral numbers. The squares and circles show two kinds of regression, response upon stimulus and stimulus upon response, respectively, but both constitute practically the same curve for each S. This fact means that the claimed difference between the two methods of regression has no practical meaning. Theoretically, the regression of stimulus upon the response (the circles in Fig. 3) would be better in this case because it is more natural to assume that stimulus intensity constitutes an interval scale and the method of interpolation is applicable, than it is to assume that the numbers reported by Ss constitute an interval or ratio scale, and that the medians or geometric means represent unbiased central tendencies.

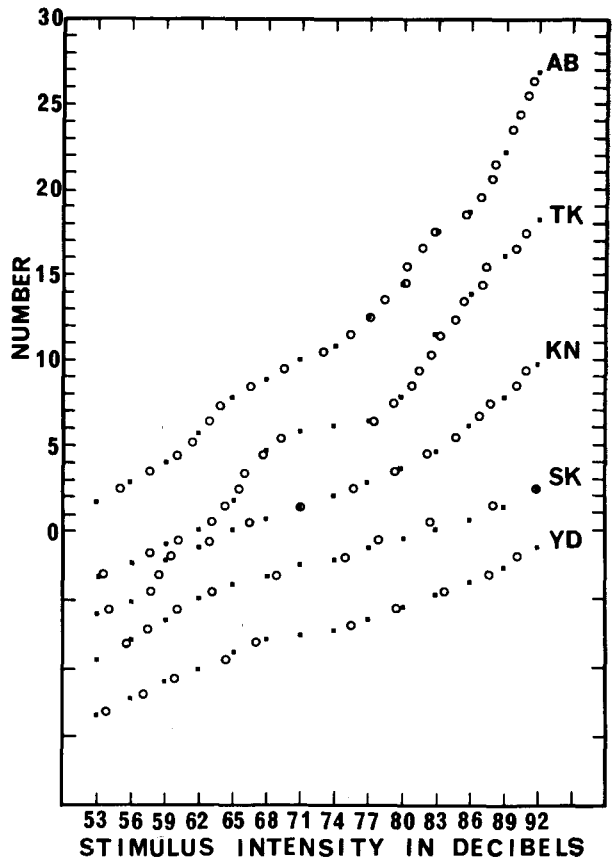


Fig. 3. Magnitude estimations as functions of stimulus intensity in decibel for five Ss. The dots show the results of Stevens type treatment, and the circles, those of Oyama-Lewis type treatment applied on the same data. Each successive curve is lowered through 4 in the scale of ordinate.

#### Subjective Magnitude of Number

Stevens (1966) presented a view that the task of Ss in magnitude estimation experiments is a kind of cross-modality matching in which numbers are assigned to stimuli in such a way that the numbers are proportional to the apparent intensity of the stimuli. This view suggests a possibility for connecting Stevens' method to Graham's analyses of psychophysical methods. The term "cross-modality matching" is ordinarily used to indicate such tasks as controlling the stimulus intensity of a sound to match its loudness with the brightness of a light (Stevens, 1966a), or squeezing a dynamometer with force whose apparent strength is matched with the apparent strength of an electric shock, sound, or light (Stevens, Mack, & Stevens, 1960). The results of such cross-modality matching are usually expressed by the stimulus intensity of matched sound, the physical strength of handgrip, or some other objective measure of matched stimuli.

If we regard the method of magnitude estimation

as an example of cross-modality matching, we have to clarify the distinction between the number as stimulus and the subjective magnitude of number. The former may be given as a form of verbal stimulus or its memory trace, to which the S may have learned many responses including arithmetical calculation. In contrast with it, the latter is a theoretical construct concerning the S's private event. It is very doubtful that the subjective magnitude of 100 is really ten times as large as that of 10. Consequently, a magnitude of sensation to which the number 100 is assigned is not necessarily ten times as large as another magnitude of sensation to which the number 10 is assigned. The numbers reported by the Ss in Stevens' experiments as well as those of Oyama and Takekawa would belong to numbers as stimuli and would not necessarily represent sensory magnitudes directly.<sup>3</sup>

Even if we observe a power function between the stimulus intensity and the number assigned to it, we cannot instantaneously assume a power function relation between the stimulus and sensation. There are many possible interpretations of this observation. For instance, we can follow Fechner to assume logarithmic functions between the stimulus intensity (I) and the sensory magnitude corresponding to it ( $\Psi_i$ ) and between the number (N) and its subjective magnitude ( $\Psi_n$ ):

$$\Psi_i = a \log I \quad (1)$$

$$\Psi_n = b \log N \quad (2)$$

When  $\Psi_n$  is equated to  $\Psi_i$  in magnitude estimation,

$$\Psi_n = \Psi_i \quad (3)$$

and then  $b \log N = a \log I$  (4)

Thus we obtain a power function between N and I:

$$N = I^k \quad (5)$$

where  $k = a/b$ .

Of course we may assume power functions between the stimuli and sensations:

$$\Psi_i = I^c \quad (6)$$

and  $\Psi_n = N^d$  (7)

When  $\Psi_n$  is matched with  $\Psi_i$ , we will get a power function again:

$$N = I^l \quad (8)$$

where  $l = c/d$ .

These considerations reveal that power functions

obtained with magnitude estimation do not directly show the relations between stimuli and sensations. We may assume functions of other types for these relations. Even when power functions are hypothesized, the obtained exponent will not represent c in Equation (6), but it will correspond to l in Equation (8). We have to know the value of d to estimate the value of c. The exponents obtained from the experimental results of magnitude estimation simply indicate relative efficiencies of stimuli in strengthening sensations of various modalities.

The method of magnitude estimation has no special advantage over other cross-matching methods. Cross-modality matchings between one standard modality (e.g., loudness, force of handgrip, or length of line) and many other modalities will be effective in the same extent as numerical estimation is. The complexity of the number scales shown in the study of Oyama and Takekawa suggests even a disadvantage in numerical estimation.

#### References

- Graham, C. H. Behavior, perception and the psychophysical methods. *Psychol. Rev.*, 1950, 57, 108-120.
- Graham, C. H. (Ed.) *Vision and visual perception*. New York: Wiley, 1965.
- Lewis, M. F. Category judgments as functions of flash luminance and duration. *J. Opt. Soc. Amer.*, 1965, 55, 1655-1660.
- Oyama, T. "Psychophysical" and "phenomenological" parameters in the construction of perception model. *Hokkaido Report of Psychology*, 1962, No. 6 (mimeographed technical report).
- Oyama, T., & Takekawa, T. Numeral-responses concerning loudness. *Proc. 26th Ann. Conv. Jap. Psychol. Assoc.*, 1962, 96, (abstract in Japanese).
- Stevens, J. C., Mack, J. D., & Stevens, S. S. Growth of sensation on seven continua as measured by force of handgrip. *J. exp. Psychol.*, 1960, 59, 60-67.
- Stevens, S. S. On the psychophysical law. *Psychol. Rev.*, 1957, 64, 153-181.
- Stevens, S. S. Psychophysics of sensory function. In W. A. Rosenblith (Ed.), *Sensory communication*. Cambridge, Mass.: The M. I. T. Press, 1961. Pp. 1-33.
- Stevens, S. S. Matching functions between loudness and ten other continua. *Percept. & Psychophys.*, 1966, 1, 5-6.
- Stevens, S. S. Duration, luminance, and the brightness exponent. *Percept. & Psychophys.*, 1966, 1, 96-100.

#### Notes

- Lewis cited Oyama and Takekawa's study in his paper, but they had an opportunity for communication after both of them had finished their studies independently.
- Lewis did not use this assumption of order, but he empirically assigned ranks to the numbers (response classes) in terms of mean stimulus intensity associated with each response distribution. It should be noticed that his operation was based on another assumption that empirically obtained means represent the central values of response populations. This assumption is not necessarily true, especially for the response classes used in both ends of the stimulus series, where only a biased sample of each response class can be observed. An extension of stimulus series will radically change the mean values of stimulus intensity for these response classes. Changes in rank order will possibly occur.
- The regression issue discussed above will disappear if we regard the magnitude estimation method as a kind of cross-modality matching because we will only treat regression of a stimulus continuum on another.

(Accepted for publication December 14, 1967.)