A harmonious note on pitch: Scales of pitch derived from subtractive model of comparison agree with the musical scale

ROBERT ELMASIAN
Children's Hospital and University of California, San Diego, La Jolla, California

and

MICHAEL H. BIRNBAUM
University of Illinois, Urbana-Champaign, Illinois

Subjects judged "ratios" and "differences" in pitch between pairs of tones that varied in frequency. Half of the subjects listened to stimuli that ranged from 191 to 844 Hz; the other half received a range that varied from 191 to 3740 Hz. For both ranges, the rank orders of judgments in the two tasks were essentially the same. If subjects actually use ratio and difference operations on the same scale values, the factorial design should yield different rank orders for each task. Instead, it was concluded that subjects use the same comparison operation for both tasks but map the subjective comparisons to overt responses differently for each dependent variable. If the comparison operation is subtraction, the data are consistent with the hypothesis that subjective pitch is a log function of frequency; that is, differences on the musical scale predict both the "ratio" and "difference" judgments. This result may resolve the previous discord between musicians and psychophysical psychologists on the subjective scaling of pitch.

It is often stated that the psychophysical scale of pitch and the Western musical scale are not linearly related. Stevens (1975, p. 168) remarked, "Musicians are often surprised to be told that the apparent size of a fifth expands to about seven times larger when it is moved from a low-frequency interval, 64 to 96 hertz, to a higher-frequency interval, 2048 to 3072 hertz . . . . This is also true with octaves." Stevens based the "apparent size" of the musical intervals on the mel scale, a psychophysical scale for pitch developed by Stevens and his associates (Stevens & Volkmann, 1940; Stevens, Volkmann, & Newman, 1937). Stevens's mel scale is nonlinearly related to the musical scale.[1] The present paper, however, derives a psychophysical scale of pitch that closely approximates the musical scale and is in harmony with previous research on stimulus comparison.

In the present study, pitch scaling is a byproduct of an investigation of stimulus comparison. Previous research has shown that judgments of psychological "ratios" and "differences" are monotonically related for continua such as heaviness of lifted weights (Birnbaum & Veit, 1974; Mellers, Davis, & Birnbaum, 1984), loudness of tonebursts (Birnbaum & Elmasian, 1977), and darkness

of gray papers (Veit, 1978). (Quotation marks are used around "ratios" and "differences" to distinguish task instructions or judgments from actual ratios and differences or models that describe the judgments.) Because actual ratio and difference operations are expected to produce different rank orders in the factorial designs employed in the cited studies, these results strongly suggest that subjects used the same comparison operation under both "ratio" and "difference" instructions (Birnbaum, 1978, 1979, 1980, 1982; Veit, 1978). For these continua, evidence agrees with Torgerson's (1961) hypothesis that the "subject perceives or appreciates but a single quantitative relation between a pair of stimuli."

Subjective continua such as heaviness, loudness, and darkness are usually classified as "prothetic" (Stevens & Galanter, 1957), because magnitude estimations ("ratios") and category ratings ("differences") are nonlinearly related on these continua. Pitch, however, is sometimes considered the foremost example of a second class of continua called "metathetic" (Stevens & Galanter, 1957), for which "ratio" and "difference" scaling are supposed to agree. Mels, for example, were said to be linearly related to category ratings and magnitude estimations of pitch as well as the summation of just noticeable differences (e.g., Stevens & Galanter, 1957; Zwislocki, 1965).

It is possible that the reported agreement between the mel scale, category ratings, and magnitude estimations occurs because subjects actually do use two operations—ratios and differences—with pitch. The present study was

designed to investigate whether the rank orders of pitch "ratios" and "differences" would be distinct (consistent with the two instructed operations) or whether the rank orders would be essentially the same for both tasks (consistent with Torgerson's hypothesis that subjects use only one comparison process).

METHOD

The method is similar to that of Birnbaum and Elmasian (1977). Subjects were presented with pairs of tonebursts. In one session, they estimated the "ratio" of the pitch of the second tone relative to the first in a pair; in another session, they judged the "difference" in pitch between the tones. Half of the subjects performed the "ratio" task first, and half did the "difference" task first. (No effect of task order was discerned.)

Task Instructions

The "ratio" task called for estimates of the "ratio of the pitch of the second tone to the pitch of the first tone." The modulus (the response for a "ratio" of unity) was designated "100," and printed examples were provided specifying that if the second tone seemed "one-fourth" as high as the first, the subject should respond "25"; if it seemed "half" as high, "50"; "twice" as high, "200"; and "four times" as high "400." Instructions encouraged the subjects to use whatever numerical values best represented the "psychological ratios."

The "difference" task instructions required ratings of the "difference in pitch between the second and the first tone." Pitch "differences" were rated on a 9-point scale with category labels varying from "1 = second tone is very, very much lower" to "9 = second tone is very, very much higher"; "5" was "equal."

Stimuli and Apparatus

Each signal was a 500-msec, 78-dB SPL tone burst generated by a Wavetek 155, gated on and off with 5-msec rise-fall times by a Grason-Stadler 1287B switch and presented monaurally over a TDH-39 earphone. The interstimulus interval was 1 sec.

Stimulus pairs were generated by a 5 × 9 factorial design (first tone × second tone frequency) in which the five levels of the first tone consisted of the first, third, fifth, seventh, and ninth frequency levels of the second tone. Four subjects received a "wide" range of stimuli for which the nine levels of the second tone were 191, 277, 402, 582, 844, 1200, 1770, 2570, and 3730 Hz. Another four subjects were presented with a "narrow" range in which the nine levels of the second tone were 191, 230, 277, 336, 402, 484, 582, 701, and 884 Hz. On a logarithmic scale of frequency, the stimuli are evenly spaced and the wide range is twice the narrow range.

Session Procedure

Written tests showed that each subject understood the intellectual distinction between ratios and differences and could calculate them numerically. Following task instructions and after a brief warm-up, each judge completed 12 repetitions of the 5 × 9 design in each session. The order of the 45 pairs within each repetition was randomized by the computer program that controlled the experiment.

Subjects

The subjects were eight paid members of the academic community of the University of California, San Diego. None of the subjects considered themselves to be musicians or thought they could be successful music majors, although most had had some informal experience attempting to play pianos, guitars, harmonicas, etc. One subject (H.C.) had participated in the Birnbaum and Elmasian (1977) study.

RESULTS

Figure 1 plots the subjects' mean "ratio" estimations on a logarithmically spaced ordinate against their mean "difference" judgments. The data for the narrow-frequency-range group are displaced three units to the left of the wide-range group. For both ranges, "ratio" judgments are nearly a monotonic function of "differences," consistent with the theory that subjects performed the same comparison operation in both tasks.

The theory of one operation can be written as follows:

$$R_{ij} = J_R(s_j - t_i) \qquad (1)$$

$$D_{ij} = J_D(s_j - t_i), \qquad (2)$$

where R_{ij} and D_{ij} are the "ratio" and "difference" judgments; s_j and t_i are the scale values of the second and first stimuli, respectively; and J_R and J_D are strictly monotonic judgment functions. It follows that $R_{ij} = J_R[J_D^{-1}(D_{ij})]$; hence, "ratios" are a monotonic function of "differences."
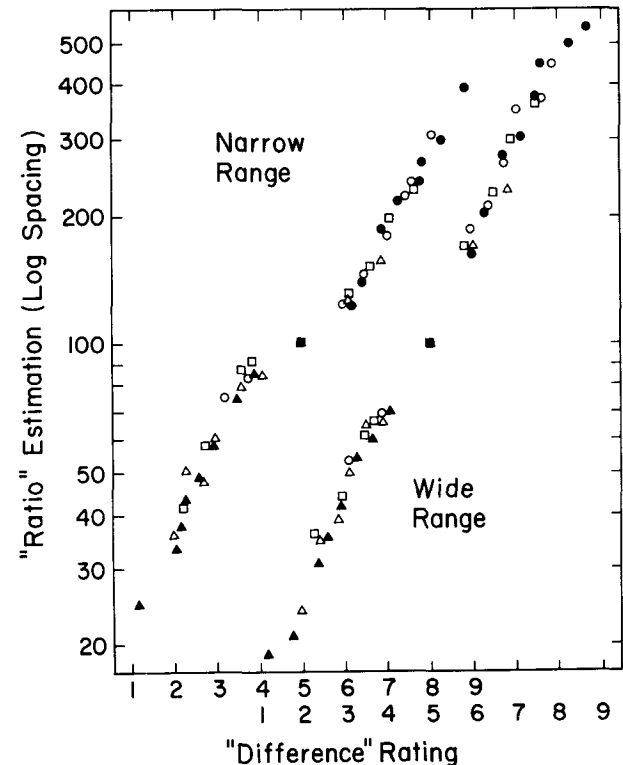

Figure 1. Mean "ratio" estimations for each stimulus pair as a function of mean "difference" judgments. Ordinate is spaced logarithmically. The results for the wide stimulus range have been displaced three units to the right of the narrow-range results. Filled circles, open circles, squares, open triangles, and filled triangles represent levels 1, 3, 5, 7, or 9 of the first stimulus, respectively. Compare data with Figure 2.

If the subjects had actually calculated both ratios and differences on the same pitch scale, the plot in Figure 1 would not be monotonic. To illustrate this prediction, numerical ratios (s_j/t_i) are plotted against numerical differences ($s_j - t_i$) in Figure 2. A 5 × 9 design used the numbers 1, 3, 5, 7, and 9 as the scale values for the first stimulus (t_i) and the integers 1 through 9 for the second stimulus (s_j). The predictions in Figure 2 are not monotonic, and no strictly monotonic (nondegenerate) transformation of either the abscissa and/or the ordinate will result in a monotone plot (Birnbaum, 1978, 1980, 1982). For example, 3/1 > 9/5 but 3 − 1 < 9 − 5. Thus, because the data in Figure 1 do not resemble the pattern of predictions in Figure 2, they provide no evidence that subjects performed both operations, ratios and differences, as instructed. Instead, the results are consistent with the theory that subjects used the same comparison operation in both tasks.

Metric Analyses

Figure 3 shows "difference" judgments with separate panels plotting the cell means for the narrow and wide ranges. The abscissa indicates the nine frequencies for the second tone, and separate curves are drawn for values of the frequency of the first tone. Assuming responses are linearly related to subjective differences (i.e., J_D in Equation 2 is linear), the curves should be parallel (Anderson,

Figure 3. "Difference" ratings as a function of the second stimulus (log spacing) with a separate curve for each level of the first stimulus. The left panel is for the narrow-range group and the right panel is for the wide-range group. The symbols indicate the different levels of the first stimulus, as in Figures 1 and 2.
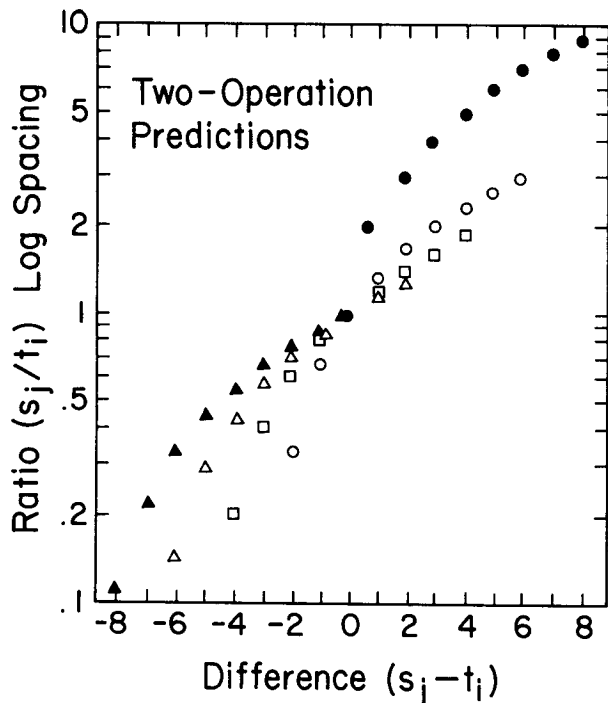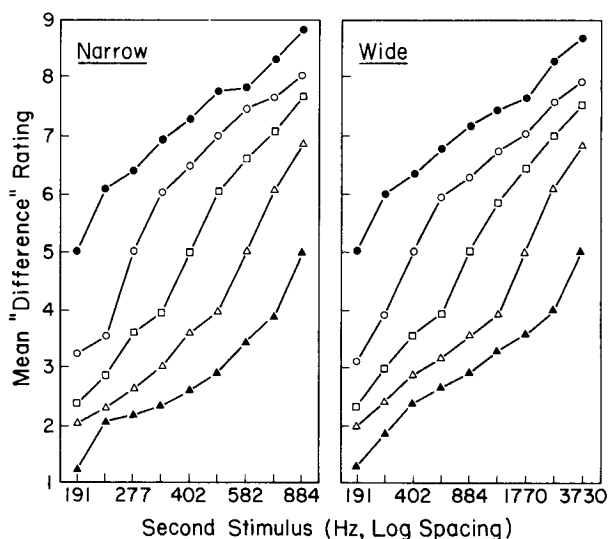

Figure 2. Actual ratio (A/B) plotted against actual difference (A−B). Filled circles, open circles, squares, open triangles, and filled triangles indicate values of 1, 3, 5, 7, or 9 for B, respectively. The values of the second stimulus (A) are 1, 2, 3, 4, 5, 6, 7, 8, or 9. Note that the plot is not monotone. Compare pattern with Figure 1.

1974; Birnbaum & Veit, 1974). The curves are roughly parallel, although there is a systematic deviation from parallelism. Subjects virtually always recognize the smallest frequency differences and rate them at least one category unit from equality, that is, "4" or "6," depending on direction. Consequently, the curves are steep near "5" on the ordinate ("equal in pitch"). An opposite deviation was observed by Birnbaum and Elmasian (1977), who found that subjects tended to say too often that a 6-dB difference in intensity was "equal" in loudness. This pattern suggests that the nonparallelism may be attributable to a slightly nonlinear judgment function, J_D, relating subjective differences to overt responses.

Figure 4 shows the "ratio" judgments as in Figure 3. The ordinate, however, is spaced according to the log of the "ratio" judgments. If the ratio model holds and the judgment function (J_R in Equation 1) is any power function, the data should plot as a set of parallel curves on the logarithmic ordinate. [The log of a ratio is a difference, $\log(s/t)^a = a\log s - a\log t$; therefore, plotting the log of ratios results in the parallelism characteristic of differences (Birnbaum & Veit, 1974).]

Given the logarithmic spacing of the ordinate, parallelism in Figure 4 would be consistent with the ratio model. The wide-range data, in particular, show a steepness near equality (100 on the ordinate) similar to the pattern for the "difference" judgments in Figure 3. These deviations from parallelism also appear attributable to the judgment function.

The range of numbers individual subjects used to report "ratios" varied widely. The largest mean "ratio" judgments, averaged over repetitions, were 8.62, 6.98, 3.17,
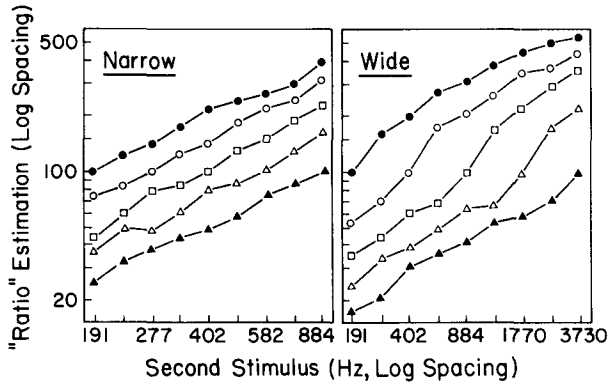
Figure 4. "Ratio" estimations plotted as in Figure 3, except that the ordinate is logarithmically spaced.

and 3.13 by each of the wide-range subjects and 5.80, 3.47, 3.41, and 3.41 by each of the narrow-range subejcts. In the same subject order, the smallest mean "ratios" reported were .17, .04, .25, .27 and .11, .29, .30, .30. Such intersubject differences in the range of "ratios" reported are common (Birnbaum, 1980; Painton, Culli-nan, & Mencke, 1977; Poulton, 1979). However, in-dividual plots drawn separately for each subject showed that individual data patterns closely resembled the pattern in Figure 4. (See Birnbaum & Elmasian, 1977, for analo-gous individual plots of loudness judgments.)

The one- versus two-operation theories were also com-pared by the numerical techniques described in Birnbaum (1980). The one-operation model of Equations 1 and 2 was further constrained to assume that J_D was linear and J_R exponential. The two-operation model replaces the difference operation in Equation 1 with a ratio operation, $R_{ij} = a(s_j/t_i)^b + c$, where J_R is a linear function of a power function. The one-operation model provided a better fit to the data for both ranges. When the "ratio" judgments are fit with a ratio model, the two-operation model im-plies that a physical ratio of 19.5 (3730/191) corresponds to a subjective ratio of only 1.8! Furthermore, to explain why this pair (3730/191) receives a mean judged "ratio" of 5.5, the two-operation model requires that judged "ra-tios" be subjective ratios raised to the 3.67 power! These values are inconsistent with previous theories. (See Birn-baum, 1980, for further discussion.)

Nonmetric Analyses

In order to remove nonlinearities in the judgment func-tions (sometimes called "response bias"), the data for both tasks (Figures 3 and 4) were transformed to parallelism via MONANOVA (Kruskal & Carmone, 1969), a computer program for nonmetric scaling. The 540 (5 × 9 × 12) judgments of each subject in each task were separately rescaled to the model:

$$T_X(X_{ijk}) = s_j - t_i + e_k, \qquad (3)$$

where T_X is the strictly monotonic inverse of the output function for task X; X_{ijk} is the response ("ratio" or

"difference") to the kth repetition of stimulus pair ij; t_i and s_j are the estimated scale values of the first and sec-ond stimuli, respectively; and e_k is an (error) effect of repetition block. This transformation simultaneously at-tempts to reduce first stimulus × second stimulus inter-actions and all interactions involving repetitions. Separate scale values are allowed for the first (t) and second (s) stimuli to allow for order effects (Birnbaum, 1980, 1981).

Equation 3 is a subtractive model and can be applied to both "differences" and "ratios." However, if the judg-ments were assumed to represent ratios, then the values derived from Equation 3 would be logarithmically related to ratio model values.

Figure 5 plots the mean of the transformed judgments for both "differences" (dots linked by lines) and "ratios" (symbols as in Figures 1-4). The similarity of the points and the lines is consistent with the hypothesis of one oper-ation. The success of the transformation can be assessed from the parallelism of the transformed scores. Parallelism is consistent with either the subtractive or the ratio model. The ordinate in Figure 5, however, would be linearly spaced if the subtractive model were assumed, and it would be regarded as logarithmically spaced if the ratio model were assumed.

Each curve in Figure 5 represents an estimated scale of pitch for the second stimulus. By averaging across the different levels of the first stimulus, the five parallel separate curves can be summarized by one curve. Simi-larly, the vertical separation of points in each column represents a scale of pitch for the first stimulus, and can be summarized by averaging across levels of the second stimulus.

Figure 6 plots the marginal means of the transformed scores from Figure 5. Marginal means are estimates of
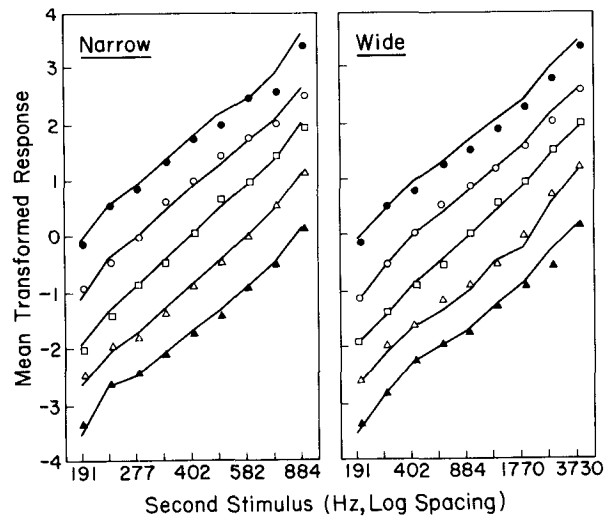

Figure 5. Mean transformed responses, plotted against the second stimulus with a separate curve for each level of the first, as in Figure 3. The dots linked by lines show the "difference" results and symbols plot the "ratio" results. The parallelism in the plots is an indication of the success of the transformation in removing response bias. Similarity of dots and lines is consistent with the theory that one operation underlies both tasks.
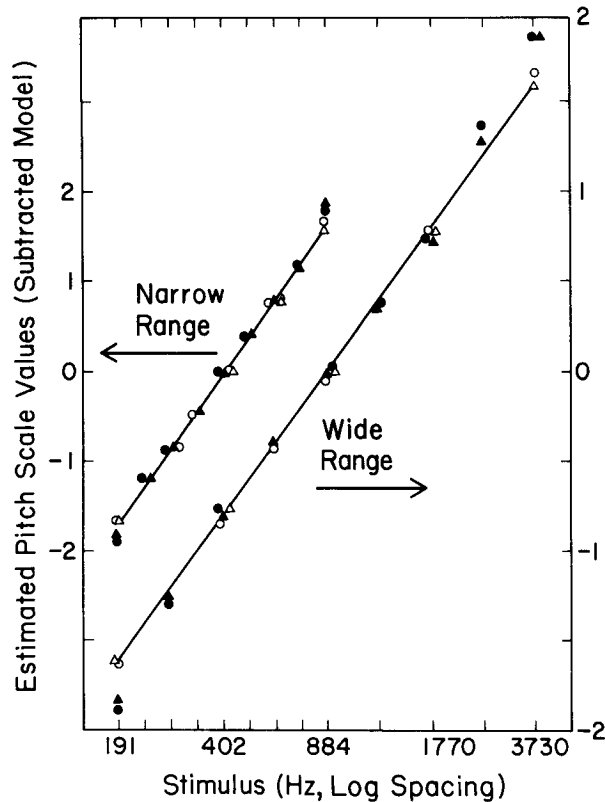
Figure 6. Subtractive model scales of pitch derived from Figure 5. Circles show data from the "difference" task and triangles show data from the "ratio" task. Filled symbols are estimates of scale values for the second stimulus; open symbols are estimates of scale values for the first stimulus. Data for the narrow-range condition should be read against the left ordinate; wide-range results should be read against the right scale. For both groups, a straight line has been drawn between the extreme open triangles. The abscissa is spaced in equal steps of log frequency.

s_j and t_i in Equation 3. Open symbols show the marginal means for the first simulus (t_i) and filled symbols for the second (s_j). The circles represent data from the "difference" task and the triangles represent data from the "ratio" task. Both the narrow-range data, displayed on the left, and the wide-range data are nearly linear when plotted against log frequency. Only the values for the highest and lowest second stimulus deviate from a line drawn between the extreme open triangles. The deviations may represent an end effect, because the irregularity at 844 Hz in the narrow range data is not evident in the wide range. If the subtractive model is assumed, then Figure 6 indicates that pitch is a log function of frequency.

Figure 6 provides a direct test of an idea suggested by Rule, Curtis, and Mullin (1981) as an explanation of the monotonic relationship between "ratios" and "differences." Rule et al. suggested that "ratio" scale values for the first stimulus (standard) would show regression to the mean, but "difference" scale values would not. Figure 6 shows no evidence of any systematic difference between circles and triangles ("differences" and "ra-

tios"): the effect of second versus first stimulus (filled vs. open) appears the same for both tasks, representing a slight order effect (Birnbaum, 1981). Mellers et al. (1984) noted that this analysis has also been done for studies of other continua with no support for the Rule et al. idea. Furthermore, Mellers et al. (1984, Appendix) developed the biased standard theory mathematically and showed that the idea cannot account for one rank order in a factorial design.

DISCUSSION

The results for comparisons of pitch are similar to previous results for loudness, heaviness, and other prothetic continua: the rank orders of "difference" and "ratio" judgments are virtually identical (Figures 1 and 5). This finding does not support the theory that subjects actually performed both ratios and differences (Figure 2). Attempts to fit the data to a two-operation model were not satisfactory. Instead, the data are consistent with the theory that subjects used only one comparison operation, despite instructions (Figure 5). The parallelism of the transformed scores (Figure 5) implies that this comparison operation can be represented by either a ratio or a subtractive model.

Differences in the metric form of the data for the two tasks (remember that Figure 4 would plot as a diverging fan if the ordinate were linear rather than log-spaced) can be explained by different judgment functions for the two tasks and the different ranges. If subjects performed differences throughout, the mapping of subjective differences to overt responses would be roughly linear for the "difference" judgments and roughly exponential for "ratios."

The conclusion that one operation underlies both tasks is consistent with the findings of Schneider, Parker, and Upenieks (1982), who concluded that "ratios" and "differences" of pitch can be represented by the same operation. Schneider et al. used 10 tones, varying from 460 to 1370 Hz, presented in a triangular design of tone pairs. The largest "ratio" judgment (1370:460) was 5.45 or 6.35 in two of their studies. Schneider et al. expressed the reservation that if their subjective range were too small, their data might be consistent with two operations. However, the present data with ranges of 191-844 Hz and 191-3730 Hz appear to reinforce the conclusion of one operation by reducing the reservation concerning stimulus range.

The Metathetic/Prothetic Classification

The data call into question the usefulness of the metathetic/prothetic classification, at least as applied to comparisons of pitch versus loudness. The conclusions for pitch are similar to our previous conclusions for loudness (Birnbaum & Elmasian, 1977). Birnbaum and Mellers (1978) found similar results for position, another so-called metathetic continuum. Since agreement between "ratio" and "difference" results is a defining feature of metathetic continua such as pitch or position (Stevens,

1957; Stevens & Galanter, 1957), it is tempting to agree with Warren and Warren's (1963) contention that the metathetic/prothetic distinction may be spurious.

The belief that the ratio model and the subtractive model yield the same pitch scales may have survived because few studies had contrasted pitch "ratios" and "differences" under comparable conditions when the assertion was made. However, there are findings that are consistent with the present results. Lewis (1942) reported that pitch scales derived from fractionation, bisection, and the equating of supraliminal pitch extents were in disagreement. Frazen, Nordmark, and Sjöberg (1972) also found nonlinearity between pitch scales derived from "ratio" techniques (which roughly approximated power functions) and pitch scales derived from category and Thurstonian scaling (which approximated logarithmic functions of frequency). Similarly, Schneider, Parker, Valenti, Farrell, and Kanow (1978) report that category ratings of pitch intervals were nonlinearly related to magnitude estimations of pitch intervals. Schneider et al. (1982) found that "ratios" and "differences" of pitch yield the same scale when one operation is assumed for both tasks.[2]

Stevens and Galanter (1957) had difficulty getting category scales to appear to agree with the mel scale. Their category scales were concave downward when plotted against mels, although magnitude estimations appeared to agree with mels. Only after differentially crowding stimuli at the higher frequencies or after using separate category scales for low and high frequencies were they able to get cateogry scales to approximate the mel scale. Since judgment functions can be altered by changes in stimulus spacing (e.g., Birnbaum, 1982; Mellers & Birnbaum, 1982), the claimed agreement between pitch "ratio" and category scaling can be questioned, especially in light of the present evidence.

Ratio or Subtraction?

While the present metric "ratio" and "difference" results clearly differ, the rank orders of judgments in the two tasks are essentially the same. Since this order is consistent with either the ratio or difference model, it is useful to consider additional evidence in deciding which model best represents the data.

Four lines of argument favor the subtractive model. First, an explanation has been offered for the exponential J_R function that makes "ratios" appear to fit the ratio model (Birnbaum, 1978, 1980; Birnbaum & Veit, 1974). This explanation can account for the effect of stimulus range and spacing on the function as well as the effect of the response examples on the J_R function (Birnbaum, 1982; Mellers & Birnbaum, 1982). It also helps explain individual differences in magnitude estimation. As yet, no alternative ratio theory of the numerical judgments (including contextual effects in estimations and ratings) has been offered. Second, if the ratio model is assumed, easterliness and westerliness of cities must be nonlinearly related. The subtractive model can explain easterliness and westerliness judgments with a single scale

that resembles the actual map (Birnbaum & Mellers, 1978). Third, "ratios of differences" and "differences of differences" show two rank orders, consistent with the instructed tasks, but "ratios of ratios" and "differences of ratios" can be represented by a difference of differences model (Birnbaum, 1978, 1979, 1980, 1982; Birnbaum, Anderson, & Hynan, 1984; Hagerty & Birnbaum, 1978; Veit, 1978). Fourth, the subtractive theory yields scales that agree across six data arrays and with range-frequency theory (Birnbaum, 1978, 1982). Eisler's (1978) attempt to save the ratio model (by postulating different scales for different tasks) requires complex post hoc arguments to explain an otherwise incorrect prediction (Birnbaum, 1979), and it does not explain the results of Birnbaum and Mellers (1978).

There is a special reason for prefering the subtractive model in the case of pitch: the resulting logarithmic psychophysical function is in agreement with the musical scale, probably the oldest and possibly the most widely used perceptual scale yet developed. On a log scale, equal physical ratios span equal subtractive distances, and in music equal ratios of frequencies span equal musical intervals. For example, for musicians, the pitch interval from 220 to 440 Hz is the same as the interval from 440 to 880 Hz, that is, one octave. With this representation, changes in key correspond to the addition of a constant and retain subjective distances among the notes within a melody. Attneave and Olson (1971) have demonstrated that nonmusicians as well as musicians transpose pitch intervals consistent with a logarithmic scale. Dowling (1982) and Shepard (1982a) also argue that psychophysical scales of pitch height should be logarithmic in order to explain perception of music.[3]

Many psychophysical studies, however, have assumed a ratio model and concluded that pitch is not a logarithmic function of frequency, much to the dismay of musicians and musically oriented psychologists. If the subtractive representation of pitch comparisons is accepted, however, perhaps psychophysicists and musicians will be able to resolve their discordant debate on a harmonious note.

REFERENCES

ANDERSON, N. H. (1974). Information integration theory: A brief survey. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), Contemporary developments in mathematical psychology (Vol. 2). San Francisco: Freeman.

ATTNEAVE, F., & OLSON, R. K. (1971). Pitch as a medium: A new approach to psychophysical scaling. American Journal of Psychology, 84, 147-166.

BALZANO, G. J. (1977). On the bases of similarity of musical intervals. Journal of the Acoustical Society of America, 61, 951.

BALZANO, G. J. (1982). Musical vs. psychoacoustical variables and their influence on the perception of musical intervals. Bulletin of the Council for Research in Music Education, 70, 1-11.

BIRNBAUM, M. H. (1978). Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), Cognitive theory (Vol. 3, pp. 33-74). Hillsdale, NJ: Erlbaum.

BIRNBAUM, M. H (1979). Reply to Eisler: On the subtractive theory of stimulus comparison. Perception & Psychophysics, 25, 150-156.

BIRNBAUM, M. H. (1980). A comparison of two theories of "ratio"

and "difference" judgments. *Journal of Experimental Psychology: General*, **109**, 304-319.

BIRNBAUM, M. H. (1981). Reason to avoid triangular designs in nonmetric scaling. *Perception & Psychophysics*, **29**, 291-293.

BIRNBAUM, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical measurement*. Hillsdale, NJ: Erlbaum.

BIRNBAUM, M. H., ANDERSON, C. J., & HYNAN, L. G. (1984). Two operations for "ratios" and "differences" of distances on the mental map. Submitted for publication.

BIRNBAUM, M. H., & ELMASIAN, R. (1977). Loudness "ratios" and "differences" involve the same psychophysical operation. *Perception & Psychophysics*, **22**, 383-391.

BIRNBAUM, M. H., & MELLERS, B. A. (1978). Measurement and the mental map. *Perception & Psychophysics*, **23**, 403-408.

BIRNBAUM, M. H., & VEIT, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics*, **15**, 7-15.

DOWLING, W. J. (1982). Musical scales and psychophysical scales: Their psychological reality. In R. Falck & T. Rice (Eds.), *Cross-cultural perspectives on music*. Toronto: University of Toronto Press.

EISLER, H. (1978). On the ability to estimate differences. A note on Birnbaum's subtractive model. *Pereception & Psychophysics*, **24**, 185-189.

FRAZEN, O., NORDMARK, D., & SJOBERG, L (1972). A study of pitch. *Goteborg Psychological Reports*, **2**, 1-31.

HAGERTY, M., & BIRNBAUM, M. H. (1978). Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics*, **24**, 121-129.

IDSON, W. L., & MASSARO, D. W. (1978). A bidimensional model of pitch in the recognition of melodies. *Perception & Psychophysics*, **24**, 551-565.

KRUMHANSL, C., & KESSLER, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, **89**, 334-368.

KRUSKAL, J. B., & CARMONE, F. J. (1969). MONANOVA: A FORTRAN-IV program for monotone analysis of variance. *Behavioral Science*, **14**, 165-166.

LEWIS, D. (1942). Pitch scales. *Journal of the Acoustical Society of America*, **14**, 127.

LINDSAY, P. H., & NORMAN, D. A. (1977). *Human information processing: An introduction to psychology*. New York: Academic Press.

MELLERS, B. A., & BIRNBAUM, M. H. (1982). Loci of contextual effects in judgment. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 582-601.

MELLERS, B. A., DAVIS, D., & BIRNBAUM, M. H. (1984). Weight of evidence supports one operation for "ratios" and "differences" of heaviness. *Journal of Experimental Psychology: Human Perception and Performance*, **10**, 216-230.

PAINTON, S. W., CULLINAN, W. L., & MENCKE, E. O. (1977). Individual pitch functions and pitch-duration cross-dimensional matching. *Perception & Psychophysics*, **21**, 469-476.

POULTON, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, **86**, 777-803.

RULE, S. J., CURTIS, D. W., & MULLIN, L. C. (1981). Subjective ratios and differences in perceived heaviness. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 459-466.

SCHNEIDER, B., PARKER, S., & UPENIEKS, E. G. (1982). The perceptual basis of judgments of pitch differences and pitch ratios. *Canadian Journal of Psychology*, **36**, 4-23.

SCHNEIDER, B., PARKER, S., VALENTI, M., FARRELL, G., & KANOW, G. (1978). Response bias in category and magnitude estimation of differences and similarity for loudness and pitch. *Journal of Experimental Psychology: Human Perception and Performance*, **4**, 483-496.

SHEPARD, R. N. (1982a). Geometrical approximations to the structure of musical pitch. *Psychological Review*, **89**, 305-333.

SHEPARD, R. N. (1982b). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music*. New York: Academic Press.

STEVENS, S. S. (1957). On the psychophysical law. *Psychological Review*, **64**, 153-181.

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

STEVENS, S. S., & GALANTER, E. H. (1957). Ratio scales and category scales for a dozen perceptual continua. *Journal of Experimental Psychology*, **54**, 377-411.

STEVENS, S. S., & VOLKMANN, J. (1940). The relation of pitch to frequency, a revised scale. *American Journal of Psychology*, **53**, 329-353.

STEVENS, S. S., VOLKMANN, J., & NEWMAN, E. B. (1937). A scale for the measurement of the psychological magnitude of pitch. *Journal of the Acoustical Society of America*, **8**, 185-190.

TORGERSON, W. S. (1961). Distances and ratios in psychological scaling. *Acta Psychologica*, **19**, 201-205.

VEIT, C. T. (1978). Ratio and subtractive processes in psychophysical judgment. *Journal of Experimental Psychology: General*, **107**, 81-107.

WARD, W. D. (1970). Musical perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 1, pp. 407-447). New York: Academic Press.

WARREN, R., & WARREN, R. P. (1963). A critique of S. S. Stevens' "new psychophysics." *Perceptual and Motor Skills*, **16**, 797-810.

ZWISLOCKI, J. (1965). Analysis of some auditory characteristics. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 3). New York: Wiley.

NOTES

1. For clear expositions of the disagreement between psychophysical and musical scales, see Lindsay and Norman (1977, p. 163) or Ward (1970).

2. Schneider et al. (1982) concluded that the subtractive model should be used to represent both "ratio" and "difference" judgments of pitch. In one experiment. Schneider et al. found that four of six subjects yielded subtractive model scales in agreement with the musical scale, but two of their subjects yielded estimated scales that were positively accelerated relative to the musical scale. Birnbaum (1981) noted that the usual nonmetric analysis of a triangular design can lead to estimated scale values that are nonlinearly related to the "true" scale values when order effects are present but unanalyzed. In particular, if scale values for the first stimulus are linearly related to scale values of the second stimulus but the first scale values show less variance (as in Figure 6), a triangular analysis that ignores this order effect will yield estimated scale values that are positively accelerated relative to "true" scale values. An order effect may therefore explain possible differences in scales between the present results and those of Schneider et al. (1978, 1982). It would be interesting to know if the musical scale with an additional parameter to represent the order effect would give a good fit to the Schneider et al. data.

3. A unidimensional scale such as the log function cannot capture all the features of the musical perception such as the repetition of tone chroma every octave, or the effect of musical key on relations among tones within a melody (Balzano, 1977, 1982; Dowling, 1982; Idson & Massaro, 1978; Krumhansl & Kessler, 1982; Shepard, 1982a, 1982b), though it provides a "well-tempered" starting point for the representation of pitch height.