

Fractionation of pitch intervals: An axiomatic study testing monotonicity, commutativity, and multiplicativity in musicians and non-musicians

Florian Kattner · Wolfgang Ellermeier

Published online: 17 June 2014
© Psychonomic Society, Inc. 2014

Abstract Direct ratio scaling (e.g., magnitude estimation; Stevens, 1956, *American Journal of Psychology*) is a widely used approach in psychophysics resting on the assumption that participants are able to refer to sensations with numbers that are valid on a ratio scale. Only recently, the necessary conditions of commutativity and multiplicativity have been formulated (Narens, 1996, *Journal of Mathematical Psychology*) and tested empirically, e.g., for the sensation of loudness. The present investigation evaluated these properties for the ratio production of pitch intervals. Musically trained ($n = 10$) and untrained ($n = 11$) participants adjusted intervals defined by two ascending pure-tone frequencies to given fractions ($1/3$, $1/2$, $2/3$), starting either from a 12 or 17-semitone standard pitch interval. The results show that the axioms of commutativity and multiplicativity held for most of the participants, irrespective of musical training. Furthermore, all participants produced larger frequency intervals in response to larger ratio numbers used in the instructions (monotonicity), but only musically trained participants were sensitive to the size of the standard interval (thus producing strictly increasing magnitudes). Overall, the results indicate that pitch intervals are ratio-scalable. However, restrictions must be made, especially for non-musical listeners, and when an octave is exceeded.

Keywords Direct scaling · Ratio production · Pitch · Commutativity · Multiplicativity · Monotonicity

F. Kattner (✉) · W. Ellermeier
Technische Universität Darmstadt, Darmstadt, Germany
e-mail: kattner@psychologie.tu-darmstadt.de

F. Kattner
Institute of Psychology, Technische Universität Darmstadt,
Alexanderstr. 10, 64283 Darmstadt, Germany

The issue of whether human sensations can be measured directly by asking the participant to assign numbers to sensations has a long history (e.g., Merkel 1888; Richardson 1929; Stevens 1956). Since the 1970s, direct scaling (e.g., magnitude or ratio estimation) has become one of the most widely used method of scaling in psychophysics (Stevens 1975; Gescheider 1997). In typical applications of direct scaling, participants are asked either to assign numerals to magnitudes of sensations (magnitude estimation), or to adjust the intensity of a stimulus corresponding to a given numeral (magnitude production; see, Stevens 1971). The methods of magnitude estimation and magnitude production have produced a large amount of data with various stimulus modalities showing that the magnitude of a sensation $\psi(s)$ can be described as a power function of the stimulus intensity s : $\psi(s) = as^b$, with $a > 0$ and $b > 0$ (Stevens 1975). However, as direct numerical scaling of sensations rests on the untested assumption that the numerical judgments made are multiplicatively related to the sensation magnitudes they are to describe, the validity of these scaling methods has been doubted (e.g., McKenna 1985; Parker & Schneider 1974; Shepard 1981). That is, even though Stevens (1946) himself distinguished between different scale types (e.g., ordinal, interval or ratio), psychophysicists have rarely tested whether participants are actually capable of processing sensation magnitudes on a ratio scale.

Only recently, and four decades after Stevens' proposal, did Narens (1996) formulate a comprehensive axiomatic theory, defining the fundamental conditions (axioms) that are inherent in the procedure of direct ratio scaling. In his 'behavioral axiomatization', an empirically testable system of axioms is provided distinguishing the numerals as used by the participant (behavior) from the mathematical numbers characterizing the sensations elicited by the physical

intensity of the stimuli. Following Narens (1996) terminology, the participant in a direct magnitude production experiment produces a stimulus x that is perceived as being \mathbf{p} times as intense as a standard stimulus t (in line with Narens (1996) terminology, ‘numerals’, i.e. number words, are written in boldface throughout this article to distinguish them from ‘true’ numbers). The participant’s behavior can thus be formalized as $(x, \mathbf{p}, t) \in E$; with the triple referring to a ratio production trial, and E referring to the set of all possible ratio productions (the triple is equally valid for magnitude estimation experiments where participants respond with the numeral \mathbf{p}).

Narens (1996) showed that the numerals as used by the participant are valid on a ratio scale only if the magnitude adjustments (or estimates) made by the participant meet the testable *commutativity property* (and a number of technical side conditions, including monotonicity; see below). Commutativity holds if the order of successive magnitude productions (e.g., first doubling and then tripling the loudness of a standard tone vs. first tripling and then doubling the loudness) is irrelevant for the outcome of the adjustments, and it can be formalized as follows:

$$\begin{aligned} \text{If } (x, \mathbf{p}, t) \in E, (z, \mathbf{q}, x) \in E, (y, \mathbf{q}, t) \in E, \\ \text{and } (w, \mathbf{p}, y) \in E, \text{ then } z \sim w. \end{aligned} \quad (1)$$

That is, a participant is adjusting a stimulus x that appears to be \mathbf{p} times as intense as the standard t , and then producing a stimulus z that is \mathbf{q} times as intense as the outcome of the first adjustment x . Equation 1 states that the same intensities should result after the second adjustments ($z \sim w$) if the adjustments were made in the opposite order (first \mathbf{q} and then \mathbf{p}).

In addition, the numerals \mathbf{p} can be interpreted as ‘true’ mathematical numbers (and as representing the corresponding sensation magnitudes) only if the stronger *multiplicativity property* holds (Narens 1996):

$$\begin{aligned} \text{If } (x, \mathbf{p}, t) \in E, \text{ and } (z, \mathbf{q}, x) \in E, \text{ and } r = pq \\ \text{then } (z, \mathbf{r}, t) \in E. \end{aligned} \quad (2)$$

Multiplicativity requires that two successive adjustments (e.g., $2 \times 3 \times$ as loud) match with a single adjustment (e.g., $6 \times$ as loud) if the number corresponding to the numeral in the single adjustment equals the product of the respective numbers in the successive adjustments. If both commutativity and multiplicativity hold, then a ratio scale is available directly by using the numbers corresponding to the numerals at face value: $\psi(x) = p\psi(t)$.

Testing these axioms for loudness production with 1000-Hz tones and ratio production factors of $\mathbf{p} = 2$ and $\mathbf{q} = 3$,

Ellermeier and Faulhammer (2000) found that most adjustments were consistent with the commutativity property, but not with the multiplicativity property. Zimmer (2005) reported very similar results for fractionations of loudness sensations ($\mathbf{p} < 1$). According to Narens (1996) axiomatization, these results indicate that loudness productions are valid on a ratio scale, but the numerals as used by the participants may not be interpreted as mathematical numbers (cf., Steingrímsson & Luce 2007), or that the actual sensation scale is not yet known. An analogous pattern of results has been obtained with brightness productions (Peißner 1999), showing that the multiplicativity axiom is violated in another modality. Moreover, Augustin and Maier (2008) recently evaluated the axioms of commutativity and multiplicativity (and monotonicity; see below) in the ratio production of area. Again, their results showed that successive area productions met the commutativity property, whereas most of the adjustments significantly violated multiplicativity. Taken together, there is evidence that magnitude productions of loudness, brightness, and area are valid on a ratio scale. However, it turned out to be difficult to derive the exact function that relates the numerals used by the participants to mathematical numbers (see, Steingrímsson & Luce 2007; Zimmer 2005). The present study is an attempt to test the axioms of ratio production with a different auditory sensation that may potentially be ratio-scalable with mathematically interpretable numeric values, namely the pitch of pure tones (e.g., due to the usage of certain numerical pitch ratios in musical intervals).

Pitch has been defined as the auditory attribute according to which sounds may be ordered from low to high, with high-frequency sounds typically being associated with higher pitch (ANSI 1973). However, such a monotonically increasing function of frequency may only apply to the pitch of pure tones (Stevens and Volkman 1940; Beck 1962), but not to the pitch of complex tones. The pitch of complex tones is best described by a two-dimensional model that distinguishes pitch height and tone chroma (e.g., Shepard 1964; Idson & Massaro 1978; Krumhansl & Shepard 1979), with pitch height referring to the frequency-dependent ordering of sounds (Stevens and Volkman 1940; Beck 1962), and tone chroma accounting for the enhanced similarity of tones that are separated by an octave (frequency relation of 1:2). That is, the pitch of a complex tone is affected by both the frequency and the periodicity of the tone (see Plack & Oxenham 2005; Plomp 1967). In a similar vein, others suggested that pitch is a mixture of (a) a ‘prothetic’ continuum (quantitative sensations) yielding psychophysical power functions of stimulus intensity (e.g., frequency) and (b) a more qualitative ‘metathetic’ continuum accounting, for instance, for octave equivalences (Stevens 1946, 1957; Stevens & Galanter 1957, arguing that ‘prothetic’ and ‘metathetic’ continua differ with regard to

several functional criteria, e.g. the size of just noticeable differences, time-order errors, or hysteresis effects). In order to avoid the two-dimensionality of pitch, the present investigation of ratio-scaling properties of pitch perception focuses on pure tones, with spectral and periodicity components being identical.

The category labels associated with musical intervals (e.g., a major third or a perfect fifth) imply a pitch model that assumes a logarithmic function of frequency. However, early efforts of direct pitch scaling cast doubt into such a simple logarithmic relation between pitch and frequency (e.g., Stevens et al. 1937; Stevens & Volkman 1940). In typical direct-scaling approaches to study pitch perception, listeners are either asked to adjust the frequency of a comparison tone until it appears to be half as high as the standard tone (method of fractionation; Stevens et al. 1937), or to adjust frequency intervals corresponding to ‘equal-sense-distances’ (Stevens and Volkman 1940). Alternatively, magnitude estimates have been obtained by asking the participants to assign numerals to frequency intervals (e.g., Beck & Shaw 1961, 1962). These methods were used, for instance, to derive the *mel*-scale of pitch height, which describes a roughly logarithmic relation between pitch and frequency with a 1000 Hz pure tone, 40 dB above threshold, being defined as 1000 mels (e.g., Siegel 1964; Stevens & Volkman 1940; Stevens & Galanter 1957, but see Painton et al. 1977). Supporting the validity of direct numerical scaling, Parker and Schneider (1974) demonstrated that similar representations of pitch magnitudes (i.e., in line with the mel scale) can be obtained with non-metric scaling techniques. Others, however, questioned the reliability of direct pitch judgments (e.g., Rasch & Plomp 1999). Since there have been no attempts so far to address the problem from the perspective of representational measurement theory, the fundamental axioms of direct ratio scaling were evaluated for the pitch of pure tones in the present study.

As the sensitivity to relations between pitches (pitch intervals) is fundamental to musical experience, we decided to investigate the fractionation of pitch intervals. Particularly, fractions of pitch intervals were obtained by using a generalized form of the ratio production procedure (as introduced by Luce 2002, 2004). In a generalized ratio production task, participants match an interval between y and x (i.e., by adjusting x) to a certain proportion \mathbf{p} of a standard interval from y to t , that is, they set x to a certain proportion \mathbf{p} of the interval from y to t (Luce’s terminology: $x := t \circ_{\mathbf{p}} y$). This procedure generalizes typical ratio production (Stevens 1957) which implicitly assumes a reference of $y = 0$ (such as, e.g., an absolute threshold in the case of loudness). As the reference on the pitch continuum may be less obvious, the generalized procedure appears to be appropriate for the purpose of pitch scaling. Like in a standard

ratio production experiment, a set of pitch-interval adjustments can also be used to derive a scale of pitch magnitudes (in fact, simple pitch adjustments would just be a special case of pitch-interval adjustments; Luce (2002)). For the sake of convenience, we will adhere to Narens (1996) terminology in this article, defining generalized ratio productions as $(x, \mathbf{p}, t - y) \in E$, where the participant is adjusting x to the ratio \mathbf{p} of the standard interval from y to t .

Another property that is inherent in ratio production measurements and may be relevant to the scaling of pitch intervals is the monotonicity axiom assuming the ordering of numerals to correspond to the usual ordering of numbers (see Augustin 2006; Narens 1996, p. 114, Axiom 3.1). Empirically, this property means that the order of two (or more) ratio productions with the same standard but different production factors should preserve the order of the mathematical numbers corresponding to the production factors (e.g., half a pitch interval should always be less than two thirds of the same pitch interval). With regard to ‘generalized-ratio-production’ experiments, it can be formalized as follows:

$$\text{If } (x, \mathbf{p}, t - y) \in E \text{ and } (z, \mathbf{q}, t - y) \in E, \text{ then } p > q \Leftrightarrow x > z, \quad (3)$$

with $>$ referring to a numerical order, and the $>$ sign denoting an ordering of empirical adjustments. However, the monotonicity property also implies that magnitude productions correspond to the order of different standards (compare the assumption of magnitudes to behave in a strictly increasing manner; Narens 1996, p. 114, Axiom 2.5). That is, two ratio productions with identical production factors but based on different standard intervals $t - y$ and $s - y$ should preserve the order of the magnitudes (frequency differences) of the standards (e.g., a given fraction of an octave should be smaller than the same fraction of an octave-exceeding pitch interval):

$$\text{If } (x, \mathbf{p}, t - y) \in E \text{ and } (z, \mathbf{p}, s - y) \in E, \text{ then } t > s \Leftrightarrow x > z. \quad (4a)$$

In principle, in generalized ratio production experiments, the size of the standard interval can also be altered by changing the magnitude of the lower reference y . Actually, the present investigation contains a manipulation of the lower reference of the standard interval (in order to test whether participants are capable of using variable pitch references; see below). The monotonicity property with regard to the order of the standards can then be assumed if Eq. 4b holds:

$$\text{If } (x, \mathbf{p}, t - y) \in E \text{ and } (z, \mathbf{p}, t - w) \in E, \text{ then } y > w \Leftrightarrow x > z. \quad (4b)$$

Both types of monotonicity (Eqs. 3 and 4b) were tested empirically for generalized pitch productions in the present study.

We expected that the fractionations of intervals within an octave might be different from the adjustments of intervals larger than an octave as the latter ones are extremely rare in musical melodies (Russo and Thompson 2005). Therefore, and to control for the effects of tone chroma on the perception of pitch intervals, fractionations of pitch intervals were obtained with two different standard intervals: an octave interval (12 semitones between the musical notes F4 and F5; Young 1939) and a larger pitch interval (17 semitones; C4 to F5). Previous studies showed that particularly the perceived size of pitch intervals up to an octave depends on musical training. For instance, the estimates of pitch intervals made by musically trained participants (on a scale ranging from 1 to 100) were more differentiated and closer to a logarithmic function of (fundamental) frequency than those made by musically untrained participants (Russo and Thompson 2005). For intervals larger than an octave, differences between musically trained and untrained participants diminished. Therefore, we decided to separately evaluate the fundamental axioms of ratio scaling in pitch adjustments for musically experienced and unexperienced participants within and beyond the frequency range of an octave.

Method

Participants

Twenty-one participants (11 female) were recruited for individual testing. Four of them were members of the psychoacoustics laboratory, the remainder primarily psychology students. Ages ranged between 19 and 55 ($M = 30.0$; $SD = 9.0$) years. Ten participants (5 female), reporting more than 7 years of musical instruction and presently continuing musical activity were categorized as ‘musicians’ or ‘musically trained participants’. The remaining 11 musically untrained participants will be referred to as the ‘non-musicians’ group. All participants had normal hearing with thresholds not exceeding 20 dB with respect to the audiometric reference (ISO 1998) in the frequency range between 125 and 8000 Hz.

Stimuli and apparatus

445 pure sine tones for all integer frequencies between 259 Hz and 703 Hz were generated digitally with a sampling rate of 44.1 kHz. Each tone had a duration of 250 ms including 20-ms cosine-shaped rise and decay ramps. The signals were D/A converted by an external sound card (RME multiface II), passed through a Behringer HA 8000 Powerplay

PRO-8 headphone amplifier and played back diotically with Beyerdynamics DT 990 PRO (250 Ohm) headphones. The sounds were attenuated to comfortable levels corresponding to approx. 70 dB SPL. Since both the frequency response of the headphone, and the equal-loudness contours at this level are relatively flat in the frequency range studied, no salient loudness differences were expected (Fletcher and Munson 1933).

The experiment was conducted in a single-walled sound-attenuated listening room (International Acoustics Company). Visual instructions were displayed on a 19” TFT monitor. The participants used a standard keyboard (Cherry) for the adjustments. Stimulus presentation and response registration was programmed in MATLAB utilizing the Psychophysics Toolbox extensions (Brainard 1997; Pelli 1997).

Procedure

Each participant was tested in three experimental sessions. In each session, the participant completed 50 pitch-adjustments trials. There was a short break after 30 trials in each session. The first session started with an additional training block consisting of 10 adjustments that were not included in the analysis.

In each trial ($x, p, t - y$), two ascending pitch intervals (a lower tone being followed by a higher tone) were presented successively to the participant via headphones. The first pair of tones defined the standard interval (frequencies f_y and f_t), and the second pair was the comparison interval (f_y and f_x). The two tones of an interval were separated by a silent gap of 450 ms, and the two intervals were separated by a 750-ms gap (see Fig. 1 for an illustration of the procedure; c.f. Steingrímsson & Luce 2005, p. 312). The participants’ task was to adjust the second tone in the comparison interval to a frequency that makes the subjective

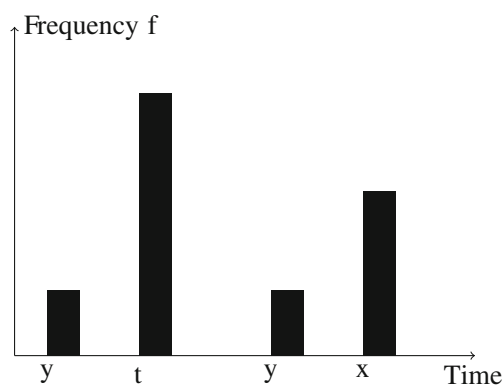


Fig. 1 Illustration of the ‘generalized ratio production’ procedure, applied to the production of pitch intervals, with x to be adjusted to a subjective fraction p of the interval from y to t

interval from f_y to f_x stand in the proportion \mathbf{p} to the standard interval. In each trial, the value of the fraction \mathbf{p} ($1/3$, $1/2$ or $2/3$) was shown on the screen, together with a short text instruction and a graphical illustration representing the pitches of the four tones (at target state) as horizontal bars. Participants were explicitly requested not to adjust the loudness but the pitch height of the tone. At the beginning of each trial, the frequency of the to-be-adjusted tone f_x was chosen randomly from a uniform distribution of frequencies between f_y and f_t . The participants could decrease or increase f_x by pressing the left or right cursor key, respectively. The resulting frequency increment or decrement of f_x was determined by a logarithmic function of the current frequency: $\Delta f_x = (10 \cdot (\log(f_x/262) + 1))\text{Hz}$. The adjustments could be accelerated by pressing the “Shift” key together with the respective cursor key (in that case, the frequency increment Δf_x was multiplied by 10). After each key press, both pitch intervals were played again with f_x having changed in frequency. The participants were encouraged to repeat the adjustments until they were satisfied with the frequency ratio. There was no time limit to the task. An adjustment was confirmed by pressing the “Enter” key. The final frequency adjustment was recorded, and the next trial started after a 1-s delay.

There were ten different adjustment conditions. The standard was either a 12-semitone interval ($f_y = 349$, $f_t = 699$ Hz; corresponding to the octave from F4 to F5 on a standard twelve-tone equally-tempered chromatic scale) or a 17-semitone interval ($f_y = 262$, $f_t = 699$ Hz; C4 to F5).¹ For both standard intervals, the participants adjusted a comparison interval that appeared to be $1/3$, $1/2$ or $2/3$ as large as the standard. In addition, there were two successive adjustments in which $2/3$ or $1/2$ of a previously fractionated interval ($1/2$ or $2/3$ of the standard interval, respectively) had to be produced. In line with Narens (1996), the notation $\mathbf{p}\bullet\mathbf{q}$ is used for two concatenated adjustments \mathbf{p} and \mathbf{q} : First, f_x is adjusted with respect to the subjective ratio \mathbf{p} of the standard interval from f_y to f_t , and then that interval from f_y to f_x serves as the standard interval in the successive adjustment trial in which the comparison tone is adjusted to the subjective ratio \mathbf{q} (see Eqs. 1 and 2). In each of the three sessions, the participants completed 5 blocks consisting of

the ten different adjustments conditions, thus accumulating 15 trials per condition. The order of the adjustments within each block was randomized (of course, with single adjustments being made prior to successive adjustments).

Results

Quality of pitch-interval adjustments

The quality of the adjustments may be evaluated based on some descriptive statistics. For each participant, standard deviations were computed for the fifteen repeated frequency adjustments within each condition. The mean standard deviation was 47.2 Hz (ranging between $SD = 3.7$ Hz for participant MH and $SD = 79.2$ Hz for participant CL). Variability was slightly smaller with the octave standard ($SD = 46.3$ Hz) than with the 17-semitone standard ($SD = 52.6$ Hz).

On average, the participants made $M = 11.8$ ($SD = 5.3$) key presses per adjustment trial before confirming the adjusted frequency (ranging between an average of 5.1 and 26.5 key presses for participants DS and XS, respectively). They used $M = 5.8$ fine steps of frequency adjustments per trial (ranging between 0.68 and 15.6 for participants DS and JB, respectively). In only 10.7 % of all trials, did participants not make any fine adjustments of frequency. There were $M = 12.4$ (7.0 fine adjustments) key presses per trial in session 1, $M = 11.2$ (5.8 fine adjustments) in session 2, and $M = 11.6$ (6.1 fine adjustments) in session 3.

Pitch adjustments of musicians and non-musicians

Figure 2 illustrates the mean adjustments of f_x (Hz) per fractionation condition and standard interval separately for musically trained and untrained participants. As the distributions of our estimates are unknown, non-parametric tests have been used predominantly for statistical evaluation in the present paper. However, we conducted an additional analysis of variance in order to test for possible interactions between Numerals, Standards, and Musical Training.

A 3 (Numeral: $1/3, 1/2, 2/3$) \times 2 (Standard: 12 and 17 semitones) \times 2 (Musical Training: trained and untrained) mixed-effects analysis of variance (Type II, Huynh-Feldt corrections applied) on the single adjustments (unconnected points above the arrows in Fig. 2) with repeated measurements on the factors Numeral and Standard revealed a significant main effect of Numeral, $F(2, 38) = 218.94$; $p < .001$; $\eta^2_{generalized} = 0.64$ (Huynh-Feldt $\epsilon = 0.87$), and a main effect of Standard, $F(1, 19) = 46.45$; $p < .001$; $\eta^2 = 0.14$, confirming that the adjustments were sensitive both to the numerals in the fractionation instructions and to the extent of the standard

¹We decided to manipulate the frequency of the lower reference of the standard f_y , rather than the upper reference f_t to compare the adjustments between two different standards. This was done in order to direct attention to the fact that the sizes of standard intervals varied. If the lower reference would have been constant in all adjustment trials, e.g. 349 Hz, then participants might have learned to ignore the lower reference. By varying the lower reference, it is possible to test whether participants are adjusting f_x with respect to f_t (indicating that participants use a constant, octave-dependent lower pitch reference f_y) or whether they refer to a variable reference as defined by the standard frequency intervals.

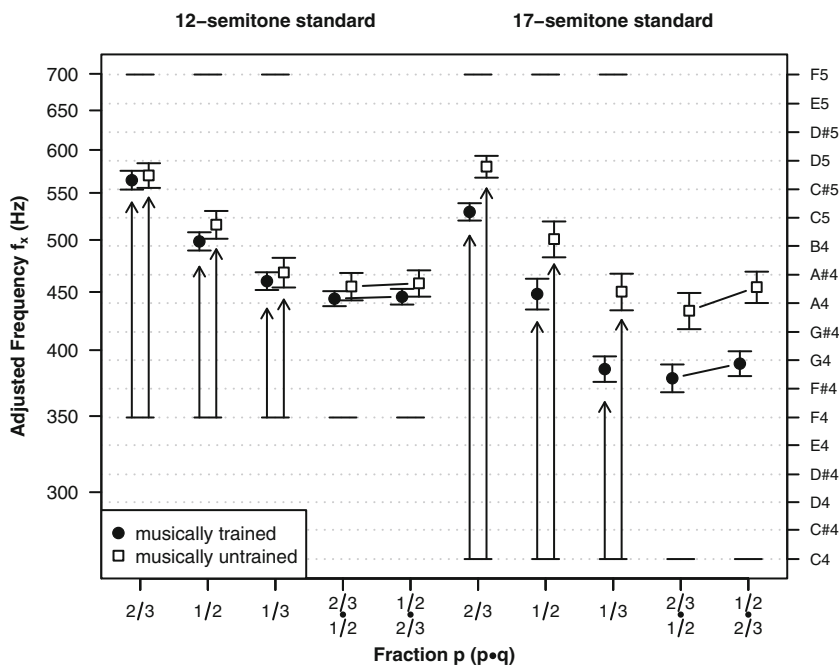


Fig. 2 Mean adjustments of f_x as a function of the standard interval $f_i - f_y$ (12 and 17 semitones) and the fraction p to be adjusted (logarithmic frequency scale). The frequencies f_y (349 and 262 Hz) and f_i (699 Hz) defining the size of the standard intervals are indicated by short horizontal lines. Vertical arrows indicate the frequency intervals resulting from single adjustments ($f_x - f_y$). Frequencies resulting from successive fractionations ($p \bullet q$) are connected by solid lines (test of commutativity). The names of the corresponding musical notes are shown on the ordinate to the right

pitch interval. There was also a significant Numeral \times Standard interaction, $F(2, 38) = 12.67$; $p < .001$; $\eta^2 = 0.03$ (Huynh-Feldt $\epsilon = 0.79$). In addition, there was a main effect of Musical Training, $F(1, 19) = 5.38$; $p = .03$; $\eta^2 = 0.17$, and a significant interaction between Musical Training and Standard, $F(1, 19) = 28.37$; $p < .001$; $\eta^2 = 0.09$, indicating that differences between musically trained and untrained participants were more prominent with the 17-semitone standard interval than with the octave standard interval (compare Fig. 2). There was no significant Musical Training \times Numeral interaction, $F(2, 38) = 0.31$; $p = .73$; $\eta^2 = 0.002$, and no three-way interaction, $F(2, 38) = 1.28$; $p = .29$; $\eta^2 = 0.003$.

F-tests showed that variances of single adjustments differed significantly between the two standard intervals, $F(62, 62) = 2.05$; $p < .01$, with the mean adjustments scattering more in the 17-semitone interval ($SD = 75.9$ Hz) than in the 12-semitone interval ($SD = 53.0$ Hz). However, there were no significant differences in variance between musically trained and musically untrained participants, $F(65, 59) = 0.80$; $p = .37$, nor did the variance of single adjustments differ between the fractions to be adjusted, Bartlett's $K^2(2) = 3.04$; $p = .22$.

It is also interesting to see (compare Fig. 2, right ordinate) that the musicians' adjustments of pitch intervals correspond better to the number of semitones defining the respective fractions of the standard interval (e.g., 4,

6 and 8 semitones within the 12-semitone standard interval). A difference between musically trained and untrained listeners is particularly evident in the octave-exceeding standard interval. That is, although they still overestimated the intervals (probably due to hysteresis; see below), the musicians' adjustments of 1/3, 1/2, and 2/3 of the 17-semitone interval fell closer to 5.7, 8.5, and 11.3 semitones, compared to those of the musically-untrained listeners.

Evaluation of the monotonicity property

To assess the monotonicity of the pitch productions, the single adjustments (unconnected points in Fig. 2) were compared with respect to (a) the fractions to be adjusted (Eq. 3), and (b) the two standard intervals (Eq. 4b). Therefore, for each participant, the mean adjustments of f_x were compared descriptively for increasing numerals p (separately for both standards) and between standards (separately for all fractions).

In order to evaluate the validity of Eq. 3, we determined for each participant and both standard intervals, whether the adjusted frequency for $p = 2/3$ was greater than for $p = 1/2$, and whether it was greater for $p = 1/2$ than for $p = 1/3$ (resulting in 84 comparisons; 4 for each of the 21 participants). In 83 of these 84 cases (98.8 %), the adjusted frequency f_x increased with the numeral of the fraction p

(participant CL adjusted higher frequencies to $p = 1/3$ than to $p = 1/2$ of the 17-semitone standard).

To evaluate Eq. 4b, there was one pairwise comparison between standards for each participant and for each fraction to be adjusted (resulting in 63 comparisons; 3 per participant). As the two standard intervals differed with regard to the lower tone (262 vs. 349 Hz), frequency adjustments with the 17-semitone standard should be lower than those made with the 12-semitone standard in order to meet Eq. 4b. In 49 of 63 cases (77.7 %), the frequency adjusted to the same numeral p was lower when based on the 17-semitone standard interval than when based on the 12-semitone standard interval (compare Table 1 and the means depicted in Fig. 2). It is interesting to see that monotonicity with regard to the size of the standard interval (4b) was satisfied in 28 of 30 cases (93.3 %) by musically trained participants, but only in 21 of 33 cases (63.6 %) by musically untrained participants. Moreover, for the latter sample, the number of these monotonicity violations increased with the fractions p to be adjusted (2 with $p = 1/3$, 3 with $p = 1/2$, and 9 with $p = 2/3$).

In addition to these descriptive comparisons, two rank-order Kruskal-Wallis analyses of variance were conducted for each participant to test for statistically significant differences in the adjusted frequencies f_x as a function of the numeral p and the standard interval $t - y$ (data were pooled across standard intervals and numerals, respectively; within each analysis, frequency adjustments were treated as independent observations). The results of these tests are listed in Table 1. As can be seen, all participants produced statistically distinguishable adjustments as a function of the fractionation condition (numeral). However, the results also confirm the descriptive differences between musicians and non-musicians regarding sensitivity to the size of the standard interval. Whereas most of the musicians' frequency adjustments systematically varied as a function of the standard interval (that is, lower adjusted frequencies of x in case of the 17-semitone standard interval, compare Table 1), only 2 of 11 musically untrained participants produced statistically distinguishable frequencies for the two different standards intervals. The lack of sensitivity to the size of the standard interval in most of the non-musicians' frequency adjustments is inconsistent with Eq. 4b and thus violates one form of the monotonicity axiom. This will be discussed in greater detail below.

Evaluation of the commutativity property

The commutativity axiom holds if the outcome frequency of successive $p \bullet q$ ($1/2 \bullet 2/3$) adjustments does not differ from $q \bullet p$ ($2/3 \bullet 1/2$) adjustments. This was tested for each participant and both standard intervals by means

of Mann-Whitney U tests (two-tailed). These 42 tests were conducted without corrections ($\alpha = .05$) for multiple comparisons, because the aim of these tests was to accept the statistical null hypothesis ($\alpha < .05$ corrections would make it more likely to accept the null hypothesis). Table 2 lists the individual mean frequencies produced by the two types of successive adjustments together with respective z -scores of U values and the p -values for the octave standard. Table 3 lists the respective values for the 17-semitone standard interval.

Tables 2 and 3 show that there were 5 violations of commutativity with the octave standard (only musically trained participants) and 3 with the 17-semitone standard (one musically trained participant). Thus, for the pitch adjustments of most of the musically trained and untrained participants, the commutativity property holds. The few significant axiom violations might have different causes in the sample of musicians and non-musicians. On the one hand, only 2 musically untrained participants showed a violation of commutativity with one of the two standards. Thus, the observed rate of violations (.09) is consistent with what is expected by chance as the a-priori probability for obtaining one of two significant violations is $p = .095$ (given $\alpha = .05$). On the other hand, the magnitude of axiom violations in musically-trained participants (the differences between the two types of successive frequency adjustments) was relatively small ($MD = 6$ Hz; range: 3–27 Hz). Particularly, the higher precision of the adjustments made by the musically trained participants (lower SD s in Tables 2 and 3) may have increased the likelihood of significant effects (4×1 violation and 1×2 violations).

However, the variance of successive adjustments did not differ significantly as a function of Musical Training, $F(43, 39) = 1.34$; $p = .35$. Variances did not differ between the two orders of successive adjustments, either, $F(41, 41) = 1.04$; $p = .89$, but they differed between the two standards, $F(41, 41) = 3.13$; $p < .001$ (with $SD_{17-semitones} = 56.4$ Hz; $SD_{12-semitones} = 31.8$ Hz) - as the single adjustments did.

Evaluation of the multiplicativity property

Multiplicativity of pitch adjustments holds if the frequency that has been adjusted to be the fraction r (with $r = p \cdot q$) ($1/3$) of the standard equals the frequency that has been adjusted to be p ($1/2$), and q ($2/3$) of that outcome, subsequently, or vice versa (see Eq. 2). This was tested by means of Mann-Whitney U tests (two-tailed) for each participant and both standard intervals (see Table 4). For the same reasons as with the commutativity tests, no corrections for multiple comparisons were made ($\alpha = .05$).

Table 4 shows that there were 7 violations of multiplicativity with the octave standard and 5 with the 17-semitone

Table 1 Single adjustments of f_x (mean frequencies in Hz) for 1/3, 1/2, or 2/3 of a 12- or 17-semitone standard interval

Participant	12-semitone standard			17-semitone standard			Numeral $\chi^2(2)$	Standard $\chi^2(1)$
	1/3	1/2	2/3	1/3	1/2	2/3		
Musically untrained participants								
AK	517	615	639	545	609	645	63.25**	1.42
CL	501	518	569	531	509	601	16.23**	0.36
CW	448	491	566	441	485	553	65.08**	0.00
FK	446	503	528	386	470	519	49.70**	4.52*
JB	463	481	534	456	483	559	11.62**	0.20
JK	445	473	533	388	468	562	36.06**	0.30
JS	458	533	598	449	477	534	24.80**	5.33*
MD	506	537	546	434	520	602	20.77**	0.02
MS	444	495	584	424	480	586	53.41**	0.12
SD	448	505	563	434	479	579	63.15**	0.38
WE	474	522	606	467	530	639	48.74**	0.22
Musically trained participants								
AR	447	472	568	360	427	526	65.13**	12.44**
AS	450	489	553	370	448	522	68.46**	12.75**
DD	474	527	562	406	462	537	30.48**	8.44**
DS	430	509	597	357	445	599	54.71**	3.91*
JM	450	455	577	378	448	557	36.91**	1.72
MH	441	496	554	364	428	498	18.17**	8.94**
SB	452	488	523	344	420	464	33.98**	34.49**
US	549	560	604	480	498	567	23.13**	16.02**
XS	422	479	540	329	394	469	36.36**	24.06**
YR	487	512	566	462	513	554	24.00**	0.08

* $p < .05$; ** $p < .001$

Kruskal-Wallis rank sum tests were used to test for the monotonicity with regard to the Numerals and Standard Intervals (last two columns)

standard. For most of these violations, the sign of the adjustment difference ($M_{1/3} - M_{1/2 \bullet 2/3}$) was positive, indicating that the single 1/3 pitch adjustments produced larger intervals than the successive adjustments.

Discussion

The aim of the present investigation was to test Narens’ (1996) fundamental axioms of direct ratio scaling for the perception of pitch intervals. The data show that, for most of the participants, *commutativity* and *multiplicativity* hold. However, differences are evident between musically trained and untrained participants with the former producing slightly more axiom violations. In addition, the data were checked for *monotonicity* with regard to (a) the fractions to be adjusted and (b) the size of the standard interval. Again, differences between musically trained and untrained participants were found, with the latter failing to produce pitch intervals that increase strictly with the extent of

the standard interval (compare the left and right panel in Fig. 2).

Commutativity and multiplicativity

In line with the commutativity axiom, in most of the cases (81.0 %), the successive concatenation of two fractionations (e.g., producing $1/2 \bullet 2/3$ of a standard pitch interval) converged on approximately the same frequency irrespective of the order of the two operations. According to Narens (1996) and Luce (2002), this implies that the participants use a ratio scale when producing intervals of pitch height. Nevertheless, the few violations of commutativity are worth considering. There were two violations of commutativity in the sample of musically untrained participants. However, as the relative frequency of observed violations nearly coincides with the probability to obtain one violation with two standards by chance ($p = .095$ given $\alpha = .05$), this pattern of results is in line with what would be expected if commutativity held.

Table 2 Successive adjustments (mean frequencies in Hz) and Mann-Whitney tests of Narens' (1996) axiom of commutativity with the octave standard interval ($z_{critical}(U) = 1.96$)

Participant	Frequency adjustment (Hz)				Commutativity violation $z(U)$
	$M_{1/2 \bullet 2/3}$	$SD_{1/2 \bullet 2/3}$	$M_{2/3 \bullet 1/2}$	$SD_{2/3 \bullet 1/2}$	
Musically untrained participants					
AK	549	63	561	46	0.52
CL	461	83	447	55	0.06
CW	440	26	424	44	0.60
FK	427	28	439	33	0.77
JB	440	56	426	52	0.81
JK	413	44	423	57	0.04
JS	467	50	443	57	1.22
MD	435	51	472	59	1.89†
MS	459	43	448	46	0.97
SD	452	34	444	32	0.46
WE	496	41	477	57	0.81
Musically trained participants					
AR	444	3	448	5	2.55*
AS	444	11	437	10	2.80*
DD	448	49	454	40	0.37
DS	442	80	451	68	0.10
JM	424	44	413	36	0.58
MH	440	1	437	2	1.98*
SB	437	10	444	11	2.20*
US	520	31	493	32	2.38*
XS	419	21	429	26	1.20
YR	439	23	435	31	0.50

† $p < .1$; * $p < .05$

In contrast, in the sample of musically trained participants, there were 5 violations of commutativity with the octave standard interval, and one violation with the 17-semitone standard. Thus, the number of musically trained listeners producing axiom violations with an octave standard is somewhat higher than what would have been expected by chance alone. The probability to observe significant axiom violations in the sample of musicians might have been enhanced by the higher precision (and reliability) of their adjustments within the octave interval (indicated by these participants' small standard deviations in Table 2). Nevertheless, the frequency adjustments observed cast doubt on the validity of a ratio-scale representation of pitch intervals in five musically-trained listeners. Potentially, their profound musical experience biased their judgments of frequency distances. According to Greenwood (1997, p. 203), particularly non-musical listeners should be able to judge pitch intervals as sensory distances (i.e., frequency differences), whereas musically trained listeners are likely to be unable to judge the distance without referring to the 'interval quality' (the octave reference). As

pitch intervals within an octave are frequently encountered in music, musically experienced participants may perceive them differently than musically unexperienced participants. In addition to an ordering from low to high, musicians may experience the intervals as consonant or dissonant or they may associate particular labels with the intervals (e.g., their names on the chromatic scale). These strategies may have biased their adjustments in ways presently not completely understood. Interestingly, with the 17-semitone standard, musicians did not produce more commutativity violations than non-musicians. This is consistent with previous studies reporting that, particularly for musically trained participants, the differentiation of pitch intervals is better within an octave than for intervals exceeding an octave (in the former case, magnitude estimates of pitch intervals increased faster in musically trained listeners; Russo & Thompson 2005).

In addition to commutativity, we tested Narens' (1996) axiom of multiplicativity which holds, if two successive $1/2 \bullet 2/3$ pitch adjustments do not differ from a single $1/3$ adjustment. This was the case for most of the pitch-interval

Table 3 Successive adjustments (mean frequencies in Hz) and Mann-Whitney U tests of Narens' (1996) axiom of commutativity with the 17-semitone standard interval ($z_{critical}(U) = 1.96$)

Participant	Frequency adjustment (Hz)				Commutativity violation $z(U)$
	$M_{1/2 \bullet 2/3}$	$SD_{1/2 \bullet 2/3}$	$M_{2/3 \bullet 1/2}$	$SD_{2/3 \bullet 1/2}$	
Musically untrained participants					
AK	583	14	584	30	1.04
CL	480	82	451	91	0.87
CW	405	58	419	42	1.51
FK	401	48	395	42	0.02
JB	401	72	418	69	0.71
JK	418	65	380	59	1.41
JS	413	58	397	69	0.89
MD	465	95	404	118	1.53
MS	444	44	432	70	0.08
SD	449	29	408	32	2.88**
WE	542	48	479	54	2.80*
Musically trained participants					
AR	355	3	355	10	0.77
AS	370	7	373	18	1.37
DD	401	61	386	68	0.75
DS	426	60	392	76	0.81
JM	413	87	384	81	1.02
MH	354	5	359	2	1.98*
SB	352	30	345	14	0.48
US	448	51	440	43	0.48
XS	355	42	340	52	0.95
YR	419	32	405	48	0.91

* $p < .05$; ** $p < .001$

fractionations produced in musically trained (70.0 %) and untrained (72.7 %) participants. The violations of multiplicativity were evident for both musically trained and untrained participants alike and with both standard intervals. This indicates that the violations were neither due to systematic effects of musical training, nor to interactions between tone chroma (or category labels) and pitch height. In contrast to the high number of multiplicativity violations obtained in ratio productions of loudness and area (Ellermeier and Faulhammer 2000; Zimmer 2005; Augustin and Maier 2008), and given Narens's own prediction that the multiplicativity property will fail in most situations of magnitude estimation (Narens 1996, p. 110), the low number of multiplicativity violations in the fractionation of pitch intervals is rather surprising. The results imply that, in most cases, the numerals used in the instructions have been treated like 'true' mathematical numbers by the listeners. Consequently, the numerals p can be taken at face value, and the perceived magnitude of the pitch intervals may be represented on a ratio scale characterized by these numerals without further transformation.

The overall consistency of (generalized) pitch productions (i.e., adjustments of pitch with respect to a frequency interval; Luce 2002) with commutativity implies that listeners are representing pitch on a ratio scale continuum. This argues for the validity of applying direct scaling techniques like magnitude production or magnitude estimation to pitch sensations of pure tones (distinctions in terms of the instructions given to the participants should be of minor importance, given that a ratio-scale representation is shown; Narens 2002). The fact that the stronger multiplicativity property held in most of the cases indicates that the numerals presented to the participants may be taken at face value, suggesting that numerals from arbitrary ranges referring to pitch sensations (e.g., absolute magnitude estimation) are likely to be interpretable on a ratio-scale level.

Monotonicity

However, the conclusion that the representation of pitch intervals is valid on a ratio scale rests on the assumption that

Table 4 Frequency differences between single $1/3$ and successive ($1/2 \bullet 2/3$ or $2/3 \bullet 1/2$) adjustments, and Mann-Whitney U tests of Narens' (1996) axiom of multiplicativity with the octave and 17-semitone standard ($z_{critical}(U) = 1.96$)

Participant	Octave standard		17-semitone standard	
	$M_{1/3} - M_{1/2 \bullet 2/3}$ (Hz)	$z(U)$	$M_{1/3} - M_{1/2 \bullet 2/3}$ (Hz)	$z(U)$
Musically untrained participants				
AK	-38	3.09**	-38	1.05
CL	47	2.02*	65	2.47*
CW	16	1.31	29	2.20*
FK	13	1.60	-12	1.00
JB	30	0.98	46	1.41
JK	27	1.53	-11	0.59
JS	3	0.16	44	1.82†
MD	53	2.28*	0	0.01
MS	-9	0.48	-14	0.61
SD	0	0.02	5	0.51
WE	-12	1.29	-43	2.37*
Musically trained participants				
AR	2	1.18	5	0.91
AS	9	2.54*	-1	0.67
DD	23	1.29	12	0.47
DS	-17	0.70	-52	2.36*
JM	31	1.63	-20	0.77
MH	2	1.59	8	1.78†
SB	11	2.11*	-4	0.95
US	42	2.71*	36	2.20*
XS	-2	0.04	-19	1.05
YR	50	2.41*	50	1.69†

† $p < .1$; * $p < .05$

the adjustments made also satisfy the basic monotonicity property. This has been tested by comparing the pitch intervals adjusted to different fractions presented in the instructions (Eq. 3) and with different standard intervals (Eq. 4b). Whereas all participants produced monotonically increasing pitch intervals as a function of the numerals presented, only the musically trained participants were able to produce pitch intervals that reflected the size of the standard interval.

Virtually, most musically untrained participants produced indistinguishable frequencies when asked to adjust either a 12-semitone interval or a 17-semitone standard to the fraction p . This suggests that they did not attend to the lower limit of the standard interval, but made their adjustments solely with regard to the upper tone (see Fig. 2). Consequently, only two musically untrained listeners satisfy monotonicity with regard to the size of the standard interval (violating Eq. 4b). Eleven listeners (including two musically trained participants), however, were unable to produce a fraction that varied as a function of the standard interval presented. Formally, this describes a violation of a basic property of a (no less than ordinal) scale, and it suggests

that direct pitch productions may be valid only within an octave interval for these participants. This is a paradox that requires further investigation. Particularly, their discounting the lower reference of the standard interval might indicate that these participants are using a constant frequency reference, e.g., one that is defined by an octave below the upper boundary of the standard interval (which varied between trials and thus potentially attracted more attention). Musical training appears to be required in order to process a reference that lies in a different octave.

Most of the musically trained participants actually satisfied monotonicity with regard to both the numerical fractions and the standard intervals presented, and their adjustments of pitch intervals seem to be representable on a ratio scale even for intervals larger than an octave. Interestingly, musically trained participants produced smaller pitch intervals on average than musically untrained participants did. This suggests that the former may have tended to produce frequency intervals that are in accordance with particular fractions of musical intervals (as defined by the number of semitones).

Altogether, these group differences indicate that musicians and non-musicians may have relied on different strategies to adjust pitch intervals. Whereas musically trained listeners might be able to count the number of semitones spanned by arbitrary musical intervals, non-musicians may have to rely on their ‘naïve’ sensations of pitch height. These standard-dependent monotonicity violations in non-musicians imply that ‘naïve’ sensations of pitch intervals may be valid on a ratio scale within but not beyond an octave. Acquired musical skills like semitone-counting could be required in order to produce meaningful pitch intervals that are larger than an octave. Deutsch (1969) argued that the perception of pitch intervals exceeding an octave (which is extremely rare in music) needs an additional process of transposing the interval to the same octave range. This transposition of pitch intervals may have been responsible for the different patterns of responses that have been obtained with the two standards in the present study. That is, it might be extremely difficult for musically untrained listeners to discriminate frequency intervals that exceed an octave.

Hysteresis and order effects

One limitation of the present study is that the pitch intervals were always presented in the same order with a lower tone being followed by a higher tone (as in all axiomatic studies known to us; for an example with loudness intervals, see Steingrimsson & Luce 2005). The data presented could thus be influenced by hysteresis effects (e.g., Stevens 1957). That is, fractionations of intervals (e.g., equisection) have been shown to depend on whether the interval is presented in an ascending or a descending order (with ascending intervals typically producing greater adjustments). For instance, bisection of ascending loudness intervals resulted in level adjustments that were about 5–8 dB higher than with descending loudness intervals (Garner and Hake 1951). Similar effects have been reported for other sensory continua (e.g., brightness or heaviness; see Stevens 1957). Stevens (1957) hypothesized that hysteresis should not occur in ‘metathetic’ continua (including pitch), but the empirical evidence is not entirely consistent. Whereas Cohen et al. (1954) found (on average) no hysteresis in pitch bisections, Greenwood (1997) reported higher equisection adjustments (1/4, 1/2, 3/4 of an interval from 400 to 7000 Hz) with ascending frequency intervals than with descending frequency intervals.

In the present data, potential hysteresis effects could be evaluated by comparing the adjusted frequencies with certain expected semitone intervals. By referring to the chromatic scale, a closer look at Fig. 2 (right ordinate) reveals that each average adjustment of a frequency inter-

val exceeded the expected number of semitones. This bias, however, was much stronger in musically untrained listeners, suggesting that pitch hysteresis can be minimized by musical experience. Moreover, if the bias was larger for smaller fractions (as for loudness equisections; see Stevens 1957), then successive adjustments should produce larger outcomes if a smaller fraction (e.g. 1/2) had been adjusted prior to a larger fraction (e.g., 2/3). In the present data, there is some indication of a trend in this direction.²

Taken together, the data suggest that the hysteresis effect with regard to pitch judgments is rather small in magnitude (if present at all), and that it should be approximately equal for different partitions made (compare Greenwood 1997, see Fig. 4). Consequently, it should not have affected axiom testing systematically in the frequency range investigated in the present study. Nevertheless, it might be an interesting issue for further research to study whether ratio-scale properties can be shown with descending pitch intervals, as well.

Order effects could also be discussed with regard to the fact that the standard interval was always presented prior to the comparison interval. Again, although it is conceivable that this may introduce a bias, the impact of that bias is likely to be the same in all adjustments made in the present experiment, and it would consequently not influence axiom testing.

Conclusions

To sum up, the present study shows that direct ratio productions of pitch intervals meet the commutativity property in most of the participants. This implies, that pitch intervals may be represented on an internal ratio scale. However, for a few musically trained participants, commutativity was violated within an octave, suggesting that these participants may not represent frequent musical intervals on a simple ratio scale (rather they might represent them by labels on the chromatic scale). Moreover, in contrast to what has been found in direct scaling of loudness, brightness, or area (Ellermeier and Faulhammer 2000; Zimmer 2005; Peißner 1999; Augustin and Maier 2008), most of the pitch adjustments were also in line with the multiplicativity property, irrespective of musical training. This demonstrates that most of the listeners used the numerals referring to fractions of pitch intervals just like scientific numbers, meaning that these numbers can be interpreted as values on a ratio scale without further transformation. That is, any adjustment represented by the triple $(x, \mathbf{p}, t - y)$ will correspond to a subjective pitch interval of $\psi(x - y) = p * \psi(t - y)$, implying that $p = \mathbf{p}$. That situation is far more favorable than merely being able to state that participants are operating on

²Thanks to one of the reviewers who pointed out this effect.

an unknown ratio scale; rather the number words used in the procedure can be taken as indicative of the actual ratio-scale values.

The present study also investigated monotonicity of pitch productions as a function of the standard - a property rarely tested explicitly in this kind of axiomatic study. Whereas musically trained listeners produced frequencies that varied as a function of the standard interval, this property was systematically violated by the musically untrained participants. They appeared to have difficulty distinguishing a standard interval spanning an octave from one exceeding it. It thus appears that, for fractionations of pitch intervals larger than an octave, some musical training is necessary in order to obtain responses that are in accordance with the monotonicity axiom.

Finally, the data show that the overall variability and the precision of pitch adjustments differed considerably between participants. This may imply that the task difficulty varied dramatically between participants, for instance as a function of the participants' prior experience with or knowledge about musical intervals. Moreover, 'the pitch continuum probably means different things to different people' (Stevens 1957, p. 161). Nevertheless, for some individuals, direct magnitude productions of pitch intervals were shown to be valid with respect to a behavioral axiomatization of ratio scaling.

Author notes Portions of the data were presented at the 28th Annual Meeting of the International Society of Psychophysics in Ottawa, Canada. The authors are indebted to Martha Teghtsoonian and an anonymous reviewer who made valuable comments and suggestions on a previous version of the manuscript.

References

- ANSI. (1973). *American national psychoacoustical terminology - S3.2*. New York: American National Standards Institute (ANSI).
- Augustin, T. (2006). Stevens' direct scaling methods and the uniqueness problem. *Psychometrika*, *71*, 469–481.
- Augustin, T., & Maier, K. (2008). Empirical evaluation of the axioms of multiplicativity, commutativity, and monotonicity in ratio production of area. *Acta Psychologica*, *129*, 208–216.
- Beck, J., & Shaw, W.A. (1961). The scaling of pitch by the method of magnitude estimation. *American Journal of Psychology*, *74*, 242–251.
- Beck, J. (1962). Magnitude estimations of pitch. *Journal of the Acoustical Society of America*, *34*, 92–98.
- Brainard, D.H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 443–446.
- Cohen, J., Hansel, C.E.M., Sylvester, J.D. (1954). Interdependence of temporal and auditory judgments. *Nature*, *174*, 642.
- Deutsch, D. (1969). Music recognition. *Psychological Review*, *76*, 300–307.
- Ellermeier, W., & Faulhammer, G. (2000). Empirical evaluation of axioms fundamental to Stevens's ratio-scaling approach: I. Loudness production. *Perception and Psychophysics*, *62*, 1505–1511.
- 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., & Hake, H.W. (1951). The amount of information in absolute judgments. *Psychological Review*, *58*, 446–459.
- Gescheider, G.A. (1997). *Psychophysics: The fundamentals*, 3rd edn. New Jersey: Lawrence Erlbaum.
- Greenwood, D.D. (1997). The Mel Scale's disqualifying bias and a consistency of pitch-difference equisections in 1956 with equal cochlear distances and equal frequency ratios. *Hearing Research*, *103*, 199–224.
- Idson, W.L., & Massaro, D.A. (1978). Bidimensional model of pitch in the recognition of melodies. *Perception and Psychophysics*, *24*, 551–565.
- ISO (1998). *Acoustics - Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supraaural earphones (ISO 389-1)*. Geneva: International Organization for Standardization (ISO).
- Krumhansl, C.L., & Shepard, R.N. (1979). Quantification of the hierarchy of tonal functions within a diatonic context. *Journal of Experimental Psychology: Human Perception and Performance*, *5*, 579–594.
- Luce, R.D. (2002). A psychophysical theory of intensity proportions, joint presentations, and matches. *Psychological Review*, *109*, 520–532.
- Luce, R.D. (2004). Symmetric and asymmetric matching of joint presentations. *Psychological Review*, *111*, 446–454.
- McKenna, F.P. (1985). Another look at the "new psychophysics". *British Journal of Psychology*, *76*, 109–129.
- Merkel, J. (1888). Die Anhängigkeit zwischen Reiz und Empfindung. *Philosophische Studien*, *4*, 541–594.
- Narens, L. (1996). A theory of ratio magnitude estimation. *Journal of Mathematical Psychology*, *40*, 109–129.
- Narens, L. (2002). The irony of measurement by subjective estimations. *Journal of Mathematical Psychology*, *46*, 769–788.
- Painton, S.W., Cullinan, W.L., Mencke, E.O. (1977). Individual pitch functions and pitch-duration cross-dimensional matching. *Perception and Psychophysics*, *21*, 469–476.
- Parker, S., & Schneider, B. (1974). Nonmetric scaling of loudness and pitch using similarity and difference estimates. *Perception and Psychophysics*, *15*, 238–242.
- Peißner, M. (1999). *Experimente zur direkten Skalierbarkeit von gesehenen Helligkeiten* (Unpublished master's thesis). Universität Regensburg.
- Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
- Plack, C.J., & Oxenham, A.J. (2005). Psychophysics of pitch. In A. N. Popper & R. Fay (Eds.), *Pitch: Neural coding and perception* (pp. 7–55). New York: Springer.
- Plomp, R. (1967). Pitch of complex tones. *Journal of the Acoustical Society of America*, *41*, 1526–1533.
- Rasch, R., & Plomp, R. (1999). *The psychology of music*. In D. Deutsch (Ed.), (2nd ed., pp. 89–112). New York: Academic Press.
- Richardson, L.F. (1929). Imagery, conation, and cerebral conductance. *Journal of General Psychology*, *2*, 324–352.
- Russo, F.A., & Thompson, W.F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin and Review*, *12*, 1068–1075.
- Shepard, R.N. (1964). Circularity of judgments of relative pitch. *Journal of the Acoustical Society of America*, *36*, 2346–2353.

- Shepard, R.N. (1981). Psychological relations and psychophysical scales: On the status of “direct” psychological measurement. *Journal of Mathematical Psychology*, *24*, 21–57.
- Siegel, J.A. (1964). A replication of the mel scale of pitch. *American Journal of Psychology*, *78*, 615–620.
- Steingrimsson, R., & Luce, D.R. (2007). Empirical evaluation of a model of global psychophysical judgments: IV. Forms for the weighting function. *Journal of Mathematical Psychology*, *51*, 29–44.
- Steingrimsson, R., & Luce, R.D. (2005). Evaluating a model of global psychophysical judgments - II: Behavioral properties linking summations and productions. *Journal of Mathematical Psychology*, *49*, 308–319.
- Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, *103*, 677–680.
- Stevens, S.S. (1956). The direct estimation of sensory magnitudes - loudness. *American Journal of Psychology*, *69*, 1–25.
- Stevens, S.S. (1957). On the psychophysical law. *Psychological Review*, *65*, 153–181.
- Stevens, S.S. (1971). Issues in psychophysical measurement. *Psychological Review*, *78*, 426–450.
- Stevens, S.S. (1975). *Psychophysics: Introduction to its perceptual, neural and social prospects*. 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., & Volkman, J. (1940). The relation of pitch to frequency: A revised scale. *American Journal of Psychology*, *53*, 329–353.
- Stevens, S.S., Volkman, J., Newman, E.B. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.
- Young, R.W. (1939). Terminology for logarithmic frequency units. *Journal of the Acoustical Society of America*, *11*, 134–139.
- Zimmer, K. (2005). Examining the validity of numerical ratios in loudness fractionation. *Perception and Psychophysics*, *67*, 569–579.