Music Perception
Winter 2002, Vol. 19, No. 2, 223–248

Asymmetries in the Discrimination of Musical Intervals: Going Out-of-Tune Is More Noticeable Than Going In-Tune

E. GLENN SCHELLENBERG
University of Toronto

Listeners were tested on their ability to discriminate "standard" and "comparison" pure-tone musical intervals that differed in size by 20 cents (1/5 of an equal-tempered semitone). Some of the intervals were prototypic, equal-tempered perfect fifths (exactly 7 semitones, or 700 cents). Others were mistuned to various degrees (660, 680, 720, or 740 cents). The intervals were melodic (sequential) in Experiments 1 and 2 and harmonic (simultaneous) in Experiment 3. Performance was neither enhanced nor impaired in comparisons that included the prototype. In other words, no "perceptual magnet" or "perceptual anchor" effects were observed. Nonetheless, performance was markedly asymmetric. Regardless of listeners' musical expertise, discrimination was superior when the standard interval was more accurately tuned than the comparison interval (e.g., 700-cent standard, 680-cent comparison), compared with when the comparison was more accurately tuned than the standard (e.g., 680-cent standard, 700-cent comparison).

Received February 28, 2001, accepted July 25, 2001.

Relative rather than *absolute* pitch processing is the norm for human listeners, except for the small minority with "perfect" or "absolute" pitch (see Takeuchi & Hulse, 1993; Ward, 1999). Thus, the identity of a melody is determined by the pitch and temporal relations between successive tones rather than the absolute values of the tones. For example, *Yankee Doodle* is recognizable regardless of whether it is sung slowly or quickly, or by a man or a woman (i.e., different pitch register and vocal timbre), provided the relations between consecutive tones conform to those of the song. The present investigation explored listeners' ability to discriminate tone patterns with different pitch relations. The goal was to examine how

Address correspondence to Glenn Schellenberg, Department of Psychology, University of Toronto at Mississauga, Mississauga, ON, Canada L5L 1C6. (e-mail: g.schellenberg@utoronto.ca)

ISSN: 0730-7829. Send requests for permission to reprint to Rights and Permissions, University of California Press, 2000 Center St., Ste. 303, Berkeley, CA 94704-1223.

the presence of a category *prototype* (i.e., an ideal instance of a category) influences discrimination abilities. The experiments tested three hypotheses about prototypes and their discriminability from other members of the same category: (1) Prototypes function as *perceptual magnets* (e.g., Kuhl, 1991), which makes them confusable with other instances and therefore difficult to discriminate, (2) Prototypes function as *perceptual anchors* (e.g., Acker, Pastore, & Hall, 1995), which makes them maximally distinct from other instances and therefore easy to discriminate, and (3) Prototypes generate *asymmetries* in discrimination (e.g., Schellenberg & Trehub, 1996b), such that performance is better when the prototypic instance is the first (i.e., the *standard*) rather than the second (i.e., the *comparison*) of two stimuli.

A category can be defined as a class of stimuli "based on common features or similarity to a prototype" (Sternberg, 1999, p. 255). *Vegetable* is a category that includes carrots, peas, broccoli, artichokes, and so on. *Peas* represent a lower-level category that includes different types of peas (snow peas, snap peas, sweet peas, etc.). Rosch questioned the idea that category membership is an all-or-none affair (Mervis & Rosch, 1981; Rosch & Mervis, 1975; Rosch, Mervis, Gray, Johnson, & Boyes-Braen, 1976). She showed that some instances of a category are better representatives than other instances, and she considered the best instances to be prototypes (Rosch, 1973, 1975b). For example, some instances of the color red (e.g., stop-sign red) are more prototypic than other instances of red (e.g., wine, crimson), and some types of cars (e.g., sedans) are more prototypic than others (e.g., Formula 1 race cars). In other words, instances from the same category vary in their degree of prototypicality. Rosch's ideas about prototypes proved to be important for many different areas of psychology (perception, cognition, social psychology) because the prototypic status of a stimulus affects how it is perceived, learned, remembered, and produced (see Mervis & Rosch, 1981).

Prototypes in Western music include particular tones (doh), intervals (octaves, perfect fifths), chords (major chords), chord progressions (V-I), and large-scale musical structures (sonata form, pop-song structures). The present investigation examined a prototypic interval: the perfect fifth. Many musical prototypes are culturally specific, acquiring their special status through repeated exposure to a particular style of music. Others appear to be universal, reflecting contributions from biology (Trehub, 2000). For example, intervals with small-integer frequency ratios (e.g., justly tuned octaves—2:1, perfect fifths—3:2, or perfect fourths—4:3), or intervals that closely approximate such ratios (as in equal temperament), are structurally important across musical cultures, including those of India, China, Africa, and Western Europe (Burns, 1999; Meyer, 1956; Schellenberg & Trehub, 1994b; Trehub, Schellenberg, & Hill, 1997). One explanation for their special status involves the harmonic series and sensory consonance. Complex tones with fundamental frequencies standing in a small-integer ratio are relatively consonant (or pleasant sounding). The smaller the integers, the greater the number of identical harmonics and the lesser the likelihood of overlapping critical bands (from adjacent harmonics) that generate dissonance (Helmholtz, 1885/1954; Kameoka & Kuriyagawa, 1969; Plomp & Levelt, 1965).

Sensitivity to sensory consonance is assumed to be universal and independent of learning (e.g., Bregman, 1990; Burns, 1999). Indeed, rhesus monkeys demonstrate enhanced memory for melodies with consonant intervals compared with melodies with dissonant intervals; they are also more likely to notice similarities between melodies transposed by consonant than by dissonant intervals (Wright, Rivera, Hulse, Shyan, & Neiworth, 2000). Moreover, infants (Schellenberg & Trainor, 1996) and songbirds (Hulse, Bernard, & Braaten, 1995) show perceptual grouping of intervals on the basis of sensory consonance. Infants also prefer consonant over dissonant music (Zentner & Kagan, 1998; Trainor & Heinmiller, 1998). Nonetheless, the prototypic status of small-integer frequency ratios generalizes to pairs of pure tones (i.e., no harmonics) presented simultaneously or sequentially, when there are no cues from sensory consonance or dissonance. For example, infants, children, and adults find pure-tone intervals with such ratios easier to process and remember than their counterparts with large-integer ratios (Schellenberg & Trehub, 1994a, 1996a, 1996b; Trainor, 1997).

The prototypic status of a stimulus is known to affect its perceived similarity to other stimuli, and, hence, its discriminability. Rosch and Mervis (1975) contend that a prototypic stimulus is maximally *distinct* from stimuli from other categories, but maximally *similar* to stimuli from the same category. They demonstrated that the most prototypic members of a category have the largest number of attributes in common with other members of the same category, but the fewest attributes in common with members of other categories. It follows, then, that the presence of a prototype should enhance discrimination of stimuli from different categories, but impair discrimination of stimuli from the same category.

The idea of poor within-category discrimination involving prototypes was extended to speech perception by proponents of the "perceptual magnet effect" (Iverson & Kuhl, 1995, 1996, 2000; Kuhl, 1991). Listeners perceive some tokens from a particular phonemic category (e.g., the vowel /i/) to be more representative of the category than other tokens from the same category (Aaltonen, Eerola, Hellström, Uusipaikka, & Lang, 1997; Frieda, Walley, Flege, & Sloane, 1999, 2000; Iverson & Kuhl, 1995, 1996, 2000; Kuhl, 1991; Lively & Pisoni, 1997; Lotto, Kluender, & Holt, 1998; Miller & Volaitis, 1989; Samuel, 1982; Sussman & Gekas, 1997; Sussman

& Lauckner-Morano, 1995; Volaitis & Miller, 1992). When listeners are asked to discriminate different tokens from the same phonemic category, performance is poorer for discrimination of a prototypic phoneme (e.g., the highest rated /i/) from slightly altered tokens than for discrimination of a nonprototypic phoneme (e.g., an /i/ with a lower rating) from identically altered tokens (Iverson & Kuhl, 1995, 1996, 2000; Kuhl, 1991). Kuhl (1991) argues that the prototype acts as a “perceptual magnet,” effectively shrinking the distance in similarity space between it and other instances of the category. Such distortions are adaptive because they focus attention away from meaningless differences toward distinctions that are relevant in a specific language. Indeed, the perceptual magnet effect appears to become operative after limited exposure to language in infancy (Grieser & Kuhl, 1989; Kuhl, Williams, Lacerda, Stevens, & Lindbloom, 1992; Polka & Werker, 1994).

Other researchers have presented evidence that category prototypes function as *perceptual anchors*, which means that the ability to discriminate two tokens from the same category is enhanced when one of the tokens is a prototype. Although the anchor hypothesis is the exact opposite of the magnet hypothesis, it provides a good explanation of response patterns observed in some music-perception experiments. For example, listeners find it easier to detect the mistuning of one tone in a major chord (a musical prototype) than an identical mistuning in a nonprototypic chord (Acker & Pastore, 1996; Acker et al., 1995; McFadden & Calloway, 1999). McFadden and Calloway propose that such response patterns are indicative of enhanced discrimination for “commonly encountered” stimuli (compared to stimuli that are encountered less frequently), which may generalize widely across the auditory domain. In support of this view, enhanced discrimination of commonly encountered stimuli extends to rhythmic patterns and to speech sounds (McFadden & Calloway, 1999).

The presence of a prototype may also produce *asymmetries* in the perceived similarity of stimuli (Rosch, 1975a). Prototypes may function as category reference points, such that other members of the category are judged and classified on the basis of their degree of similarity to the prototype. Rosch claims that “stimuli slightly different from reference stimuli are more easily assimilated to and, thus, judged metaphorically ‘closer to’ the reference stimuli than vice versa” (1975a, p. 533). For example, her participants agreed that the nonprototypic number 101 is “essentially” 100, but they did not agree that prototypic 100 is essentially 101. Similar asymmetries were identified in a spatial-distance task that required participants to position a comparison stimulus in relation to a fixed standard stimulus. The distance between stimuli was smaller when the fixed standard was a prototype and the comparison was less prototypic than when the standard and comparison were reversed (Rosch, 1975a).

Asymmetries in performance have also been identified in interval-discrimination experiments with infant, child, and adult listeners (Schellenberg & Trehub, 1994a, 1996a; Trainor, 1997). The "category" in this case was the set of Western musical intervals. The experiments included prototypic musical intervals such as octaves, perfect fifths, and perfect fourths, as well as nonprototypic intervals such as tritones, major sevenths, and minor ninths. In each case, listeners were required to detect a 1-semitone change in interval size without any cues from sensory consonance or dissonance (i.e., the intervals consisted of two pure tones). In some conditions, standard intervals (presented first) and comparison intervals (presented second) were simply the reverse of other conditions, yet performance was markedly asymmetric. For example, discriminating a perfect-fifth standard interval (7 semitones) from a tritone comparison interval (6 semitones) was relatively easy for listeners of all ages. By contrast, discriminating a tritone standard from a perfect-fifth comparison was much more difficult.

We can also consider different intervals in Western music (e.g., octaves, perfect fifths) to represent distinct lower-level categories, just as *apple*—a member of the fruit category—also represents a distinct lower-level category (consisting of Macintosh, Granny Smith, etc.). Particular intervallic categories have many instances, with some instances being more prototypic than other instances from the same category. In musical performance, intervals vary in size for instruments that do not have fixed pitches (e.g., the human singing voice, stringed instruments such as the violin or cello) as a consequence of expressive intentions, limitations in motor control, and perceptual limitations. One can consider exactly tuned instances (700 cents, an equal-tempered fifth) to be prototypic within the perfect-fifth category, slightly mistuned instances (e.g., 690 cents) to be less prototypic, and instances with greater mistunings (e.g., 680 cents) even less so. In fact, the distinction between prototypic and nonprototypic instances within an interval category parallels the distinction between prototypic and nonprototypic instances within a phonemic category. The parallelism implies that prototypic musical intervals would function as perceptual magnets. As with speech, many within-category variations in interval size go unnoticed. Rather, they are assimilated to listeners' internalized interval categories. Trained musicians have *explicit* knowledge of such categories, but untrained listeners have only *implicit* knowledge (see Smith, 1997). For example, many untrained listeners would notice when the interval between the first and second words in *Twinkle Twinkle Little Star* is sung incorrectly, yet they cannot specify the problem—substantial deviation from a perfect fifth.

The present set of experiments tested listeners' ability to discriminate different instances from the same intervallic category. The goal was to determine whether prototypic intervals function as perceptual magnets or perceptual anchors, or whether the asymmetries reported in discriminating instances from different intervallic categories would extend to smaller changes in size between intervals from the same category. Listeners were required to detect a small change (20 cents) in interval size, which is close to the just-noticeable-difference (JND) in interval-discrimination tasks (Burns, 1999). Accordingly, the task was expected to be challenging. As illustrated in Figure 1, the stimuli included intervals of 660, 680, 700, 720, and 740 cents (i.e., an equal-tempered perfect fifth plus fifths mistuned "flat" or "sharp" by 20 or 40 cents). Note that "sharp" in this context meant that, given the pitch of an interval's low tone, the pitch of the high tone was higher than it would be in equal temperament. Conversely, "flat" meant that the high tone was tuned lower than in equal temperament.

All of the stimulus intervals were from the perfect-fifth category—6.5 to 7.5 semitones (650-750 cents)—assuming that the boundaries between adjacent smaller (tritone, 6 semitones) and larger (minor sixth, 8 semitones) categories fall at the midpoint. Support for this assumption is provided by interval identification functions from musically trained listeners, which are steep and cross very close to the midpoint between categories (Burns & Campbell, 1994; Burns & Ward, 1978; Rakowski, 1990; Siegel & Siegel, 1977a, 1977b; Smith, Kemler Nelson, Grohskopf, & Appleton, 1994). Indeed, these "identification functions are characterized by sharp category boundaries, high test-retest reliability, and a resistance to contextual effects" (Burns, 1999, p. 222). By contrast, untrained listeners have no explicit knowledge of musical categories and are typically excluded from interval labeling tasks (but see Smith et al., 1994). Hence, differences between musically trained and untrained listeners were examined because it was unclear whether a category prototype—or "best" instance of a category—would affect the discrimination abilities of untrained listeners.

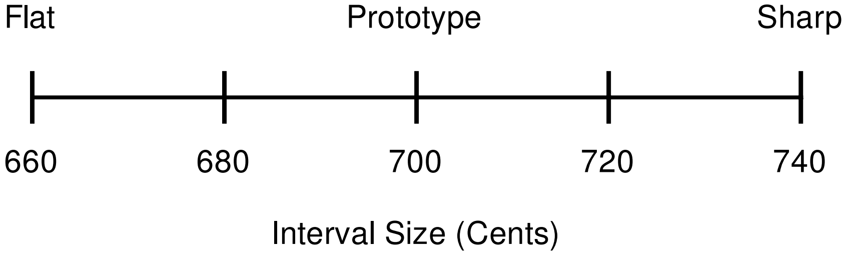

Fig. 1. Schematic drawing of the five stimulus intervals: the prototypic, equal-tempered perfect fifth (700 cents), two intervals tuned flat (680 and 660 cents), and two tuned sharp (720 and 740 cents). Listeners were tested on their ability to discriminate pairs of adjacent intervals (i.e., a change in size of 20 cents). Half of the testing conditions included the prototypic interval; the other half did not. In half of the testing conditions, the standard interval was more accurately tuned than the comparison interval; in the other half, the comparison was more accurately tuned than the standard.

The design of the experiments is outlined in Table 1. Each testing condition included a "standard" interval and a "comparison" interval. One manipulation involved the presence or absence of a prototype. Half of the testing conditions included the prototypic, equal-tempered perfect fifth (as standard or comparison); the other half did not. If a prototypic interval functions as a perceptual magnet—as prototypic phonemes do—listeners should have difficulty discriminating it from other instances from the perfect-fifth category. Thus, performance should be *worse* in conditions that include the exactly tuned interval compared to other conditions. By contrast, if the perfect-fifth prototype functions as a perceptual anchor—as other musical prototypes do—performance should be *better* in conditions that include the prototype than in other conditions. Note that neither the magnet nor the anchor perspective considers order of presentation, which formed the basis of a second manipulation (orthogonal to the first). Specifically, the standard interval was more in-tune (more prototypic) than the comparison interval in half of the conditions; the standard and comparison were reversed for the other half. If the prototype generates asymmetries in discrimination performance in within-category contexts—such as those observed in between-category tests of interval discrimination—performance should be relatively good when the standard interval is more in-tune than the comparison interval, but relatively poor when the comparison is more in-tune than the standard. We had no predictions about whether any of the hypothesized effects (magnet, anchor, or asymmetry) would be a *linear* function of deviation in interval size (in cents) from the equal-tempered fifth.

In Experiment 1, listeners (recruited without regard to musical training) were tested with sequential intervals in a between-subjects design. Potential differences in performance due to musical training were held constant by statistical means. In Experiment 2, musically trained listeners were tested

TABLE 1
Design of Experiments 1, 2, and 3

Condition	Interval Size in Cents (Standard-Comparison)	Prototype	Interval Tuned More Accurately
Flat	700-680	Present	Standard
(Experiments 1, 2, and 3)	680-700	Present	Comparison
	680-660	Absent	Standard
	660-680	Absent	Comparison
Sharp	700-720	Present	Standard
(Experiments 2 and 3)	720-700	Present	Comparison
	720-740	Absent	Standard
	740-720	Absent	Comparison

with sequential intervals in a within-subjects design. In Experiment 3, musically trained and untrained listeners were tested with simultaneous intervals in a within-subjects design. The use of pure tones (sine waves) across experiments ensured that the discrimination tasks could not be accomplished by attending to differences in sensory consonance or dissonance.

Experiment 1

Pilot testing revealed that conventional same/different tasks for testing discrimination of sequential pure-tone intervals were excessively difficult, even for listeners with many years of musical training and familiarity with music-perception experiments. Fortunately, the go/no-go method, which reduces memory demands, has been used successfully for testing adults in a variety of auditory discrimination tasks (Kuhl, 1991; Lynch & Eilers, 1992; Lynch, Eilers, Oller, & Urbano, 1990; Lynch & Steffens, 1994; Schellenberg & Trehub, 1994a; Trainor, 1997; Trainor & Trehub, 1992, 1993a, 1993b). The procedure, developed originally to test infants' perception of speech (Eilers, Wilson, & Moore, 1977), is a particularly sensitive measure of discrimination abilities.

METHOD

Participants

The listeners were 120 members of the university community who received partial course credit or token remuneration for their participation. They were recruited without regard to musical training, which varied widely. They were asked about the number of years they had taken music lessons (mean = 2.84, *SD* = 3.93, median = 1.50) and the number of years they had played music on a regular basis (mean = 3.27, *SD* = 4.33, median = 2.00).

Apparatus

Tones were sine waves generated with SoundEdit 16 software installed on a Macintosh PowerPC 7100/66AV computer. Tone sequences were stored as 16-bit monaural files (sampling rate = 22.05 kHz) on the hard disk of the same computer. Stimuli were presented over headphones (SONY CD 550) while the listeners sat in a sound-attenuating booth (Eckel Industries). Listeners used a button-box connected to the computer to make their responses. A window in the booth allowed them to see the computer monitor.

Stimuli

Each condition had a *standard* tone sequence and a *comparison* sequence. All sequences contained five contiguous pure tones, with two alternating tones in a low-high-low-high-low pattern (as in Schellenberg & Trehub, 1994a, 1996a). Each tone had a duration of 400 ms with 10-ms linear onsets and offsets, such that the total duration of sequences was 2 s. The size of the interval between the low and high tones of the sequences varied across the four testing conditions (see Table 1, "flat" intervals). In one condition (700-680), the standard interval was 700 cents and the comparison interval was 680 cents. In another condition (680-700), the standard and comparison intervals were reversed. In the third and fourth conditions, the standard was 680 cents and the comparison was 660 cents (680-660), and vice versa (660-680).

In each condition, listeners heard the standard interval presented repeatedly in *transposition*, such that absolute frequencies changed from presentation to presentation but the frequency ratio between the high and low tones remained constant. Transpositions between successive presentations were additive increases or decreases in frequency of 60 Hz presented in a random-walk pattern, with the low tone of sequences having a frequency of 310, 370, 430, 490, or 550 Hz. This constant difference in frequency (rather than frequency ratio) meant that transposed stimulus tones did not belong to any equal-tempered scale and listeners had to concentrate solely on the interval between high and low tones to succeed at the task. On average, these 60-Hz transpositions represented shifts in pitch of 2.48 semitones. Consecutive presentations of the sequence were separated by 1200 ms.

Procedure

Thirty listeners were assigned at random to each of the four testing conditions. The entire procedure was computer-controlled using a custom-made program developed with PsyScope 1.1 software (Cohen, MacWhinney, Flatt, & Provost, 1993).

Listeners were asked to indicate when they heard a change in the repeating standard interval. In each condition, there were 30 *change* trials, during which the comparison interval was substituted for the standard. As with the repeating standard interval, the comparison interval was always presented at a different pitch level than the preceding and subsequent standard intervals. Listeners pressed a button on the button-box to signal when they heard a change. There were also 30 *no-change* trials, during which the computer monitored false alarms (i.e., indicating a change when none was present). No-change trials, which involved another presentation of the repeating (and transposed) standard interval, provided an estimate of incorrect responding and response bias. Listeners' responses were recorded during a 2.8-s window that began at the onset of the first (potentially changed) high tone of trial sequences (change or no-change) and ended at the onset of the subsequent sequence. Consecutive trials (change or no-change) were always separated by 2, 3, or 4 repeating standard sequences, with the number of intervening sequences selected randomly. The word "correct" was displayed on the computer screen for 0.5 s as feedback for responding on change trials. There was no feedback for failing to notice a change on change trials, or for signaling a change when none was present. In the latter case, however, the lack of positive feedback indicated an incorrect response.

The test phase was preceded by a brief training phase designed to familiarize listeners with the procedure and equipment, and with the idea that exact transpositions did *not* constitute a "change" in the present experiment. The training session was identical to the test session except that: (1) the to-be-detected change was more noticeable (the high tones were raised by an octave, or 12 semitones), and (2) there were only six trials, all of which were change trials. Before the training session, listeners were told about musical transpositions and how an exact transposition is considered "the same" in a musical sense. Specifically, they were asked to consider a man and a woman singing the same song, and how the woman's rendition would be higher in pitch. They also heard a simple melody (e.g., *Happy Birthday*) played at several different pitch levels on a keyboard.

RESULTS AND DISCUSSION

A discrimination (d') score was derived separately for each listener using the yes/no formula from signal-detection theory (Macmillan & Creelman, 1991, p. 9). Before calculating d' scores, hit rates (the proportion of correctly identified change trials) and false-alarm rates (the proportion of incorrectly identified no-change trials) were adjusted to avoid the possibility of indeterminate *d′* scores that can arise from perfect responding. Specifically, 0.5 was added to the numerator (the number of hits or false alarms) and 1 to the denominator (the number of change or no-change trials), following Thorpe, Trehub, Morrongiello, and Bull (1988). This transformation has a minor effect on *d′* scores and no effect on their rank ordering. The maximum possible transformed *d′* was 4.28.

Table 2 provides descriptive statistics separately for each condition. As expected, discrimination was relatively poor in all conditions, with mean *d′* scores under 0.5. Nonetheless, one-sample *t* tests comparing performance with chance levels ($d' = 0$) revealed above-chance levels of responding in the 700-680 condition, $t(29) = 5.12$, $p < .001$, the 680-660 condition, $t(29) = 3.67$, $p = .001$, and the 660-680 condition, $t(29) = 2.86$, $p = .008$, but not in the 680-700 condition.

The effect of musical expertise on performance was tested by examining correlations between *d′* scores and (1) years of music lessons, and (2) years of playing music on a regular basis. A small but significant positive association was uncovered in the latter case, $r = .162$, $p = .034$ (one-tailed).

Differences between conditions were examined with an analysis of covariance that included a covariate (number of years playing music regularly) and two independent variables: presence of the prototype (present or absent) and direction of mistuning (standard interval more in-tune than comparison interval or vice versa). The main effect of the prototype manipulation was not significant and did not interact with the mistuning variable. In other words, the presence of the exactly tuned perfect fifth did not affect performance. By contrast, the main effect of the direction-of-mistuning manipulation was reliable, $F(1, 115) = 4.27$, $p = .041$ (see Figure 2). Performance was better when the standard interval was more in-tune than the comparison interval (mean = 0.38) compared with the reverse situation (mean = 0.20).

TABLE 2

Descriptive Statistics From Experiment 1 (Melodic Intervals, Listeners Recruited Without Regard to Musical Background)

Condition	N	Mean	SD	Adjusted Mean
700-680	30	.359	.384	.380
680-700	30	.085	.457	.093
680-660	30	.404	.604	.394
660-680	30	.309	.593	.291

Adjusted means have individual differences in musical expertise partialled out.

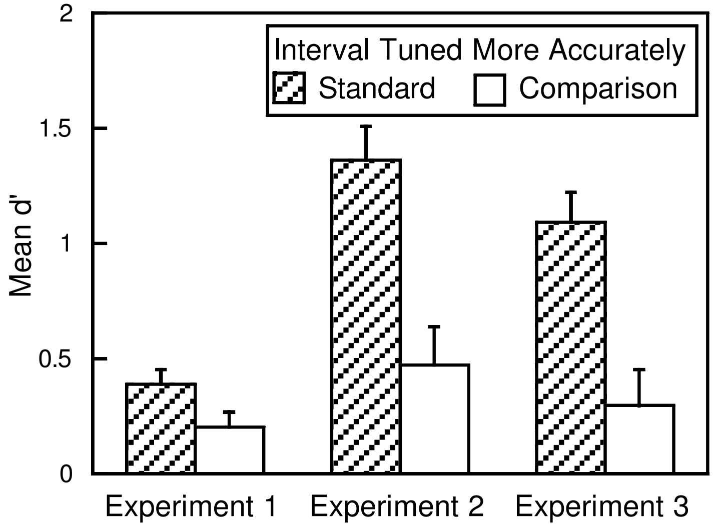

Fig. 2. Performance asymmetries in Experiments 1, 2, and 3 as a function of the direction-of-mistuning manipulation. For each experiment, the stimuli to be discriminated were the same for both levels of the manipulation—the standard and comparison intervals were simply reversed. Performance was superior when the standard interval was more accurately tuned than the comparison interval, compared with when the comparison interval was more accurately tuned than the standard. In Experiment 1, listeners were recruited without regard to musical training and tested with melodic intervals in a between-subjects design. In Experiment 2, musically trained listeners were tested with melodic intervals in a within-subjects design. In Experiment 3, musically trained and untrained listeners were tested with harmonic intervals in a within-subjects design. Error bars represent standard errors.

In sum, the data were consistent with the performance asymmetries observed in between-category tests of interval discrimination (Schellenberg & Trehub, 1994a, 1996a; Trainor, 1997). Performance was better when the standard interval was more prototypic (tuned more accurately) than the comparison interval compared with the reverse situation. There was no evidence for poorer or superior discrimination abilities in conditions that included the prototype, and, consequently, no evidence that the prototypic perfect fifth functions as either a perceptual magnet or a perceptual anchor.

Experiment 2

The purpose of the present experiment was to extend the findings of Experiment 1 by using a different design, a different group of listeners, and a larger set of intervals. Instead of the large-sample between-subjects design of Experiment 1, a small-sample within-subjects design was used. Above-chance levels of performance in Experiment 1 plus a positive association between performance and musical expertise implied that a repeated-measures design would be appropriate for musically trained listeners. Hence, the sample was restricted to listeners with many years of formal training in music or playing music regularly. Half of the listeners were tested with the same intervals used in Experiment 1 (700, 680, and 660 cents); the other half were tested with a new set of intervals equal to or larger than an equal-tempered perfect fifth (700, 720, and 740 cents).

METHOD

Participants

The participants were 16 members of the university community who had taken music lessons or played music regularly for at least 8 years. Participants were tested on two different days within a period of two weeks. They received partial course credit or token remuneration for participating.

Apparatus

The apparatus used was identical to that used in Experiment 1.

Stimuli

For half of the listeners, the stimuli were identical to those used in Experiment 1. For the other half, a set of "sharp" intervals (compared with the "flat" intervals used in Experiment 1; see Table 1) was created. These sharp intervals were 700, 720, or 740 cents in size.

Procedure

The procedure was identical to that of Experiment 1, except that each listener was tested in four experimental conditions rather than one. Half of the listeners were tested in all four conditions from Experiment 1: 700-680, 680-700, 680-660, and 660-680. The other half were tested in four analogous conditions that incorporated sharp rather than flat mistunings: 700-720, 720-700, 720-740, and 740-720 (Table 1). Listeners were tested in two conditions on the first testing day and in another two on the second testing day. In order to eliminate possible effects of testing order, four different orders were formed separately for sharp and flat intervals using a digram-balanced Latin square. Specifically, the order of the testing conditions was randomized for an initial listener. The remaining three orders were formed such that each condition appeared in a different serial position in each order, and a particular condition was directly preceded and followed by any other condition in only one order. Two listeners were assigned to each of the four testing orders in the sharp conditions, and to each of the four orders in the flat conditions.

RESULTS AND DISCUSSION

A *d´* score was calculated separately for each listener in each experimental condition, as in Experiment 1. Means and standard deviations are provided in Table 3. Comparisons with chance levels of responding revealed a clear and interpretable pattern. Performance exceeded chance levels in the four conditions in which the standard interval was more accurately tuned

TABLE 3
Descriptive Statistics From Experiment 2 (Melodic Intervals, Musically Trained Listeners)

Flat Conditions				Sharp Conditions			
Condition	N	Mean	SD	Condition	N	Mean	SD
700-680	8	1.261	0.727	700-720	8	0.919	0.502
680-700	8	0.715	0.946	720-700	8	0.829	1.134
680-660	8	1.788	1.058	720-740	8	1.461	0.818
660-680	8	0.452	0.934	740-720	8	-0.123	0.508

than the comparison interval [700-680: $t(7) = 4.91$, $p = .002$; 680-660: $t(7) = 4.78$, $p = .002$; 700-720: $t(7) = 5.17$, $p = .001$; and 720-740: $t(7) = 5.05$, $p = .001$], but not in the other four conditions.

A 2 × 2 × 2 mixed-design analysis of variance (ANOVA) with two within-subjects factors (presence of the prototype, direction of mistuning) and one between-subjects factor (sharp or flat intervals) uncovered two significant effects. The strongest was the main effect of direction of mistuning, $F(1, 14) = 21.16$, $p < .001$ (see Figure 2), which confirmed that asymmetries in performance generalized to the present experiment. Performance was better when the standard interval was more accurately tuned than the comparison interval (mean = 1.36) compared with the reverse situation (mean = 0.47). This effect was qualified, however, by an interaction between direction of mistuning and presence of the prototype, $F(1, 14) = 12.58$, $p = .003$ (see Figure 3). When the standard interval was more in-tune than the comparison interval, performance was superior when the prototype was absent (680-660, 720-740; mean = 1.62) rather than present (700-680, 700-720; mean = 1.09), $F(1, 14) = 8.41$, $p = .012$. When the comparison interval was more in-tune than the standard, performance was superior when the prototype was present (680-700, 720-700; mean = .77) rather than absent (660-680, 740-720; mean = .16), $F(1, 14) = 4.74$, $p = .047$. In short, the best and worst levels of performance were observed in conditions without the prototype.

Comparisons with performance in Experiment 1 (mostly untrained listeners) were conducted by calculating a 95% confidence interval for each of the four conditions that were retested in Experiment 2 and determining whether the relevant mean from Experiment 1 fell inside or outside of the interval. For the two conditions in which the standard interval was more accurately tuned than the comparison interval (700-680 and 680-660), performance of the trained listeners in the present experiment proved to be significantly better than it was in Experiment 1 (i.e., the mean from Experiment 1 fell outside of the confidence interval for Experiment 2). For the

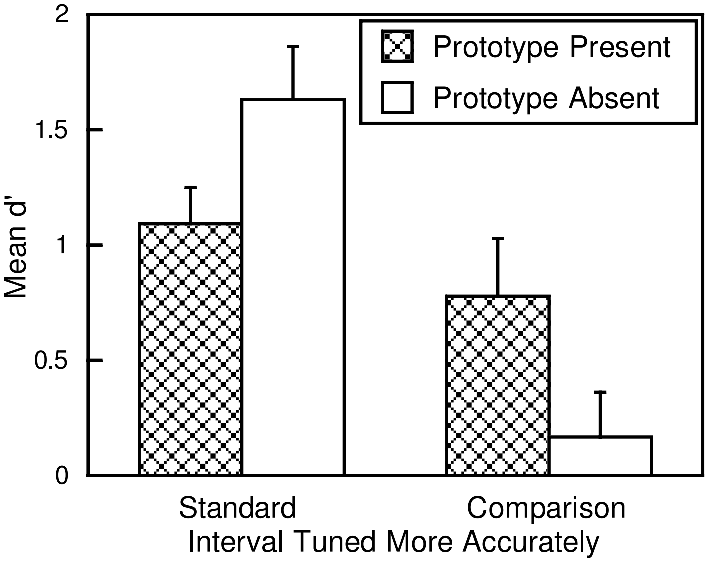

Fig. 3. Discrimination (*d′*) scores illustrating the interaction between the direction-of-mistuning manipulation and the presence or absence of the prototypic, equal-tempered perfect fifth (Experiment 2). Error bars represent standard errors.

two conditions in which the standard interval was more mistuned than the comparison (680-700, 660-680), the two groups of listeners did not differ.

The results from Experiments 1 and 2 indicate that prototypic musical intervals do not function as perceptual magnets or as perceptual anchors. Rather, pronounced asymmetries in discrimination performance are similar to those found in tests of between-category interval discrimination. Listeners more readily detect differences between relatively well-tuned standard intervals and poorly tuned comparison intervals than between poorly tuned standard intervals and well-tuned comparison intervals. These asymmetries extend to intervals both smaller and larger than an equal-tempered perfect fifth, and across listeners who vary in musical training. Interestingly, musically trained listeners in the present experiment outperformed their relatively untrained counterparts from Experiment 1 only in the “easy” conditions, when the standard interval was relatively well tuned. The more difficult conditions—with a more accurately tuned comparison interval—appear to be impervious to effects of musical training. Indeed, performance failed to exceed chance levels in any of these conditions (i.e., 680-700, 660-680, 720-700, 740-720).

The results from the present experiment also indicate that performance asymmetries are more pronounced, at least with musically trained listeners, for discriminations that do not include the prototypic, equal-tempered perfect fifth as either the standard or the comparison interval. In other words, such asymmetries appear to be a nonlinear function of deviation (in cents) from the prototype. Although one might be tempted to attribute this finding to "interval boundary" effects (i.e., conditions are easier if they include intervals near the boundary between adjacent smaller or larger intervals), it is important to keep in mind that performance was not superior for nonprototypic intervals (i.e., there was no perceptual magnet effect). Rather, the largest performance asymmetries were observed for nonprototypic intervals: remarkably *good* performance when the standard interval was more accurately tuned than the comparison interval, but remarkably *poor* performance when the standard and comparison intervals were reversed.

Experiment 3

In the present experiment, listeners were tested on their ability to make within-category discriminations of *harmonic* (simultaneous) intervals, in contrast to the *melodic* (sequential) intervals of Experiments 1 and 2. Interval discrimination is known to be easier for simultaneous than for sequential intervals because of extraneous cues from interactions between harmonic-distortion components and the tones (Burns, 1999). Indeed, pilot testing suggested that a more conventional same-different task would be manageable for simultaneous intervals. Participants included musically trained as well as completely untrained listeners.

METHOD

Participants

The participants were 32 members of the university community who received partial course credit or token remuneration. None participated in Experiments 1 or 2. Half were musically trained, with at least 8 years of music lessons or playing music regularly. The other half, designated musically untrained, had no music lessons and had never played music regularly.

Apparatus

The apparatus used was identical to that used in Experiment 1.

Stimuli

The stimuli were identical to those used in Experiment 2, with the following exceptions. Instead of tone sequences, the stimulus intervals consisted of a low tone and a high tone presented simultaneously. Tones were 1 s in duration with 10-ms linear onsets and offsets.

Procedure

On each trial, listeners heard three simultaneous intervals presented consecutively: an initial "standard" interval, a second interval, and a third interval. Either the second or the third interval was the changed interval, differing from the standard by 20 cents. The other interval was the same size as the standard. Listeners were asked to identify the interval (second or third) that differed from the first. The second and third intervals were presented in transposition, each 2.25 semitones higher than the preceding interval. The low tone of the initial standard in¥terval was selected from a set of five frequency values (250, 310, 370, 430, and 490 Hz). The frequency was chosen randomly from trial to trial, constrained such that each frequency was presented an equal number of times. Feedback was provided after each trial.

As in Experiment 2, there were eight testing conditions with each listener tested in four. Half of the listeners (counterbalanced for degree of musical training) were assigned to the four flat conditions (700-680, 680-700, 680-660, and 660-680); the other half were tested in the four sharp conditions (700-720, 720-700, 720-740, and 740-720). Testing order was determined as in Experiment 2.

Listeners were initially told about transpositions, as in the previous experiments. In each condition, a training phase preceded the test phase. The changed interval incorporated a larger change during the training phase than in the test phase, with the top tone displaced upward by an octave from the displacement in the test phase. After four practice trials, listeners proceeded to the test phase, which had 60 trials. The second interval was the changed interval on 30 trials, whereas the third interval was changed on the other 30 trials. "Second" and "third" trials were ordered randomly.

RESULTS AND DISCUSSION

A separate d' score was derived for each listener for each testing condition by using adjusted hit and false-alarm rates (as in the previous experiments), and the formula for ABX designs (Macmillan & Creelman, 1991, Table A5.3). The present design (XAB) is a type of ABX design. Descriptive statistics are provided in Table 4 for musically trained listeners (upper) and for untrained listeners (lower). For musically trained listeners, performance was better than chance in the four conditions in which the standard interval was more accurately tuned than the comparison interval [700-680: $t(7) = 4.86$, $p = .002$; 680-660: $t(7) = 3.31$, $p = .013$; 700-720: $t(7) = 5.47$, $p < .001$; and 720-740: $t(7) = 7.19$, $p < .001$], but at chance levels in all but one of the other four conditions [680-700: $t(7) = 2.38$, $p = .049$]. For untrained listeners, performance was better than chance in three of four conditions in which the standard interval was more in-tune than the comparison interval [680-660: $t(7) = 2.68$, $p = .032$; 700-720: $t(7) = 2.46$, $p = .049$; and 720-740: $t(7) = 3.39$, $p = .010$]. Performance failed to exceed chance levels in the other four conditions, in which the standard and comparison intervals were reversed.

The data were analyzed with a 2 × 2 × 2 × 2 mixed-design ANOVA that included two between-subjects variables (musical training, flat or sharp intervals) and two within-subjects variables (presence of prototype, direction of mistuning). Again, the main effect of the direction-of-mistuning variable was highly significant, $F(1, 28) = 21.84$, $p < .001$ (see Figure 2). Performance was better when the standard interval was more accurately tuned than the comparison interval (mean = 1.09) compared to the reverse

TABLE 4
Descriptive Statistics From Experiment 3 (Harmonic Intervals, Musically Trained and Untrained Listeners)

Musically Trained Listeners							
Flat Conditions				Sharp Conditions			
Condition	N	Mean	SD	Condition	N	Mean	SD
700-680	8	2.181	1.269	700-720	8	0.864	0.446
680-700	8	1.455	1.726	720-700	8	-0.325	0.945
680-660	8	1.626	1.388	720-740	8	0.981	0.386
660-680	8	0.350	0.783	740-720	8	-0.061	0.921
Untrained Listeners							
Flat Conditions				Sharp Conditions			
Condition	N	Mean	SD	Condition	N	Mean	SD
700-680	8	0.291	0.806	700-720	8	0.969	1.114
680-700	8	0.635	1.143	720-700	8	0.911	1.397
680-660	8	0.945	0.999	720-740	8	0.847	0.688
660-680	8	-0.601	0.732	740-720	8	0.004	1.124

situation (mean = .30). Two other results were significant: the two-way interaction between musical training and flat or sharp intervals, *F*(1, 28) = 11.97, *p* = .002, and the two-way interaction between direction of mistuning and presence of the prototype, *F*(1, 28) = 4.74, *p* = .038. As with sequential intervals (Experiments 1 and 2), the ANOVA provided no evidence that simultaneous musical intervals function as perceptual magnets or as perceptual anchors (i.e., there was no main effect of presence of prototype).

The interaction between musical training and flat or sharp intervals was investigated by analyzing the flat and sharp intervals separately (see Figure 4). For flat intervals, the trained group (mean = 1.40) had significantly better discrimination performance than the untrained group (mean = .318), *F*(1, 14) = 13.41, *p* = .003. There was no difference between groups for sharp intervals.

Further investigation of the interaction between direction of mistuning and presence of the prototype (see Figure 5) revealed that when the standard interval was more in-tune than the comparison interval, the presence of the prototype had no effect on performance (present: mean = 1.08; absent: mean = 1.10). Indeed, performance was reasonably good in all four conditions (700-680, 680-660, 700-720, 720-740) in which the standard interval was more accurately tuned than the comparison interval. For conditions in which the comparison interval was more in-tune than the standard, performance was superior when the prototype was present (680-700, 720-700; mean = .67) rather than absent (660-680, 740-720; mean = -.08),

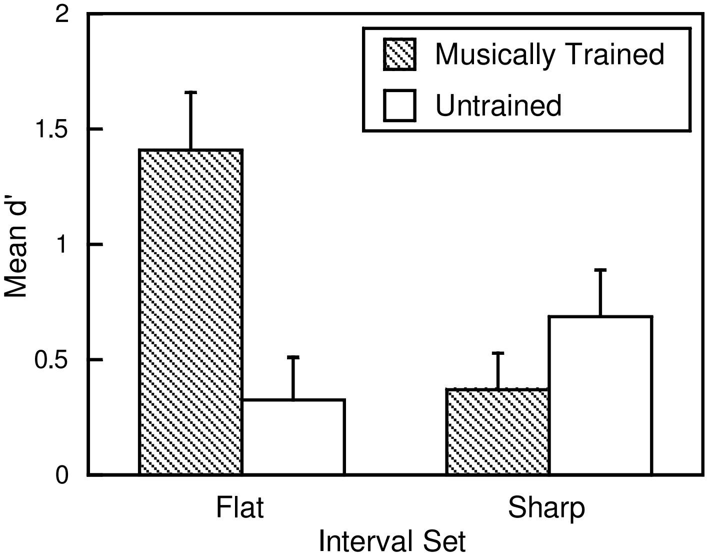

Fig. 4. Discrimination (*d′*) scores illustrating the interaction between musical training and the set of stimulus intervals (Experiment 3). Intervals in the "flat" set were 700, 680, and 660 cents. Intervals in the "sharp" set were 700, 720, and 740 cents. Error bars represent standard errors.

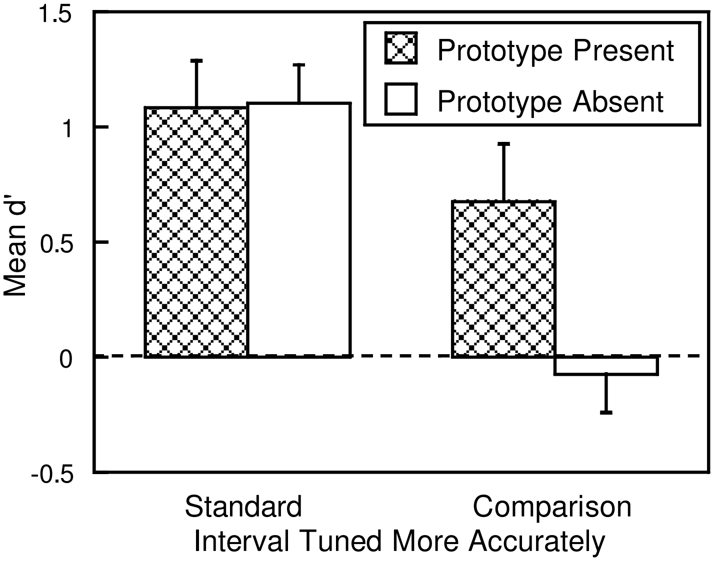

Fig. 5. Discrimination (*d′*) scores illustrating the interaction between the direction-of-mistuning manipulation and the presence or absence of the prototypic, equal-tempered perfect fifth (Experiment 3). Error bars represent standard errors.

$F(1, 29) = 7.16$, $p = .012$. In other words, although performance was generally poor when the comparison interval was more in-tune than the standard interval, it was particularly poor for conditions without the prototype, as it was in Experiment 2.

General Discussion

Musically trained and untrained listeners were tested on their ability to make within-category discriminations of musical intervals belonging to the perfect-fifth category. Pure-tone intervals were presented sequentially in Experiments 1 and 2 and simultaneously in Experiment 3. Across experiments and listener groups, performance asymmetries were observed. Listeners found it much easier to discriminate a pair of intervals when the standard interval was more accurately tuned than the comparison interval, compared with when the comparison interval was more in-tune than the standard. Discriminations involving the prototypic, equal-tempered perfect fifth were neither easier nor more difficult than other discriminations. In short, there was no evidence that perceptual-magnet or perceptual-anchor effects extend to within-category discrimination of musical intervals.

The present findings are notable for revealing that discrimination asymmetries for perfect fifths extend to within-category contexts. Such asymmetries can be explained by considering musical prototypes to be particularly stable—or well perceived and remembered—auditory events (e.g., Acker et al., 1995; Schellenberg & Trehub, 1999). Although it is relatively easy to detect a subtle alteration to a stimulus that has a stable representation, it is more difficult to detect the same alteration to a stimulus with an unstable representation. Hence, when listeners are asked to discriminate a stable stimulus from another, slightly different but relatively unstable stimulus, performance is much better when the stable stimulus is the to-be-remembered standard and the unstable stimulus the to-be-detected comparison, compared with the reverse situation with an unstable standard and stable comparison. The present findings make it clear that such asymmetries do not require the presence of a prototypic category member. Indeed, they were somewhat larger when slightly mistuned intervals (680 or 720 cents) were discriminated from intervals with greater mistunings (660 or 740 cents). Whereas musical intervals vary *continuously* in their degree of prototypicality, the numbers tested by Rosch (1975a) were either prototypes or nonprototypes. Hence, asymmetries in agreement with statements involving two numbers that are equally nonprototypic ("101 is essentially 102" vs. "102 is essentially 101") seem unlikely.

The results imply that any pairwise comparison involving stimuli that differ in degree of prototypicality, stability, goodness, and so on, could yield asymmetries in perceived similarity space, whether or not one of the stimuli is actually a prototype. Support for this suggestion is provided by evidence of asymmetries across a range of auditory stimuli, including rhythmic sequences and sentences. In each case, discrimination of relatively "regular" (e.g., isochronous rhythms, coherent sentences) and "irregular" (e.g., disrupted rhythms, anomalous sentences) stimuli is better when the standard stimulus is regular and the comparison irregular compared with the reverse situation (Bharucha, Olney, & Schnurr, 1985; Bharucha & Pryor, 1986). Moreover, when listeners provide similarity ratings instead of making discrimination judgments, identical asymmetries are apparent. For example, when listeners rate the similarity of pairs of melodies, they give higher ratings when the comparison melody is more typical (i.e., conforming to Western scale structure) than the standard melody compared to when the standard and comparison are reversed (Bartlett & Dowling, 1988). These asymmetries in similarity ratings mirror the asymmetries observed among 5-year-old children and adults in melody-discrimination tasks (Schellenberg & Trehub, 1999).

A slightly different view of asymmetries in music perception concerns particular tones presented in a musical context. When a musical key is established, some tones are perceived to be more or less stable than others in the key. Tones that are unstable in the key (e.g., F♯ in the key of C) are considered to be more closely related to stable tones (e.g., G in the key of C) than stable tones are to unstable tones (Krumhansl, 1979, 1990). Krumhansl (1990) describes this phenomenon as the "contextual asymmetry" principle. Listeners also find it relatively easy to detect when a stable tone (e.g., *doh*) in a musical context (e.g., a short melody) is replaced by an unstable tone (e.g., *ti*), but relatively difficult to detect when an unstable tone is replaced by a stable tone (Bharucha, 1984). They appear to encode stable *doh* in a straightforward and reliable manner, whereas *ti* may be encoded as "almost *doh*" or "leaning towards *doh*". In the latter case, substitution of the unstable tone for the stable tone resolves the perceived instability, obscuring the difference between the tones. This perspective attributes poor performance to a relatively stable comparison stimulus, instead of attributing good performance to a relatively stable standard stimulus. In the present series of experiments, a comparison interval that was more accurately tuned than the standard interval may have served a similar function, in spite of the fact that no musical context was established. In general, then, alterations to a stimulus that make it a "better" stimulus (e.g., more stable, coherent, familiar) may be relatively difficult to detect.

Perceptual asymmetries for within-category discriminations are consistent with the idea that within-category similarity space is warped or distorted because of the presence of a reference point, or prototype. Nonetheless, asymmetries in the present investigation reveal a distortion quite different from that described by perceptual magnets or perceptual anchors. The magnet effect describes a shrunken similarity space surrounding the prototype, such that perceptual differences among stimuli close to the prototype are *smaller* than physical differences would otherwise dictate. Conversely, anchor effects describe an expanded similarity space around the prototype, such that perceptual differences among stimuli close to the prototype are *greater* than physical differences would otherwise dictate. Neither perspective considers dynamic aspects, or how stimuli are presented over time.

The asymmetries reported here point to distortions that depend on order of stimulus presentation (i.e., which stimulus is the standard—presented first, and which is the comparison—presented second). Similarity space is expanded (i.e., discrimination is good) when the standard pattern is more prototypic than the comparison pattern, consistent with a perceptual anchor effect. But similarity space is contracted (i.e., discrimination is poor) when the comparison pattern is more prototypic than the standard pattern, consistent with a perceptual magnet effect. When stimuli are tested in both directions (by reversing the order of presentation), magnet and anchor effects appear to cancel each other out. The present findings also provide some evidence that asymmetric distortions in similarity space increase as one moves away from the category center.

On the one hand, the perceptual magnet effect may fail to extend to music perception because it is specific to speech. On the other hand, its status in speech perception is also questionable. Despite demonstrations of the effect with adults and infants (Iverson & Kuhl, 1995, 1996, 2000; Kuhl, 1991; Kuhl et al., 1992; Polka & Werker, 1994), there have been several replication failures (Lively & Pisoni, 1997; Lotto et al., 1998; Sussman & Gekas, 1997) or equivocal findings (Aaltonen et al., 1997; Frieda et al., 1999; Polka & Bohn, 1996; Sussman & Lauckner-Morano, 1995). The original finding (Kuhl, 1991) may be attributable to comparisons without the prototype involving many between-category contrasts, whereas comparisons involving the prototype were always within-category (Lively & Pisoni, 1997; Sussman & Lauckner-Morano, 1995). Hence, relatively poor discrimination with comparisons involving the prototype may have resulted from a design flaw.

Previous reports of perceptual anchors in music perception (Acker & Pastore, 1996; Acker et al., 1995, McFadden & Calloway, 1999) may also benefit from further examination. In each of these reports, the possibility of asymmetric performance was not examined. When a prototypic stimulus was included in a discrimination task, it was always presented as the standard pattern. Similar anchor effects would have been uncovered in the present study if the same experimental design were adopted. In other words, the previous results are entirely consistent with the present findings. Nonetheless, the present findings imply that the story is more complicated than the conclusions proposed earlier. Future research could test the prediction of asymmetric discrimination performance with the stimuli used in previous investigations.

Effects of musical training are relatively robust across a wide variety of musical tasks, including tests of categorical perception, octave generalization, awareness of melodic transpositions, sensitivity to the diatonic hierarchy, and so on (for a review, see Smith, 1997). One might therefore have expected substantial advantages for musically trained listeners in discrimination tests that rely on relative pitch perception, such as the ones used in the present experiments. In Experiment 1, however, although discrimination performance was positively correlated with time spent playing music on a regular basis, the association accounted for just 2.6% of the variance in the data. Comparisons of results from Experiments 1 and 2 revealed an advantage for trained listeners in the relatively easy conditions (standard interval more in-tune) but not in the more difficult conditions (comparison interval more in-tune). Finally, in Experiment 3, trained listeners outperformed untrained listeners only with flat intervals.

The subtle differences listeners were asked to detect may explain the relatively small effects of musical training. Such differences (1/5 of a semitone) have no structural meaning in Western music, in which relevant distinctions must be at least 1 semitone in size. Rather, when within-category deviations in relative pitch are performed intentionally, they are considered to be changes in intonation that help to convey the emotional meaning of a musical piece (Sloboda, 1985). Musicians also exhibit categorical perception of musical intervals (for a review, see Burns, 1999), which might have counteracted the usual advantage over their untrained counterparts. The present findings make it clear, however, that musically trained and untrained listeners can discriminate musically "irrelevant" differences, at least in some contexts. Similarly, in the appropriate experimental contexts, listeners can often detect linguistically irrelevant differences, that is, differences between tokens that belong to the same phonemic category (e.g., Carney, Widin, & Viemeister, 1977; Pisoni & Tash, 1974; Werker & Logan, 1985). The present findings also imply that asymmetries in discrimination do not require explicit knowledge of category membership or prototypic status.

Although these findings provide no evidence that prototypic musical intervals function as perceptual magnets or anchors, such effects could emerge with different stimuli, different listeners, or different methods. Regardless, the performance asymmetries observed here were robust across such differences. At the very least, future attempts to measure perceived similarities or differences between pairs of stimuli should consider the potential for asymmetries. The apparent similarity—and hence discriminability—between stimulus X and stimulus Y may often be quite different from the similarity between stimulus Y and stimulus X. Clarification of the function of perceptual magnets, perceptual anchors, and perceptual asymmetries could improve our knowledge of speech and music perception as well as our understanding of the structure of psychological categories.[1]

References

Aaltonen, O., Eerola, O., Hellström, Å, Uusipaikka, E., & Lang, A. H. (1997). Perceptual magnet effect in the light of behavioral and psychophysiological data. *Journal of the Acoustical Society of America, 101*, 1090–1105.

Acker, B. E., & Pastore, R. E. (1996). Perceptual integrality of major chord components. *Perception & Psychophysics, 58*, 748–761.

Acker, B. E., Pastore, R. E., & Hall, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception & Psychophysics, 57*, 863–874.

Bartlett, J. C., & Dowling, W. J. (1988). Scale structure and similarity of melodies. *Music Perception, 5*, 285–314.

Bharucha, J. J. (1984). Anchoring effects in music: The resolution of dissonance. *Cognitive Psychology, 16*, 485–518.

Bharucha, J. J., Olney, K. L., & Schnurr, P. P. (1985). Detection of coherence-disrupting and coherence-conferring alterations in text. *Memory & Cognition, 13*, 573–578.

Bharucha, J. J., & Pryor, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception & Psychophysics, 40*, 137–141.

Bregman, A. S. (1990). *Auditory scene analysis.* Cambridge, MA: MIT Press.

Burns, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 215–264). San Diego, CA: Academic Press.

Burns, E. M., & Campbell, S. L. (1994). Frequency and frequency-ratio resolution by possessors of absolute and relative pitch: Examples of categorical perception? *Journal of the Acoustical Society of America, 96*, 2704–2719.

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of musical intervals. *Journal of the Acoustical Society of America, 63*, 456–468.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America, 62*, 961–970.

Cohen, J. D., MacWhinney, B., Flatt, M., & Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments & Computers, 25*, 257–271.

Eilers, R. E., Wilson, W. R., & Moore, J. M (1977). Developmental changes in speech discrimination in infants. *Journal of Speech and Hearing Research, 20*, 766–780.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (1999). Adults' perception of native and nonnative vowels: Implications for the perceptual magnet effect. *Perception & Psychophysics, 61*, 561–577.

Frieda, E. M., Walley, A. C., Flege, J. E., & Sloane, M. E. (2000). Adults' perception and production of the English vowel /i/. *Journal of Speech, Language, and Hearing Research, 43*, 129–143.

---

[1] Funding for this research was provided by the Natural Sciences and Engineering Research Council of Canada. I thank Margaret McKinnon and Jane Campbell for testing the listeners, Paul Pilon for writing the computer program, and Ed Burns, Jay Dowling, Paul Iverson, Wayne Slawson, Bill Thompson, and Sandra Trehub for helpful comments on earlier versions of the manuscript.

Grieser, D. L., & Kuhl, P. K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology, 25*, 577–588.

Helmholtz, H. L. F. von (1954). *On the sensations of tone as a physiological basis for the theory of melody* (rev. ed., A. J. Ellis, Ed. and Trans.). New York: Dover. (Original work published 1885)

Hulse, S. H., Bernard, D. J., & Braaten, R. F. (1995). Auditory discrimination of chord-based spectral structures by European starlings (*Sturnus vulgaris*). *Journal of Experimental Psychology: General, 124*, 409–423.

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America, 97*, 553–561.

Iverson, P., & Kuhl, P. K. (1996). Influence of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *Journal of the Acoustical Society of America, 99*, 1130–1140.

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception & Psychophysics, 62*, 874–886.

Kameoka, A., & Kuriyagawa, M. (1969). Consonance theory: Part I. Consonance of dyads. *Journal of the Acoustical Society of America, 45*, 1451–1459.

Krumhansl, C. L. (1979). The psychological representation of musical pitch in a tonal context. *Cognitive Psychology, 11*, 346–374.

Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New York: Oxford University Press.

Kuhl, P. K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics, 50*, 93–107.

Kuhl, P. K., Williams, K. A., Lacerda, F., Stevens, K. N., & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science, 255*, 606–608.

Lively, S. E., & Pisoni, D. B. (1997). On prototypes and phonetic categories: A critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1665–1679.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1998). Depolarizing the perceptual magnet effect. *Journal of the Acoustical Society of America, 103*, 3648–3655.

Lynch, M. P., & Eilers, R. E. (1992). A study of perceptual development for musical tuning. *Perception & Psychophysics, 52*, 599–608.

Lynch, M. P., Eilers, R. E., Oller, D. K., & Urbano, R. C. (1990). Innateness, experience, and music perception. *Psychological Science, 1*, 272–276.

Lynch, M. P., & Steffens, M. L. (1994). Effects of aging on processing of novel musical structure. *Journal of Gerontology: Psychological Sciences, 49*, P165–P172.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, UK: Cambridge University Press.

McFadden, D., & Calloway, N. L. (1999). Better discrimination of small changes in commonly encountered than in less commonly encountered auditory stimuli. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 543–560.

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*, 89–115.

Meyer, L. B. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

Miller, J. L., & Volaitis, L. E. (1989). Effects of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics, 46*, 505–512.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics, 15*, 285–290.

Plomp, R., & Levelt, W. J. M. (1965). Tonal consonance and critical bandwidth. *Journal of the Acoustical Society of America, 38*, 548–560.

Polka, L., & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America, 100*, 577–592.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of nonnative vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421–435.

Rakowski, A. (1990). Intonation variants of musical intervals in isolation and in musical contexts. *Psychology of Music, 18*, 60-72.

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). New York: Academic Press.

Rosch, E. (1975a). Cognitive reference points. *Cognitive Psychology, 7*, 532–547.

Rosch, E. (1975b). Universals and cultural specifics in human categorization. In R. W. Brislin, S. Bochner, & W. J. Lonner (Eds.), *Cross-cultural perspectives on learning* (pp. 177–206). Beverly Hills, CA: Sage.

Rosch, E., & Mervis, C. B. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology, 7*, 573–605.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382–439.

Samuel, A. G. (1982). Phonetic prototypes. *Perception & Psychophysics, 31*, 307–314.

Schellenberg, E. G., & Trainor, L. J. (1996). Sensory consonance and the perceptual similarity of complex-tone harmonic intervals: Tests of adult and infant listeners. *Journal of the Acoustical Society of America, 100*, 3321–3328.

Schellenberg, E. G., & Trehub, S. E. (1994a). Frequency ratios and the discrimination of pure tone sequences. *Perception & Psychophysics, 56*, 472–278.

Schellenberg, E. G., & Trehub, S. E. (1994b). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review, 1*, 191–201.

Schellenberg, E. G., & Trehub, S. E. (1996a). Children's discrimination of melodic intervals. *Developmental Psychology, 32*, 1039–1050.

Schellenberg, E. G., & Trehub, S. E. (1996b). Natural musical intervals: Evidence from infant listeners. *Psychological Science, 7*, 272–277.

Schellenberg, E. G., & Trehub, S. E. (1999). Redundancy, conventionality, and the discrimination of tone sequences: A developmental perspective. *Journal of Experimental Child Psychology, 74*, 107–127.

Siegel, J. A., & Siegel, W. (1977a). Absolute identification of notes and intervals by musicians. *Perception & Psychophysics, 21*, 143–152.

Siegel, J. A., & Siegel, W. (1977b). Categorical perception of tonal intervals: Musicians can't tell *sharp* from *flat*. *Perception & Psychophysics, 21*, 399–407.

Sloboda, J. A. (1985). *The musical mind: The cognitive psychology of music*. Oxford, UK: Clarendon Press.

Smith, J. D. (1997). The place of musical novices in music science. *Music Perception, 14*, 227–262.

Smith, J. D., Kemler Nelson, D. G., Grohskopf, L. A., & Appleton, T. (1994). *What child is this*? What interval was that? Familiar tunes and music perception in novice listeners. *Cognition, 52*, 23–54.

Sternberg, R. J. (1999). *Cognitive psychology* (2nd ed.). Fort Worth, TX: Harcourt Brace College.

Sussman, J. E., & Gekas, B. (1997). Phonetic category structure of [I]: Extent, best exemplars, and organization. *Journal of Speech, Language, and Hearing Research, 40*, 1406–1424.

Sussman, J. E., & Lauckner-Morano, V. J. (1995). Further tests of the "perceptual magnet effect" in the perception of [i]: Identification and change/no-change discrimination. *Journal of the Acoustical Society of America, 97*, 539–552.

Takeuchi, A. H., & Hulse. S. H. (1993). Absolute pitch. *Psychological Bulletin, 113*, 345–361.

Thorpe, L. A., Trehub, S. E., Morrongiello, B. A., & Bull, D. (1988). Perceptual grouping by infants and preschool children. *Developmental Psychology, 24*, 484–491.

Trainor, L. J. (1997). The effect of frequency ratio on infants' and adults' discrimination of simultaneous intervals. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 1427–1438.

Trainor, L. J., & Heinmiller, B. M. (1998). The development of evaluative responses to music: Infants prefer to listen to consonance over dissonance. *Infant Behavior & Development, 21*, 77–88.

Trainor, L. J., & Trehub, S. E. (1992). A comparison of infants' and adults' sensitivity to Western musical structure. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 394–402.

Trainor, L. J., & Trehub, S. E. (1993a). Musical context effects in infants and adults: Key distance. *Journal of Experimental Psychology: Human Perception and Performance, 19*, 615–626.

Trainor, L. J., & Trehub, S. E. (1993b). What mediates infants' and adults' superior processing of the major over the augmented triad? *Music Perception, 11*, 185–196.

Trehub, S. E. (2000). Human processing predispositions and musical universals. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 427–448). Cambridge, MA: Bradford.

Trehub, S. E., Schellenberg, E. G., & Hill, D. (1997). Music perception and cognition: A developmental perspective. In I. Deliège & J. A. Sloboda (Eds.), *Music perception and cognition* (pp. 121–162). Sussex, UK: Psychology Press.

Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America, 92*, 723–735.

Ward, W. D. (1999). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (2nd ed., pp. 265–298). San Diego, CA: Academic Press.

Werker, J. F., & Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics, 37*, 35–44.

Wright, A. A., Rivera, J. J., Hulse, S. H., Shyan, M., & Neiworth, J. J. (2000). Music perception and octave generalization in rhesus monkeys. *Journal of Experimental Psychology: General, 129*, 291–307.

Zentner, M. R., & Kagan, J. (1998). Infants' perception of consonance and dissonance in music. *Infant Behavior & Development, 21*, 483–492.