ABSOLUTE JUDGMENT OF MUSICAL INTERVAL WIDTH

CHRISTOPHER ARUFFO
McMaster University, Hamilton, Ontario, Canada

ROBERT L. GOLDSTONE
Indiana University

DAVID J. D. EARN
McMaster University, Hamilton, Ontario, Canada

WHEN A MUSICAL TONE IS SOUNDED, MOST LISTENERS are unable to identify its pitch by name. Those listeners who can identify pitches are said to have *absolute pitch* perception (AP). A limited subset of musicians possesses AP, and it has been debated whether musicians' AP interferes with their ability to perceive tonal relationships between pitches, or *relative pitch* (RP). The present study tested musicians' discrimination of relative pitch categories, or *intervals*, by placing absolute pitch values in conflict with relative pitch categories. AP listeners perceived intervals categorically, and their judgments were not affected by absolute pitch values. These results indicate that AP listeners do not infer interval identities from the absolute values between tones, and that RP categories are salient musical concepts in both RP and AP musicianship.

Received: September 12, 2011, accepted March 31, 2014.

Key words: absolute pitch, relative pitch, categorical perception, music perception, music cognition

YOU CAN PROBABLY RECOGNIZE THE BEEPING of a car horn—but can you name the musical tone it produces? If you can, you possess the uncommon ability of *absolute pitch*: the ability to identify individual sounds by their musical note names, such as "F-sharp" or "B-flat" (Ward & Burns, 1982). The majority of people, including most musicians, do not possess absolute pitch (Profita & Bidder, 1988), but the ability to name musical tones is not necessary for hearing or making music. Melodies are perceived as successive changes from one tone to another, which makes the tones' absolute values irrelevant. Harmony is produced by the interactions between tones, which depends on tones' relationships to each other, not their absolute values. Recognizing the *relative* values between tones, therefore, is essential to musical experience, whereas recognizing their absolute values appears to be a superfluous novelty. Yet absolute pitch ability can impact musical experience by interfering with relative pitch tasks (for a review, see Miyazaki, 2004a). For example, in recognizing a *transposed* melody—a song that has been moved upwards or downwards in absolute value, but with tones' relative values remaining the same—it is easy to remember the pattern of relationships. Yet a musician with absolute pitch may instead try to remember every pitch from the original melody and calculate its new position. What we do not yet know is why musicians with absolute pitch would use an inefficient strategy, relying on absolute pitch, instead of an efficient strategy that uses relative pitch. Perhaps, because absolute pitch tends to appear in the early stages of human development (Sergeant, 1969), musicians might learn at an early age to rely on absolute pitch strategies, and these strategies could persist into adulthood out of habit. Alternatively, early absolute pitch perception might prevent the natural development of relative pitch perception; adult musicians might then be forced to apply absolute pitch strategies because relative pitch values are difficult to recognize or, perhaps, may not be perceived at all (Parncutt & Levitin, 2001). The present experiment was designed to test whether musicians with absolute pitch show a deficit in relative pitch perception because it suffers from competition with absolute pitch perception.

It is not known whether absolute pitch ability exerts a mandatory influence on relative pitch perception (henceforth relative perception, or RP). Musicians who possess absolute pitch perception (AP), or *absolute listeners*, are unable to ignore the absolute quality of a pitch (Miyazaki, 2004b). The primacy of absolute pitch quality could therefore weaken access to relative perception, interfering with tasks that require relative judgment. For example, absolute listeners find it difficult to match a piece of sheet music to a transposed melody (Miyazaki & Rakowski, 2002), detect a changed tone in transposition (Miyazaki, 2004c), or identify tonal relationships, or *intervals*, within an unusual context (Miyazaki, 1993, 1995), whereas musicians who exhibit only RP, or *relative listeners*, perform these tasks easily. Yet relative perception does not necessarily suffer in the presence of AP.

Music Perception, VOLUME 32, ISSUE 2, PP. 184–198, ISSN 0730-7829, ELECTRONIC ISSN 1533-8312. DOI: 10.1525/MP.2014.32.2.184

In fact, in tasks that do not require absolute judgment, absolute listeners have shown excellent performance, such as transcribing musical dictation (Dooley & Deutsch, 2010) or identifying intervals in various musical keys and contexts (Dooley & Deutsch, 2011). However, while these tasks did not require AP, they were also not in conflict with AP, because the musical pitches whose relationships were to be identified did conform to standard absolute pitch values. When RP and AP present conflicting information, the ability to make relative judgments may be inversely related to the ability to make absolute judgments (Benguerel & Westdal, 1991). In other words, it is not certain whether absolute listeners' relative judgments represent their actual perception of relative musical qualities or, instead, represent logical inferences derived from absolute qualities.

CATEGORICAL PERCEPTION

Absolute and relative perception may be disambiguated by testing categorical perception (CP) of musical intervals. CP is a phenomenon in which a continuum is perceived to be divided into discrete categories. CP was originally observed in linguistics as an explanation for the acoustic variability of phonemes (Liberman, Harris, Hoffman, & Griffith, 1957). That is, speakers differ widely in their physical production of speech, but a listener perceives only a limited inventory of meaningful sounds. As stimuli proceed along a physical continuum, each stimulus is identified with the same label until a category boundary is reached, at which point perception shifts relatively abruptly to the next category. Discrimination of small differences improve as a stimulus moves toward a categorical boundary, because people can use category identity to help them make the discrimination (Goldstone & Hendrickson, 2010). Given that stimuli on opposite sides of a category boundary possess distinct categorical identities, it is easiest to discriminate differences that cross category boundaries. Thus a categorical boundary may be recognized, along a physical continuum, by a peak in discrimination accuracy occurring at the same location as an abrupt change in identification (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970).

Categories for pitches and intervals are defined along the continuum of sound frequency. On this continuum, the "width" between any two points is perceived logarithmically such that, for example, values of 220 and 440 Hz are perceived to be the same "distance apart" as 440 and 880 Hz. Units of scale are derived from a special distance, the *octave*. An octave is defined by any two frequencies that form a ratio of two to one (e.g., 220 and 440 Hz). On octave may be divided into twelve equal widths to produce twelve *semitones*. A semitone may be further divided into 100 equal widths, or *cents*. Pitch categories and interval categories each differ by one semitone in width. Each category is defined by a median value (or *prototype*) and extends 50 cents in either direction, and category members above or below the prototype are referred to as *sharp* or *flat,* respectively. A pitch category prototype is defined as a single specific absolute frequency (e.g., A = 440 Hz), whereas interval prototypes are widths defined as the multiples of 100 cents between two frequencies (e.g., octave = 1,200 cents). Interval and pitch categories may therefore be placed in conflict by selecting intervals in which frequencies do not conform to the standard absolute pitch values of the Western musical scale.

Absolute listeners' internal representations of pitch categories may not necessarily conform precisely to the standard absolute pitch values of the Western musical scale (Burns & Campbell, 1994). Nonetheless, standard values for absolute pitches were established in 1955, with "concert pitch" as A = 440 Hz (Mendel, 1978), so modern listeners have substantial experience with these standards, and a majority of absolute listeners will accurately identify mistuned musical sounds according to these standard pitch categories (Miyazaki, 1988). Moreover, all participants were recruited for the present experiment from the Jacobs School of Music at Indiana University, and therefore shared these standards in performance and practice. Standard absolute pitch values were therefore presented in this design.

Trained musicians exhibit categorical perception for musical sounds. Absolute listeners demonstrate CP for pitch (Harris & Siegel, 1975; Rakowski, 1993) and relative listeners demonstrate CP for intervals (Burns & Ward, 1978). However, these same results show that relative listeners do not demonstrate CP for pitch, and that untrained listeners do not demonstrate CP for either pitches or intervals. Because music production and perception is interval based, it may be expected that continued musical practice would strengthen a musician's representation of interval categories, and this is the case for relative listeners (Howard, Rosen, & Broad, 1992). However, to the extent that absolute listeners rely on AP to perceive melody and harmony, their representations of interval categories would remain weak—not only would absolute listeners' experience with intervals remain inferential and indirect, but their ongoing practice would serve to strengthen CP for pitches instead. Prior to the present investigation, absolute listeners have not been tested for interval CP. A test of CP for intervals, featuring nonstandard pitch frequencies, should determine whether AP musicianship is or is not independent of relative perception.

PROTOTYPE EFFECTS

The present design could be subject to *prototype effects*, independently of listeners' absolute or relative perception. A prototype represents a strong exemplar of a particular category and may therefore cause a "magnet" or an "anchor" effect. A *magnet* prototype assimilates its neighboring stimuli, making small deviations difficult to perceptually discriminate from the prototype. As stimuli move away from a magnet prototype, small differences are easier to discriminate, because they assimilate less strongly (Iverson & Kuhl, 1995). By contrast, an *anchor* prototype is a value so clearly fixed in long-term memory that small deviations from it are easily detected. As stimuli move away from an anchor prototype, small differences become more difficult to detect, because comparisons to the prototype become more difficult (Braida, 1984). A musical prototype, defined as the mathematical center of a pitch or interval category, may produce either a magnet or anchor effect.

Pitch and interval prototypes may behave as anchors, magnets, or neither. A musical prototype can assimilate an area of approximately 25 cents in either direction (Vurma & Ross, 2006), i.e., musically trained listeners may perceive all stimuli within 25 cents of a prototype to be identical (Perlman & Krumhansl, 1996; Siegel & Siegel, 1977a). Alternatively, a prototype may show enhanced discrimination in its immediate vicinity compared to the same area surrounding a non-prototype (Acker, Pastore, & Hall, 1995; McFadden & Callaway, 1999). Neither of these effects must necessarily occur, even among trained listeners (Schellenberg, 2002); however, it may be that with increasing musical experience, a prototype gradually changes from magnet to anchor, as detecting fine mistunings becomes more important to professional performance (Barrett, 1999). A magnet effect would assimilate stimuli to the prototype, reducing discrimination accuracy, whereas an anchor effect would show greatest accuracy near the prototype. The possibility of observing either effect may be accommodated here by testing discrimination within 25 cents of a prototype.

EXPERIMENTAL DESIGN

Our experimental design follows a two-alternative forced-choice paradigm for testing CP discrimination. Each trial presented a standard stimulus X followed by two comparison stimuli A and B. Participants then indicated which of A or B was identical to X. Scores were measured as the percentage of accurate judgments made.

Participants were separated into four groups by labeling ability for pitches and intervals. Labeling ability was used as a between-groups determinant because labeling skill is strongly correlated with CP for intervals (Siegel & Siegel, 1977b) as well as pitches (Miyazaki, 1993). Following Miyazaki (1993), we implemented thresholds of 90% and 40% pitch labeling accuracy to classify participants as absolute listeners (AL) or partial-absolute listeners (PL), respectively. This distinction arises from the fact that absolute identification can be accomplished by strategies other than direct judgment of absolute value (Stewart, Brown, & Chater, 2005). For example, absolute pitch values can be accurately inferred from kinesthetic feeling (Bachem, 1937) or association to a familiar tone (Brady, 1970; Cuddy, 1968, 1970), making it possible to identify pitches significantly above chance level without actually "possessing AP" (Bermudez & Zatorre, 2009). We adopted the 90% threshold because musicians whose labeling skill is less accurate than 90% show weaker evidence of CP for pitches (Miyazaki, 1993), and we wanted to test participants with strong CP.

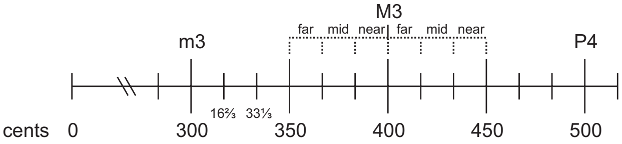

FIGURE 1. The stimulus range tested for interval discrimination.

Stimuli were selected to test CP of interval categories. Discrimination was tested within a range of widths from 300 to 500 cents (Figure 1), which encompassed three common interval prototypes: 300 cents, or "minor third" (m3); 400 cents, or "major third" (M3); and 500 cents, or "perfect fourth" (P4). Because the just noticeable difference for interval width is 16 cents (Burns, 1999), the range was divided into 12 segments of equal width, with each segment 16⅔ cents wide. It was not necessary to make the task unusually difficult, because both AL and RL demonstrate equivalent skill at discriminating musical sounds within a category (Fujisaki & Kashino, 2002). Dividing the range into 12 segments provided discrimination within three different categories at three different distances from their prototypes: *near, middle*, and *far.* If listeners possessed CP for interval categories, discriminations should become easier, and accuracy should increase, as stimuli moved away from the prototype and toward the boundary.

The intervals featured in the present design are those featured in the most common standard of Western musical systems, known as "equal tempered" intervals. In an equal-tempered system, intervals are devised by dividing a 1200-cent interval, or *octave*, into 12 logarithmically equal segments. Although musicians may deviate from this standard in performance (Mason, 1960), listeners with significant exposure to the equal-tempered scale will judge equal-tempered intervals as "preferable" to other standards (Ward & Martin, 1961), suggesting that for our listening task, equally tempered intervals will present the most salient categories.

Stimuli were designed to test categorical discrimination of interval width. Each stimulus was an ascending trichord with a variable middle tone. A *trichord* is a musical structure of three tones, in which the lowest- and highest-frequency tones are separated by a 700-cent interval, or *perfect fifth*, and a third "middle" tone is located between the other two. Interval comparisons were generated by varying the middle tone, because a trichord's middle tone is the most salient to its identity (Acker & Pastore, 1996). Trichords were used, rather than two-tone intervals, because trichords have been explicitly observed to function as categorical prototypes, exhibiting either magnet or anchor effects depending on the manner of their presentation and a listener's level of musical experience (Barrett, 1999). The top tone of a trichord also forms an interval with its middle tone, providing more sensory information and musical context to a listener than would a two-tone interval. Even so, a trichord is defined relative to its bottom tone, or "root" tone. Thus a trichord whose root and middle tones form a 300-cent interval, or *minor third*, is perceived as a minor-third chord, despite the fact that the middle and top tones form a 400-cent interval, or *major third*. Our design and analysis therefore also defined each trichord based on the interval formed by its bottom and middle tones. A trichord is "ascending" when its tones are played sequentially from lowest to highest, "descending" when its tones are played from highest to lowest, and "harmonic" when the tones are sounded simultaneously. Ascending trichords were used because ascending intervals are easiest to identify versus descending or harmonic (Samplaski, 2005), providing the greatest opportunity for categorization. Harmonic intervals were also rejected because harmonic intervals can draw attention away from tonal relationships (Zatorre & Halpern, 1979), and because harmonic intervals create *beating*—an audible phenomenon that occurs when harmonic tones interact. When beats are present, they can make discrimination difficult for sensitive listeners (Vos, 1982) and overpower fine differences (McFadden & Callaway, 1999); alternatively, some listeners can use beats instead of width to judge interval identity (Hall & Hess, 1984). Ascending intervals were therefore selected as the most likely to promote comparison of musical interval sounds.

The range of tested stimuli allowed a balanced comparison among interval categories, although different interval categories may not be equally salient. Certain intervals are privileged in musical experience and may be better learned, leading to a stronger categorical representation. That is, an interval may be more or less "stable," depending on how smoothly or roughly its component pitches harmonize with each other, and stable intervals enjoy a privileged status in music perception (Krumhansl & Keil, 1982). Stable intervals are sounded more frequently than unstable intervals, and tend to appear in more prominent positions, such as the beginning or end of a melody, making them more easily recognized and learned. Of the intervals featured in the present design, a perfect fourth is considered stable; a major third is also stable, but less so than a perfect fourth; and a minor third is considered less stable than a major third. If interval stability corresponds to category strength, then discrimination performance may be different among these categories. Additionally, within the range of stimuli tested in the present experiment, half of the stimuli were flat relative to their category, and half were sharp. Therefore the overall range tested included two *sharp* areas and two *flat* areas, and musically trained listeners are more able to judge small differences between flat intervals than sharp ones (Schellenberg, 2002). Therefore, although three named intervals were tested, the present design featured four recognizably different categories: minor third *sharp*, major third *flat*, major third *sharp*, and perfect fourth *flat*.

Conflict between AP and RP categories was introduced by shifting stimulus frequencies upward by 25 cents (an "eighth tone"). In standard musical performance, an interval is formed by two pitches located at their standard (prototypical) absolute frequency values. Shifting the pitches' absolute frequency values upwards introduces a conflict between the interval category, as indicated by the width between pitches, and the pitch categories, as indicated by the absolute pitch frequencies (Figure 2). Frequencies that are *near* to an interval prototype become *far* from a pitch prototype, and vice versa. If listeners find pitch categories more salient than interval categories, then their discrimination accuracy will be different for shifted intervals than unshifted because the distances from the pitch prototype will be different.

"Incorrect" absolute pitches do not prevent absolute listeners from correctly identifying intervals (Benguerel & Westdal, 1991), but can slow their judgment (Miyazaki, 1992), possibly due to a "Stroop-like" effect (Miyazaki, 2004b) in which a pitch frequency automatically activates

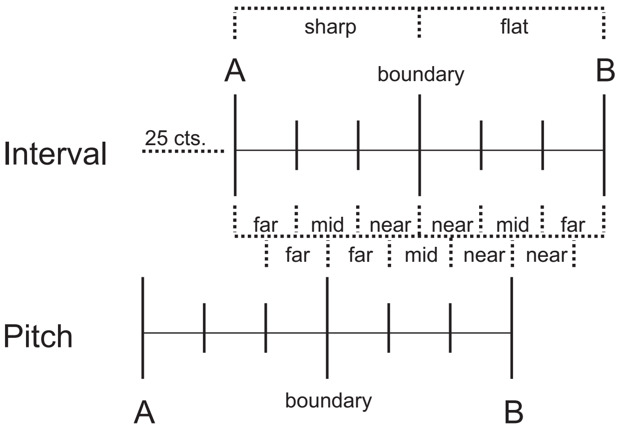

FIGURE 2. Intervals were shifted by 25 cents in half the trials, causing a conflict between interval and pitch categories.

its categorical label, thus interfering with other labeling tasks (Arao, Itoh, Suwazono, Nakada, & Miyazaki, 2002). However, pitch labeling interference may be restricted principally to musicians trained in Asian-style "fixed-do" traditions rather than Western-style "movable-do" (Hseih & Saberi, 2008), and interference appears to decrease with increasing musical skill (Ikeda, 2010); therefore, we do not expect that absolute listeners in the current experiment will find it difficult to identify intervals when stimuli are shifted. Rather, because of the conflict introduced by shifting, a discrimination task should indicate whether pitch or interval categories are the more salient.

A third possibility was that listeners might remember the middle tone's fundamental frequency, ignoring ascending interval sounds and disregarding pitch classes, to make a direct comparison of one frequency to another. The present design accommodated this possibility. If listeners did make direct comparisons between absolute vibratory frequencies, without regard to interval or pitch class, those listeners would be unaffected by any aspect of the present design. That is, such listeners would demonstrate no categorical perception for intervals, and no effect of pitch conflict, showing instead consistent performance across all stimuli. If listeners were unaffected by the present design, this would be interpreted as a demonstration of absolute judgment being made preferentially to relative perception.

ADDITIONAL FACTORS

Additional considerations of music perception guided our design. For one, absolute listeners may demonstrate an advantage for identifying intervals whose bottom tone is C (Miyazaki, 1995). This advantage may also be attributable to "fixed-do" tradition, which inculcates C-major as a privileged tonality (Gregersen, Kowalsky, & Li, 2007; Miyazaki & Ogawa, 2006). This potential advantage was controlled by varying the bottom tone for each trial among eight different pitches, a manipulation that has no effect on relative listeners (Siegel & Siegel 1977a). If absolute listeners are more capable of identifying intervals whose bottom tone is an accurately tuned C, this advantage would be applicable only in a minority of all trials. Two other features of our design—ascending intervals and order of presentation—were more likely to have an overall effect, so these features were controlled by incorporating the factors upon which they could potentially exert an influence.

Ascending intervals may distort interval widths to make them seem wider, a phenomenon referred to as "perceptual enlargement" (Hartmann, 1993; Russo & Thompson, 2005). If perceptual enlargement occurred here, it would cause categorical boundaries to shift upwards, distorting the interval widths perceived around each boundary. This is controlled in our design by the width of the overall range tested. Enlargement is proportional to interval width (Rakowski, 1976; Vurma & Ross, 2006), so our range features three different intervals (m3, M3, P4). If enlargement distorted categorical boundaries, judgment accuracy would be different for each interval at *far* distances. Furthermore, within the range tested, stimuli are at different heights relative to their category; half the stimuli are situated below a prototype (*flat*) and half are above (*sharp*). Enlargement would cause flat sounds to seem nearer a prototype and sharp sounds to seem further, making judgment of *far* distances different between flat and sharp conditions. If enlargement were to occur in the current design, it would manifest as differences in performance at *far* distances for each interval. If enlargement did not occur, no such differences would be observed.

Stimulus order of presentation is likely to influence perception. When a stimulus moves away from a prototype, the difference is easier to detect than an equivalent change in the opposite direction. This asymmetry occurs in multiple domains (Bharucha & Pryor, 1986; Hanley & Roberson, 2011) and has been observed for musical intervals (Schellenberg & Trehub, 1994). In musical terms, a prototype musical sound is considered "well-tuned," and moving away from a prototype represents a change from *better* tuning to *worse* tuning. In the present (XAB) design, discrimination is expected to be less accurate when the standard (X) is worse tuned (Schellenberg, 2002), because the change from X to A presents movement toward a prototype. Although the

TABLE 1. *Independent Variables Featured In Experimental Design and Used as ANOVA Factors.*

Design	Factor	Levels	Measure
Within-subjects	Distance	near, middle, far	Distance from interval prototype
	Interval	m3, M3f, M3s, P4	Interval category label
	Shift	unshifted, shifted	Stimuli shifted upward 25 cents
	Tuning	better, worse	Standard closer to or further from prototype
Between-subjects	Group	AL, RL, NL, PL	Labeling accuracy

subsequent change from A to B would present a movement away from the prototype, and be easier to detect, detecting that change would serve to reinforce the mistaken judgment that X and A had been identical. Research with infants shows that asymmetry for musical tuning is innate, not learned (Schellenberg & Trehub, 1996); therefore asymmetry may exert an influence on all participants regardless of musical ability.

Participants could become fatigued. The procedure was expected to take approximately three hours, in a single session, and the repetitive task could induce fatigue. To help mitigate this possibility, participants were allowed to take breaks. Additionally, results were analyzed to determine whether participants were subject to fatigue.

The current design included a total of five factors (see Table 1). Factor one, *distance*, represented different regions of each interval category relative to a prototype: *near* included the prototype; *middle* was equidistant from prototype and category boundary; *far* included the category boundary. Factor two, *interval*, indicated the three featured categories of m3, M3-flat, M3-sharp, and P4. Factor three, *shift*, designated whether stimuli frequencies were *unshifted* and thus congruent with pitch categories, or *shifted* and in conflict. Factor four, *tuning*, acknowledges a natural asymmetry in discrimination for standards that are *better* or *worse* tuned. Factor five is the between-groups factor, *group*, organized by labeling ability into AL, RL, PL, and NL. Our predictions were made with respect to these factors.

PREDICTIONS

Most important to CP is *distance*. CP for intervals would be indicated by better discrimination at *far* distances compared to both *middle* and *near*. *Shift* creates a conflict between AP and RP categories, such that *middle* judgments for intervals are *far* judgments for pitch (see Figure 2). If listeners are influenced by absolute pitch categories when discriminating musical intervals, performance for shifted stimuli would instead be best at *middle* distance and decline at *far*. The greatest accuracy would be predicted at the middle distance in sharp areas—which, when shifted, is within-category for interval but cross-category for pitch. If intervals are not influenced by pitch categories, then shifting pitch values would have no effect.

If a prototype effect were observed, it would be seen at the *near* distance. Because a prototype can assimilate stimuli within a 25-cent range in either direction, and *near* comparisons are made here at a distance of 16⅔ cents from the prototype, these stimuli could be assimilated by the prototype; a magnet effect would therefore result in the lowest level of discrimination accuracy for *near* comparisons. An anchor effect, by contrast, would show the highest level of discrimination accuracy for *near* comparisons.

Our primary interest was to identify differences observed among groups. Non-labelers were not expected to show CP for either intervals or pitches, and their performance should therefore control for any effects of natural perception. Relative listeners should show the influence of music training on interval discrimination. Any differences observed between these groups and absolute labelers should, therefore, demonstrate whether AP exhibits an independent influence on relative perception of intervals.

Method

PARTICIPANTS

Forty-one participants, age 16-39 ($M = 22.14$), were recruited with flyers placed in the Indiana University Jacobs School of Music and around campus. Participant groups included absolute listeners (AL), partial absolute listeners (PL) relative listeners (RL), and non-labelers (NL). Fourteen participants self-identified as AL, 13 as RL, and 14 as nonmusicians. All participants were paid $50 for their participation.

To classify participant groups, participants named musical tones and intervals. Participants were asked to name the pitch class of 24 tones from C3 (130.81 Hz) to B4 (493.88 Hz), randomly ordered. Participants were then asked to name 24 intervals, comprising two instances each of the 12 musical intervals from minor

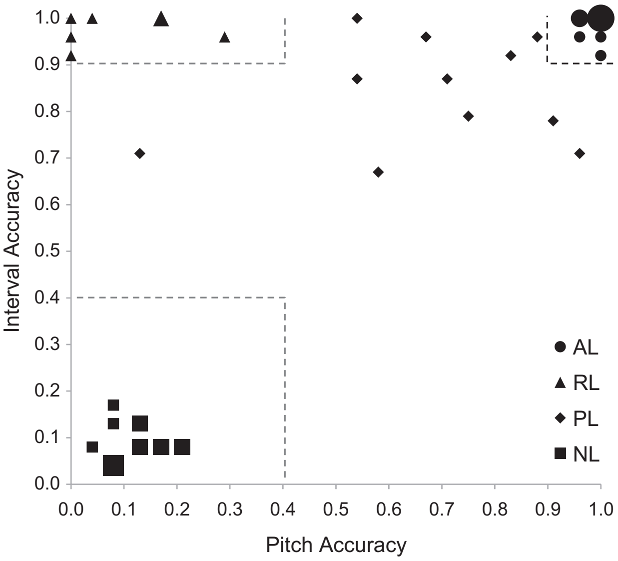

FIGURE 3. Labeling accuracy for all participants. Data points represent one to four participants depending on size.

second (m2) to octave (P8), also in random order, whose tones were randomized within the range of C3 to B4. Participants with 90% or greater accuracy for musical pitches were classified as AL. Participants scoring between 40% and 90% accuracy for pitches were classified as PL. Participants with 90% or greater accuracy for musical intervals, but who did not achieve 40% accuracy for tones, were classified as RL. Participants who failed to achieve either accuracy level were classified as NL.

Participants were classified by demonstrated accuracy, not by self-report (Figure 3). Thus one self-identified RL was classified AL; four self-identified AL were classified PL; one self-identified RL was classified PL; and one self-identified RL was classified NL. All nonmusicians accurately self-identified as NL; neither their judgment of pitch (mean = .16, SD = .13) nor intervals (mean .13, SD = .16) were different from chance (.08).

The assumption that PL performance may be due to weaker category representations was tested by evaluating "semitone errors"; i.e., errors that occur from selecting a neighboring pitch category. Among the eight participants initially classified as PL, 62% of reported errors were semitone errors. If semitone errors had been counted as accurate judgments, five of these eight participants would have scored above 90%. This result supported the assumption of weaker categories; and, in doing so, recommended the classification of three additional participants as PL. Two participants had demonstrated greater than 90% accuracy for pitch labeling, but less than 90% for intervals; a third had achieved neither 40% accuracy for pitches nor 90% accuracy for intervals, but had achieved greater than 40% for intervals. These three participants were reclassified as partial *relative* listeners and included in the PL group.

Thus participants were classified as 14 NL (5 male, 9 female), 7 RL (4 male, 3 female), 9 AL (6 male, 3 female), and 11 PL (5 male, 6 female). Each participant's primary instrument was noted; some indicated an ability to play multiple instruments. AL comprised 8 pianists and 1 violinist. RL comprised 2 pianists and 1 each of specialists in trumpet, violin, french horn, cello, and double bass. PL comprised 6 pianists, 2 vocalists, and 1 each of specialists in percussion, cello, and viola. Formal music training began, on average, for AL at age 6.44 (median = 7.0, SD = 2.19), giving 16.0 years of experience (median = 14.0, SD = 6.74); for RL at age 4.71 (median = 5.0, SD = 0.76), giving 16.86 years of experience (median = 17.0, SD = 1.95); for PL at age 7.8 (median = 7.5, SD = 3.88), giving 14.7 years of experience (median = 16, SD = 7.51). The self-reported RL classified as NL had received no music training, but had played guitar for 8 years starting at age 12.

STIMULI

Each stimulus presented three musical tones played in ascending sequence. These tones were complex tones, rather than pure tones, and were generated using QuickTime MIDI synthesis with the instrument set to "Grand Piano." Each tone was played for a duration of 450 ms and then silenced. There was no pause between tones.

The first tone was selected from among the eight tones of the "middle octave." The middle octave comprises F3, G3, A3, B3, C4, D4, E4, and F4. A 676-item list of these tones was generated and randomized prior to the experiment (i.e., 84 iterations of all 8 tones, plus one more of the 4 tones F3–B3). All experimental sessions followed this list in selecting the first tone of each trial.

The second tone was placed at a variable position above the first. The tone's position varied within the range described by the interval categories *minor third* (m3), *major third* (M3), and *perfect fourth* (P4), encompassing a width of 3 to 5 semitones, or 300 to 500 cents, as measured from the first tone. This 200-cent range was divided into equally spaced positions (16⅔ cents width). Therefore, each position was located at one of four distances relative to an interval's prototype, or categorical center: 0*d*, 1*d*, 2*d*, or 3*d*, where 0*d* represented the category center and 3*d* the category boundary.

There were thus a total of 13 positions available to select for the second tone.

The third tone always formed a "perfect fifth" interval above the first. The tone was therefore selected among C4, D4, E4, F#4, G4, A4, B4, and C5, dependent on the pitch class of the first tone. Together, the sequence of three tones in each stimulus formed a *trichord*.

Stimuli were presented as either *shifted* or *unshifted*. *Unshifted* stimuli conformed to the frequencies of the standard musical scale. *Shifted* stimuli were transposed an eighth-tone (25 cents) higher. When transposed, stimuli retained their interval relationships, but their tone frequencies no longer corresponded to standard pitch values.

APPARATUS

The procedure was controlled with a custom interface programmed in Realbasic and presented on a Macintosh G4 Powerbook. The interface provided instructions, presented stimuli, and recorded responses. Stimuli were presented over headphones (Sony MDR-7506). Participants sat in a private room.

PROCEDURE

Each trial presented three stimuli in an XAB design. Listeners were presented with a standard stimulus (X) followed by three seconds of silence. Two comparison stimuli (A and B) were then presented in sequence with one second of silence between each stimulus. Participants indicated which of A or B was identical to X by clicking an on-screen button or pressing the corresponding letter on the keyboard (A or B).

In each trial, only the second tones of the stimuli were changed. All three stimuli used the same first tone, selected from the pre-randomized list already described, and the same third tone, always a perfect fifth above the first. The second tone of X was randomized at runtime among the 13 positions within the range of interest. One of either A or B was identical to X; the other was a *foil,* whose second tone was 1*d* above or below the second tone of X. Participants' task in each trial was to detect a change in the second tone.

Four independent variables were present in each trial. Stimuli could be either *shifted* or *unshifted*. The second tone of each standard fell within an *interval* category (m3, M3f, M3s, P4). Each trial represented a certain *distance* from the prototype of the featured interval; standard–foil comparisons could be *near* the prototype (0*d*–1*d*), in the *middle* of the category (1*d*–2*d*), or *far* from the prototype (1*d*–2*d*). Musical tones become better *tuned* as they approach a prototype; therefore, in each comparison, the standard stimulus X exhibited either *better* or *worse* tuning than the comparison stimulus. Thus the four within-subjects factors were *interval, distance, height*, and *tuning*.

There were a grand total of 676 trials. Each trial tested one of the three distances (near, middle, far). Each distance was presented 13 times in each shift condition (shifted, unshifted) for each tuning condition (better, worse) in each of the interval areas (m3, M3f, M3s, P4). Of each set of 13 presentations, 6 were "no change" trials (X=A) and 7 were "change" trials (X=B). In a discrimination task, participants may recognize end stimuli and respond differently to those stimuli (Eriksen & Hake, 1957). To avoid end effects, two *near* distances were added above and below the primary range; these were also presented 13 times for each condition of shift and tuning, for a total of 52 trials, but these trials were excluded from analysis. Thus the number of trials presented were 3 distances × 13 presentations × 2 shifted conditions × 4 height areas × 2 tuning conditions = 624 + 52 end trials = 676 trials.

The experimental session lasted approximately three hours. Participants were encouraged to take breaks by clicking a "Take a Break" button which was visible between trials. The length of a break was left to participants' discretion, but no break lasted longer than 20 minutes, and participants did not leave the immediate area. The computer interface provided a progress bar indicating the proportion of trials remaining in the experiment. At the conclusion of the experiment, participants were given the option to provide a narrative description of their strategy in performing the task; 38 of the 41 participants elected to do so, but no consistent patterns, with respect to strategy, were observed from these descriptions.

Results

Scores were calculated for each participant as a percentage of correct responses. These scores were used in all tests. All ANOVAs performed were repeated-measures, and all *t*-tests were two-tailed paired samples, except where otherwise indicated.

To avoid end effects, trials involving the highest and lowest tones were excluded. These two tones, outside the range of interest, had been presented 26 times, thereby excluding 52 trials from analysis. A total of 624 trials were analyzed.

Labeler groups were not different from each other, indicating that AP does not exhibit an independent influence on the relative perception of intervals. A five-factor mixed-design ANOVA (4 × 2 × 4 × 3 × 2) was performed with between-subjects factor *group* (absolute labeler, relative labeler, partial labeler, non-labeler) and within-subjects factors *shift* (shifted, unshifted), *interval* (m3, M3-flat, M3-sharp, P4), *distance* (near, middle, far), and *tuning* (better, worse). Labelers and non-labelers performed differently, indicated by a main effect of *group*, $F(3, 37) = 25.65$, $p < .01$, $\eta^2 = .68$. To determine whether this effect was driven by the difference between musicians and nonmusicians, the same ANOVA was repeated, excluding data from non-labelers. This test also showed a main effect of *group*, $F(2, 24) = 4.08$, $p = .03$, $\eta^2 = .25$, but *group* did not interact with any other factor. The main effect of *group* indicated that PL had, overall, performed less accurately than AL (PL, $M = .60$, $SD = .03$; AL, $M = .65$, $SD = .06$); however, this difference was not supported by an independent-samples *t*-test corrected for multiple comparisons, $t(18) = 2.50$, *ns*, and RL performance was not different from PL or AL (RL, $M = .62$, $SD = .03$). Data were therefore collapsed into two groups, *labelers* and *non-labelers*, for subsequent tests. This means that AL, RL, and PL were analyzed as a single group.

Non-labelers were affected only by *tuning*. To determine whether non-labelers were affected by any musical factors, the five-factor mixed-design ANOVA was repeated, but with only two levels of *group* (labeler, non-labelers). A difference between labelers and non-labelers was supported by a main effect of *group*, $F(1, 39) = 55.87$, $p < .01$, $\eta^2 = .59$, so a four-factor ANOVA was performed on non-labeler data only (the same test, removing the factor *group*). Only one effect was observed for non-labelers. Non-labelers were better able to recognize a difference moving away from a prototype, as indicated by a main effect of *tuning*, $F(1, 13) = 18.29$, $p < .01$, $\eta^2 = .59$. No other effects or interactions were observed. Subsequent tests were performed on labeler data only, and any effects not related to *tuning* could be attributed to musical experience.

Labeler data were analyzed by performing a four-factor within-subjects ANOVA ($2 \times 4 \times 3 \times 2$) using factors *shift, interval, distance,* and *tuning*. As has been mentioned, these data collapsed all labeling groups together, comprising AL, RL, and PL; therefore, any effects produced by *shift* would indicate all labelers' sensitivity to absolute pitch categories, and any effects produced by *interval* and *distance* would indicate all labelers' sensitivity to relative pitch categories. Effects related to *tuning* would reflect natural tuning asymmetry, as demonstrated by non-labeling participants. The results reported below were all drawn from this four-factor ANOVA. All *t*-tests performed were paired samples *t*-tests, corrected for multiple comparisons where necessary.

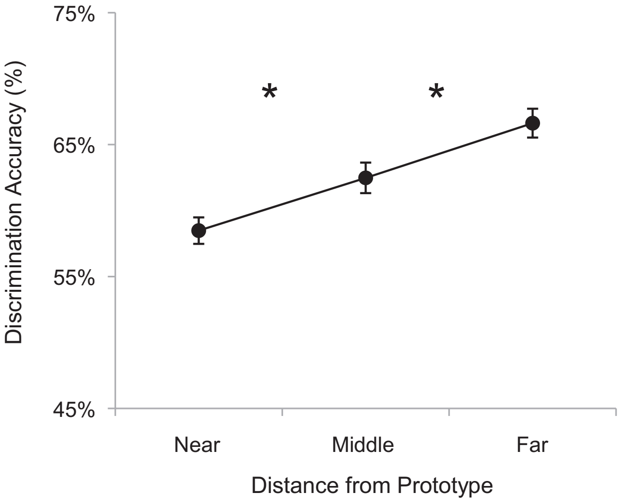

FIGURE 4. Main effect of *distance*, showing categorical perception of intervals. Significant differences are marked with asterisks. Error bars are standard error of the mean.

Labelers exhibited categorical perception for intervals. Discrimination accuracy improved as stimuli moved toward a category boundary (Figure 4), as indicated by a main effect of *distance*, $F(2, 52) = 36.22$, $p < .01$, $\eta^2 = .58$. Differences between distances were confirmed by *t*-tests: between *near* and *middle*, $t(26) = -4.11$, $p < .01$; and between *middle* and *far*, $t(26) = -5.28$, $p < .01$. No interval enlargement was observed, which would have manifested as a two-way interaction between *distance* and *interval*. The increase of discrimination accuracy as stimuli moved toward a category boundary confirmed categorical perception of intervals.

Category prototypes behaved like magnets rather than anchors. An anchor effect would have produced greatest discrimination near a prototype, and this was not observed in any condition. Instead, discrimination was least accurate near a prototype and increased as stimuli moved toward a category boundary.

Discrimination accuracy was equivalent for all intervals except the major third flat area (M3f), for which discrimination was more accurate, as indicated by a main effect of *interval*, $F(3, 78) = 14.00$, $p < .01$, $\eta^2 = .35$. This result was supported by *t*-tests comparing M3f to the other three interval areas (m3, M3s, P4), which were, in turn, not different from each other. Greater accuracy for M3f is likely a consequence of musically trained listeners' greater accuracy for flat sounds (Schellenberg, 2002) as well as the privileged status of major third intervals in musical experience (Krumhansl & Keil, 1982).

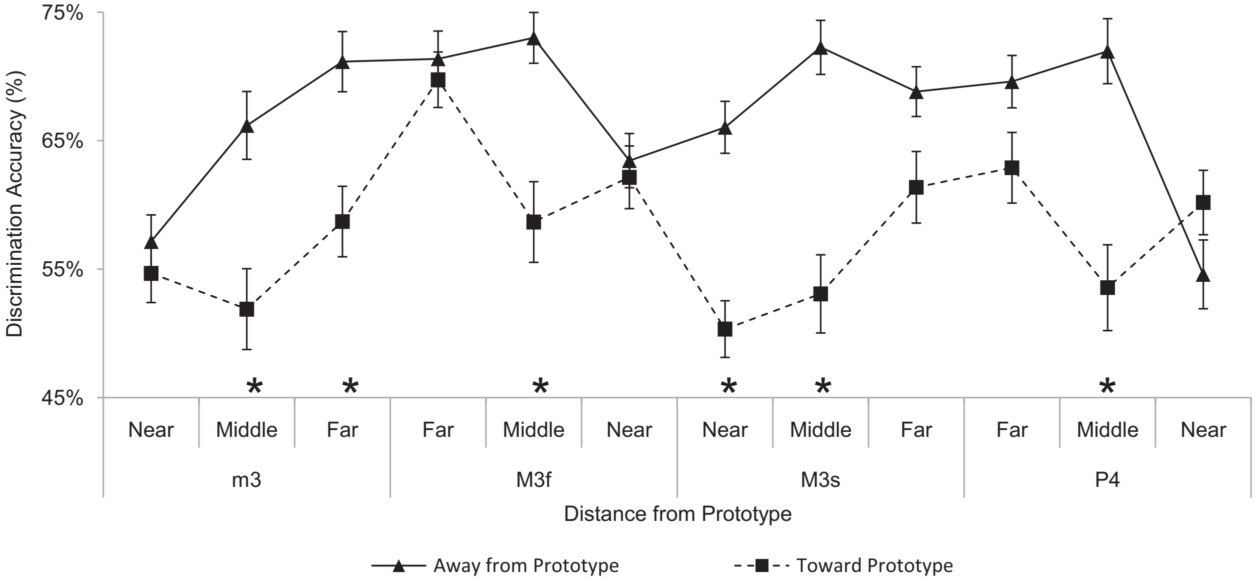

FIGURE 5. Tuning asymmetry for all four interval areas, labeler data only. Asymmetry affected *middle* distance stimuli as well as stimuli bordering the major third flat area (i.e., m3 *far*, M3s *near*). Significant differences are marked with asterisks. Error bars are standard error of the mean.

Labelers were susceptible to natural tuning asymmetry, as indicated by a main effect of *tuning*, $F(1, 26) = 11.76$, $p < .01$, $\eta^2 = .31$. Judgments performed in the major third sharp area (M3s) were most benefited by stimuli moving away from a prototype, as indicated by an interaction between *tuning* and *interval*, $F(3, 78) = 3.45$, $p = .02$, $\eta^2 = .12$; this result may also be attributable to the major third interval's privileged status. *Near* and *far* distances were generally unaffected by tuning asymmetry, with a consistent effect appearing only at the *middle* distance, as indicated by an interaction between *tuning* and *distance*, $F(2, 52) = 14.74$, $p < .01$, $\eta^2 = .36$. This may be because asymmetry arises from hearing stimuli "going out of tune," or becoming worse exemplars of their category. If *near* stimuli, being proximal to the prototype, were perceived as equally "good" representatives of their category, and *far* stimuli, being at the category boundary, were perceived as equally "poor," then at both *near* and *far* distances, these within-category movements might not produce a change in "goodness" substantial enough to generate asymmetry. If the magnitude of asymmetry is determined by each stimulus' perceived goodness (Schellenberg, 2002), this could also explain why *near* and *far* distances were affected by asymmetry when they bordered the major third flat area (M3f), as indicated by a three-way interaction between *tuning*, *distance*, and *interval*, $F(6, 156) = 3.96$, $p < .01$, $\eta^2 = .13$ (Figure 5). The major third interval is privileged in Western music because of its "good" sound versus other intervals, and discrimination here was most accurate in the M3f area, suggesting that M3f may be the more privileged than major third sharp (M3s). In each trial featuring minor-third *far* and major third sharp *near*, a stimulus was presented at the M3f border. These boundary stimuli might have been perceived as crossing into the privileged M3f category, and therefore be judged to sound more "good" than comparison stimuli, creating a strong asymmetry.

Labelers were not susceptible to shifted frequencies. Frequencies were shifted to introduce conflict between absolute pitch and relative interval categories, such that a stimulus' distance from a pitch prototype was different from its distance to an interval prototype. An effect of this conflict would therefore be represented by increased discrimination accuracy, when shifted, for *middle* distance stimuli versus *far* distance stimuli, supported by an interaction between *shift* and *distance*. The two-way interaction between *shift* and *distance* was not significant, $F(2, 52) = 0.88$, *ns*, $\eta^2 = .03$. *Shift* did show interactions with the other three factors: *shift* and *interval*, $F(3, 78) = 2.83$, $p = .04$, $\eta^2 = .10$; *shift*, *interval*, and *tuning*, $F(3, 78) = 3.35$, $p = .02$, $\eta^2 = .11$; and *shift*, *interval*, *tuning*, and *distance*, $F(6, 156) = 3.38$, $= .004$, $\eta^2 = .12$. However, simple effects analyses performed on each level of *interval*, using factors *shift*, *tuning*, and *distance*, failed to support these interactions. This was not surprising, because any effect of *shift* represents sensitivity to absolute pitch categories, and the labeler group included relative listeners. It is improbable that relative listeners, who cannot label absolute pitches and do not demonstrate categorical perception for absolute pitch, would have demonstrated sensitivity to absolute pitch categories. Nonetheless, the presence of interactions for *shift* made it necessary to address our *a priori* assumption that absolute listeners and relative listeners would respond differently to absolute pitch categories. Therefore, simple effects analyses were performed on each labeler group, looking particularly for effects of *shift*.

Absolute listeners were not affected by shifted stimuli. Simple-effects ANOVAs were performed on each labeler group, using factors *shift*, *distance*, *interval,* and *tuning*, correcting for multiple comparisons ($\alpha = .016$). No effect or interaction for *shift* was observed within the RL group. PL showed a four-way interaction among all four factors, $F(6, 48) = 3.91$, $p < .01$, $\eta^2 = .33$; a simple-effects ANOVA was subsequently performed on each level of *interval* ($\alpha = .0125$), using factors *shift*, *distance*, and *tuning*. The only interaction observed from these tests was a three-way interaction within major third flat (M3f), $F(2, 16) = 6.51$, $p = .009$, $\eta^2 = .45$, which indicated that tuning asymmetry could be reversed for unshifted *far* stimuli; that is, unshifted stimuli moving toward the prototype would be easier to discriminate than unshifted stimuli moving away. This indicates PL were not affected by shifting. Not only did this interaction implicate unshifted stimuli rather than shifted, but this difference was also not supported by a paired samples *t*-test ($\alpha = .016$). AL did not show an interaction between *distance* and *tuning*. Rather, AL showed an interaction between *shift* and *tuning*, $F(1, 10) = 9.17$, $p = .013$, $\eta^2 = .48$, which indicated a lesser magnitude of tuning asymmetry for shifted than unshifted stimuli. This interaction was not supported by paired samples *t*-tests, which showed no differences among tuning conditions. It may be noted that, even if this difference had been supported, it would not have represented a conflict between interval and pitch categories. Rather, it would have suggested that absolute listeners were sensitive to stimuli moving toward or away from a pitch prototype, even though they were performing judgments on interval sounds. From these results, it may be concluded that all listeners were indifferent to shifted stimulus frequencies. Absolute listeners performed judgments on interval categories, and did not infer their interval judgments from absolute pitch categories.

Lastly, RL was the only group subject to fatigue. Each participant's data were organized into six chronological segments, with each segment representing approximately half an hour (104 trials). A mixed-design ANOVA was performed on these data, with between-subjects factor *group* (NL, RL, AL, PL) and within-subjects factor *segment* (6 segments). RL became fatigued over the course of the session, whereas the other groups did not, supported by an interaction between *group* and *segment*, $F(15, 185) = 1.72$, $p = .05$, $\eta^2 = .12$. RL performance declines significantly in the first two segments and then does not recover (Figure 6), making RL the only group whose performance in Segment 6 is significantly lesser than in Segment 1, $t(6) = 2.70$, $p = .04$.

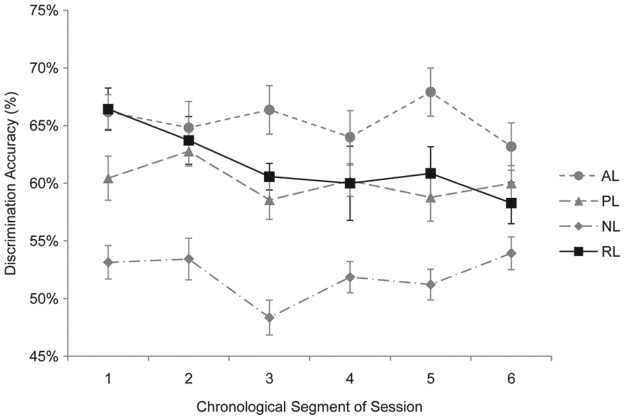

FIGURE 6. Participant fatigue. Each data point represents average overall accuracy for that half hour segment of the overall session. Each line represents a different participant group. Error bars are standard error of the mean.

Discussion

Listeners were tested for categorical perception (CP) of musical intervals via a discrimination task that presented conflicting relative pitch and absolute pitch values. Listeners were either nonmusicians who were unable to identify musical sounds (NL), or were trained musicians who could label tones absolutely (AL), relatively (RL), or with partial accuracy (PL). NL did not demonstrate CP, but exhibited natural sensitivity to well-tuned interval sounds. AL, RL, and PL did not perform differently from each other; all three groups showed CP for intervals and demonstrated relative judgment unaffected by absolute pitch identities.

All musicians exhibited categorical perception (CP) for intervals. Discrimination accuracy was weakest near each interval's prototype and increased as stimuli moved toward the category border, which is a result evidencing categorical perception. Absolute listeners (AL) did not perform differently in this relative listening task, despite the fact that interval and pitch categories were in conflict for half the trials. Pitch and interval categories were placed in conflict by shifting absolute frequencies, such that each stimulus was a different distance from a pitch prototype than from an interval prototype. If AL had been unable to perceive relative pitch categories, or if absolute pitch categories were perceptually dominant versus relative pitch sounds, then AL discrimination accuracy for shifted trials would have represented distance from an absolute pitch prototype, and this did not occur. Moreover, if AL had performed the task by remembering and comparing absolute pitch frequencies, discrimination accuracy would have been equivalent for all stimuli, regardless of shifting or distance, and this did not occur. Instead, AL accuracy for shifted stimuli was affected by distance from an interval prototype, demonstrating CP for intervals. It cannot be ruled out that AL may have used absolute pitch categories to perform judgments—perhaps in only a proportion of trials—but such an account would leave unexplained why interval-CP effects are found in shifted stimuli. The current results therefore indicate that musicians with absolute pitch do perceive relative sounds directly, and not by inference.

Interval category prototypes exhibited what could be interpreted as a magnet effect, because discrimination was least accurate for stimuli nearest a prototype. The present results therefore do not support musical interval prototypes as anchors, which would predict greatest discrimination accuracy near a prototype. However, a magnet effect makes no prediction for increased accuracy at a category border (Iverson & Kuhl, 2000), as was observed in the present experiment. The present results are therefore more indicative of categorical perception.

Discrimination accuracy was greatest for the major third flat area (M3f), but only for musicians, indicating an effect of music training. This result likely reflects the major third interval's privileged status in music, combined with musicians' orchestral experience, which inculcates a bias to value sharp sounds as qualitatively better than flat (Geringer, 1976). That is, from their experience playing in groups, musicians may become particularly attentive to detecting and correcting less desirable flat sounds, and to be less exacting with more desirable sharp sounds. This tendency, combined with greater experience with the major third interval, would suggest a stronger internal representation of the major third flat area, contributing to more-accurate discrimination in this area.

The order of presentation effect was also affected by music training. Differences between musical intervals are easier to detect when they go out of tune rather than in tune (Schellenberg, 2002). That is, a change is easier to detect when it becomes qualitatively worse. This asymmetry in judgment is natural, as it is evidenced by nonmusicians as well as musicians. However, music training affected this asymmetry. Musicians did not exhibit asymmetry at categorical boundaries or near prototypes, indicating that listeners had become sensitive to these categories and therefore learned to judge these stimuli as equally good. The area bordering the major third flat area, from m3 *far* to M3s *near*, although near a category boundary and prototype, respectively, did exhibit asymmetry, but this asymmetry may have arisen from musicians' perception of sounds moving out of or into the privileged major third flat category. Tuning asymmetries, therefore, provided further evidence that musical listeners, including those with absolute pitch, made judgments based on category membership.

Musicians with absolute pitch did not show a deficit in relative perception; nor did they show a bias for absolute perception. These data argue against the supposition that absolute pitch perception interferes with the deployment of relative pitch perception, or that musicians with absolute pitch abilities are inherently less able to perform relative pitch tasks. Rather, absolute listeners, early in their musicianship, may develop listening and reading strategies based on their absolute pitch abilities. Subsequently, when solving musical problems whose solutions are more efficient when accomplished with relative pitch, it may be easier to apply habitual strategies than to learn a new approach. Alternatively, as asserted by Dooley and Deutsch (2011), previous experiments purporting to demonstrate interfering effects of absolute pitch on musicianship may be attributable to "artificial" laboratory conditions and procedures, rather than natural perception and typical musical tasks. To more fully explore this hypothesis, the same absolute listeners could be subjected both to musical tasks at which they may excel (e.g., Dooley & Deutsch, 2010, 2011) and laboratory tests designed to confound absolute listeners (see Miyazaki, 2004a). The results of the present experiment do not address how absolute pitch ability may relate to musicianship, but argue against the assertion that absolute pitch ability is detrimental to musicians' relative pitch perception.

That relative pitch musicians became fatigued, and musicians with absolute pitch labeling ability did not, supports the assertion of Dooley and Deutsch (2011) that we have more to understand about the positive effects of absolute pitch ability on musicianship. After testing sessions were complete, participants' verbal comments suggested, anecdotally, that relative listeners found the task difficult and frustrating, whereas absolute listeners viewed the experience as more of a fun challenge.

However, verbal comments were not formally recorded or coded; and written comments, because they described strategy and not physical experience, could not corroborate the impression suggested by verbal comments. Future investigations of absolute pitch ability may find it beneficial to analyze not only participants' performance in the musical task being tested, but also participants' subjective impression of that task's difficulty.

Further research may also be desirable to investigate the effect of shifted stimuli that was observed when AL were tested separately. Shifting stimuli did not affect AL categorical perception of intervals, but AL appeared to be less susceptible to asymmetry when stimuli were shifted, suggesting that their perception of stimulus "goodness" might have been influenced by pitch prototypes. Alternatively, because a simple effects test showed no differences among any conditions, AL might not be affected by tuning asymmetry. These tests may have failed to find differences because the current design did not allow sufficient power. Nonetheless, if AL do not experience tuning asymmetry, then that would raise the possibility that AL do not base their judgments on "goodness" of sound, but could instead make objective judgments of interval width. These hypotheses cannot be tested with the current data, but may be worthy questions for later investigation.

Conclusion

The current study provides evidence that absolute listeners (AL) perceive musical intervals categorically, and that they perform judgments of musical intervals independently of absolute pitch categories. This indicates that the relative interval between musical tones is a salient concept in AL musicianship, and therefore implies that previous observations of absolute pitch as a musical "inability," such as difficulty in recognizing transposed melodies (Miyazaki 1993), are not caused by AL musicians having a weaker ability to perceive relative sounds, but should be attributed to AL musicians using inappropriate absolute pitch strategies to achieve the relative pitch task at hand.

Author Note

This research was supported in part by National Science Foundation REESE grant 0910218 and Department of Education IES grant R305A1100060. David J.D. Earn is supported by NSERC.

Correspondence concerning this article should be addressed to Christopher Aruffo, Department of Psychology, Neuroscience, & Behaviour, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada. E-mail: acoustic@aruffo.com

References

ACKER, B. E., & PASTORE, R. E. (1996). Perceptual integrality of major chord components. *Perception and Psychophysics*, *58*, 748-761. doi: 10.3758/BF03213107

ACKER, B. E., PASTORE, R. E., & HALL, M. D. (1995). Within-category discrimination of musical chords: Perceptual magnet or anchor? *Perception and Psychophysics*, *57*, 863-874. doi: 10.3758/BF03206801

ARAO, H., ITOH, K., SUWAZONO, S., NAKADA, T., & MIYAZAKI, K. (2002). Auditory Stroop interference induced by sung syllables. *IEIC Technical Report* (*Institute of Electronics, Information and Communication Engineers)*, *102*, 1-6.

BACHEM, A. (1937). Various types of absolute pitch. *Journal of the Acoustical Society of America*, *9*, 146-151. doi: 10.1121/1.1915919

BARRETT, S. (1999). The perceptual magnet effect is not specific to speech prototypes: New evidence from music categories. *Speech, Hearing and Language: Work in Progress*, *11*, 1-16.

BENGUEREL, A. P., & WESTDAL, C. (1991). Absolute pitch and the perception of sequential musical intervals. *Music Perception*, *9*, 105-119. doi: 10.2307/40286161

BERMUDEZ, P., & ZATORRE, R. J. (2009). A distribution of absolute pitch ability as revealed by computerized testing. *Music Perception*, *27*, 89-101. doi: 10.1525/mp.2009.27.2.89

BHARUCHA, J. J., & PRYOR, J. H. (1986). Disrupting the isochrony underlying rhythm: An asymmetry in discrimination. *Perception and Psychophysics*, *40*, 137-141. doi: 10.3758/BF03203008

BRADY, P. T. (1970). Fixed-scale mechanism of absolute pitch. *Journal of the Acoustical Society of America*, *48*, 883-887. doi: 10.1121/1.1912227

BRAIDA, L. D. (1984). Intensity perception. XIII. Perceptual anchor model of context-coding. *Journal of the Acoustical Society of America*, *76*, 722-731. doi: 10.1121/1.391258

BURNS, E. M. (1999). Intervals, scales, and tuning. In D. Deutsch (Ed.), *Psychology of music* (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. doi: 10.1121/1.411447

Burns, E. M., & Ward, W. D. (1978). Categorical perception—phenomenon or epiphenomenon: Evidence from experiments in the perception of melodic musical intervals. *Journal of the Acoustical Society of America*, *63*, 456-468. doi: 10.1121/1.381737

Cuddy, L. L. (1968). Practice effects in the absolute judgment of pitch. *Journal of the Acoustical Society of America*, *43*, 1069-1076. doi: 10.1121/1.1910941

Cuddy, L. L. (1970). Training the absolute identification of pitch. *Perception and Psychophysics*, *8*, 265-269. doi: 10.3758/BF03212589

Dooley, K., & Deutsch, D. (2010). Absolute pitch correlates with high performance on musical dictation. *Journal of the Acoustical Society of America*, *128*, 890-893. doi: 10.1121/1.3458848

Dooley, K., & Deutsch, D. (2011). Absolute pitch is correlated with high performance on relative pitch tasks. *Journal of the Acoustical Society of America*, *129*, 2582. doi: 10.1121/1.3588532

Eriksen, C. W., & Hake, H. W. (1957). Anchor effects in absolute judgments. *Journal of Experimental Psychology*, *53*, 132-138. doi: 10.1037/h0047421

Fujisaki, W., & Kashino, M. (2002). The basic hearing abilities of absolute pitch possessors. *Acoustical Science and Technology*, *23*, 77-83. doi: 10.1250/ast.23.77

Geringer, J. M. (1976). Tuning preferences in recorded orchestral music. *Journal of Research in Music Education*, *24*, 169-176. doi: 10.2307/3345127

Goldstone, R. L., & Hendrickson, A. T. (2010). Categorical perception. *Wiley Interdisciplinary Reviews: Cognitive Science*, *1*, 69-78. doi:10.1002/wcs.026

Gregersen, P. K., Kowalsky, E., & Li, W. (2007). Reply to Henthorn and Deutsch: Ethnicity versus early environment: Comment on "Early childhood music education and predisposition to absolute pitch: Teasing apart genes and environment" by Peter K. Gregersen, Elena Kowalsky, Nina Kohn, and Elizabeth West. *American Journal of Medical Genetics Part A*, *143A*, 104-105. doi: 10.1002/ajmg.a.31595

Hall, D. E., & Hess, J. T. (1984). Perception of musical interval tuning. *Music Perception*, *2*, 166-195. doi: 10.2307/40285290

Hanley, J. R., & Roberson, D. (2011). Categorical perception effects reflect differences in typicality on within-category trials. *Psychonomic Bulletin and Review*, *18*, 355-363. doi: 10.3758/s13423-010-0043-z

Harris, G., & Siegel, J. A. (1975). Categorical perception and absolute pitch. *Journal of the Acoustical Society of America*, *57*, S11. doi: 10.1121/1.1995063

Hartmann, W. M. (1993). On the origin of the enlarged melodic octave. *Journal of the Acoustical Society of America*, *93*, 3400-3409. doi: 10.1121/1.405695

Howard, D., Rosen, S., & Broad, V. (1992). Major/minor triad identification and discrimination by musically trained and untrained listeners. *Music Perception*, *10*, 205-220. doi: 10.2307/40285607

Hsieh, I. H., & Saberi, K. (2008). Language-selective interference with long-term memory for musical pitch. *Acta Acustica united with Acustica*, *94*, 588-593. doi: 10.3813/AAA.918068

Ikeda, S. (2010). Accuracy and encoding of absolute pitch: The effect of phonetic interference on absolute pitch identification. *Japanese Journal of Cognitive Psychology*, *8*, 41-51.

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*, 553-562. doi: 10.1121/1.412280

Iverson, P., & Kuhl, P. K. (2000). Perceptual magnet and phoneme boundary effects in speech perception: Do they arise from a common mechanism? *Perception and Psychophysics*, *62*, 874-886. doi: 10.3758/BF03206929

Krumhansl, C. L., & Keil, F. C. (1982). Acquisition of the hierarchy of tonal functions in music. *Memory and Cognition*, *10*, 243-251. doi:10.3758/BF03197636

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*, 358-368. doi: 10.1037/h0044417

Mason, J. A. (1960). Comparison of solo and ensemble performances with reference to Pythagorean, just, and equi-tempered intonations. *Journal of Research in Music Education*, *8*, 31-38. doi: 10.2307/3344235

McFadden, D., & Callaway, 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. doi: 10.1037/0096-1523.25.2.543

Mendel, A. (1978). Pitch in Western music since 1500: A re-examination. *Acta Musicologica*, *50*, 1-93.

Miyazaki, K. (1988). Musical pitch identification by absolute pitch possessors. *Perception and Psychophysics*, *44*, 501-512. doi: 10.3758/BF03207484

Miyazaki, K. (1992). Perception of musical intervals by absolute pitch possessors. *Music Perception*, *9*, 413-426. doi: 10.2307/40285562

Miyazaki, K. (1993). Absolute pitch as an inability: Identification of musical intervals in a tonal context. *Music Perception*, *11*, 55-72. doi: 10.2307/40285599

Miyazaki, K. (1995). Perception of relative pitch with different references: Some absolute-pitch listeners can't tell musical interval names. *Perception and Psychophysics*, *57*, 962-970. doi: 10.3758/BF03205455

Miyazaki, K. (2004a). How well do we understand absolute pitch? *Acoustical Science and Technology, 25*, 426-432. doi: 10.3758/BF03205455

Miyazaki, K. (2004b). The auditory Stroop interference and the irrelevant speech/pitch effect: Absolute pitch listeners can't suppress pitch labeling. Paper presented at the *18th International Congress on Acoustics* (*ICA 2004),* Kyoto, Japan, April 4-9, 2004.

Miyazaki, K. (2004c). Recognition of transposed melodies by absolute-pitch possessors. *Japanese Psychological Research*, *46*, 270-282. doi: 10.1111/j.1468-5584.2004.00260.x

Miyazaki, K., & Ogawa, Y. (2006). Learning absolute pitch by children. *Music Perception*, *24*, 63-78. doi: 10.1525/mp.2006.24.1.63

Miyazaki, K., & Rakowski, A. (2002). Recognition of notated melodies by possessors and nonpossessors of absolute pitch. *Perception and Psychophysics*, *64*, 1337-1345. doi: 10.3758/BF03194776

Rakowski, A. (1976). Tuning of isolated musical intervals. *Journal of the Acoustical Society of America*, *59*, S50. doi: 10.1121/1.2002737

Rakowski, A. (1993). Categorical perception in absolute pitch. *Archives of Acoustics*, *18*, 515-523.

Russo, F. A., & Thompson, W. F. (2005). The subjective size of melodic intervals over a two-octave range. *Psychonomic Bulletin and Review*, *12*(6), 1068-1075. doi: 10.3758/BF03206445

Parncutt, R., & Levitin, D. J. (2001). Absolute pitch. In S. Sadie (Ed.), *The new Grove dictionary of music and musicians* (2nd ed., pp. 37-39). New York: St. Martin's Press.

Perlman, M., & Krumhansl, C. L. (1996). An experimental study of internal interval standards in Javanese and Western musicians. *Music Perception*, *14*, 95-116. doi: 10.2307/40285714

Profita, J., & Bidder, T.G. (1988). Perfect pitch. *American Journal of Medical Genetics*, *29*, 763-771. doi: 10.1002/ajmg.1320290405

Samplaski, A. (2005). Interval and interval class similarity: Results of a confusion study. *Psychomusicology*, *19*, 59-74. doi: 10.1037/h0094040

Schellenberg, E. G. (2002). Asymmetries in the discrimination of musical intervals: Going out-of-tune is more noticeable than going in-tune. *Music Perception*, *19*, 223-248. doi: 10.1525/mp.2001.19.2.223

Schellenberg, E. G., & Trehub, S. E. (1994). Frequency ratios and the discrimination of pure tone sequences. *Perception and Psychophysics*, *56*, 472-478. doi: 10.3758/BF03206738

Schellenberg, E. G., & Trehub, S. E. (1996). Natural musical intervals: Evidence from infant listeners. *Psychological Science*, *7*, 272-277. doi: 10.1111/j.1467-9280.1996.tb00373.x

Sergeant, D. (1969). Experimental investigation of absolute pitch. *Journal of Research in Music Education, 17*, 135-143. doi: 10.2307/3344200

Siegel, J. A., & Siegel, W. (1977a). Categorical perception of tonal intervals: Musicians can't tell sharp from flat. *Perception and Psychophysics*, *21*, 399-407. doi: 10.3758/BF03199493

Siegel, J. A., & Siegel, W. (1977b). Absolute identification of notes and intervals by musicians. *Perception and Psychophysics*, *21*, 143-152. doi: 10.3758/BF03198717

Stewart, N., Brown, G. D. A., & Chater, N. (2005). Absolute identification by relative judgment. *Psychological Review*, *112*, 881-911. doi: 10.1037/0033-295X.112.4.881

Studdert-Kennedy, M., Liberman, A. M., Harris, K. S., & Cooper, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, *77*, 234-249. doi: 10.1037/h0029078

Vos, J. (1982). The perception of pure and mistuned musical fifths and major thirds: Thresholds for discrimination, beats, and identification. *Perception and Psychophysics*, *32*, 297-313. doi: 10.3758/BF03206236

Vurma, A., & Ross, J. (2006). Production and perception of musical intervals. *Music Perception*, *23*, 331-344. doi: 10.1525/mp.2006.23.4.331

Ward, W. D., & Burns, E. M. (1982). Absolute pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 141-182). New York: Academic Press.

Ward, W. D., & Martin, D. W. (1961). Psychophysical comparison of just tuning and equal temperament in sequences of individual tones. *Journal of the Acoustical Society of America*, *33*, 586-588. doi:10.1121/1.1908732

Zatorre, R. J., & Halpern, A. R. (1979). Identification, discrimination, and selective adaptation of simultaneous musical intervals. *Perception and Psychophysics*, *26*, 384-395. doi: 10.1121/1.2017241