Pitch-interval discrimination and musical expertise: Is the semitone a perceptual boundary?

Jean Mary Zarate,a) Caroline R. Ritson, and David Poeppel

Department of Psychology, New York University, 6 Washington Place, New York, New York 10003

(Received 2 September 2011; revised 11 June 2012; accepted 14 June 2012)

The ability to discriminate pitch changes (or intervals) is foundational for speech and music. In an auditory psychophysical experiment, musicians and non-musicians were tested with fixed- and roving-pitch discrimination tasks to investigate the effects of musical expertise on interval discrimination. The tasks were administered parametrically to assess performance across varying pitch distances between intervals. Both groups showed improvements in fixed-pitch interval discrimination as a function of increasing interval difference. Only musicians showed better roving-pitch interval discrimination as interval differences increased, suggesting that this task was too demanding for non-musicians. Musicians had better interval discrimination than non-musicians across most interval differences in both tasks. Interestingly, musicians exhibited improved interval discrimination starting at interval differences of 100 cents (a semitone in Western music), whereas non-musicians showed enhanced discrimination at interval differences exceeding 125 cents. Although exposure to Western music and speech may help establish a basic interval-discrimination threshold between 100 and 200 cents (intervals that occur often in Western languages and music), musical training presumably enhances auditory processing and reduces this threshold to a semitone. As musical expertise does not decrease this threshold beyond 100 cents, the semitone may represent a musical training-induced intervallic limit to acoustic processing. © 2012 Acoustical Society of America.

PACS number(s): 43.66.Fe, 43.66.Hg, 43.75.Cd, 43.75.St [DD] Pages: 984–993

I. INTRODUCTION

Pitch is a central perceptual attribute for both speech and music. Pitch changes in the context of an utterance’s intonation contour indicate the linguistic intent of a sentence (e.g., declarative versus interrogative), set the emotional context (e.g., happy, angry, sad), or distinguish lexical meanings in tonal languages. In music, pitch changes create melody. The magnitude of pitch changes in music can be measured by the frequency ratio between two sounds or the interval size; these intervals have specific labels in Western music composition, such as the minor third, the fifth, and the octave. In contrast, pitch intervals are not as precisely measured or used in speech—the overall contour (pattern of rising or falling intervals) of the utterance is more important to derive the intent or lexical meaning of speech (Dowling and Harwood, 1986).

Given the importance of processing pitch intervals, many researchers have examined pitch-interval discrimination extensively (for a comprehensive review, see Burns, 1999) in purely musical contexts, since, as one author noted, comparing pitch-interval sizes alone is an artificial task with no musical or ecological validity (Burns, 1999). These investigations of interval discrimination employed tasks of categorizing intervals according to Western musical labels and/or interval discrimination around musically relevant intervals (Burns and Ward, 1978; Hill and Summers, 2007; Houtsma, 1968; Zatorre, 1983; Zatorre and Halpern, 1979), adjusting mistuned intervals (Rakowski, 1976; Ward, 1954), and analysis of performance intonation (Dowling, 1978; Ward, 1970). Given these musically relevant constraints, musicians should discriminate and categorize pitch intervals better than non-musicians. However, studying pitch-interval discrimination only in musical contexts overlooks the importance of pitch changes in speech, where pitch intervals have no standardized categories or labels.

In our study, we assessed pitch-interval discrimination in both musicians and non-musicians to assess the effects of extensive musical training (or the lack thereof) in a less musical context—i.e., without presenting the intervals to be compared within a musical phrase, or as done in previous work listed earlier, musical categorization of intervals or adjustment of mistuned intervals. More specifically, we wanted to examine whether musicians and non-musicians would exhibit similar patterns of interval discrimination after removing musical contexts (and thus biases toward individuals with musical training; see the following discussion). Similar to Houtsma (1968) and McDermott and colleagues (2010), we employed a two-alternative forced-choice paradigm in which subjects had to indicate which pure-tone pitch interval was larger. However, our tasks were not designed to find discrimination thresholds like the experiments of Houtsma and McDermott et al. Rather, our ultimate goal was to determine the neural substrates underlying elementary operations—namely, calculations and comparisons of various aspects of auditory stimuli—thought to be executed in particular brain regions within the dorsal auditory stream (see Dehaene, 2009; Griffiths and Warren, 2002). In this study, we specifically targeted calculations and comparisons...
of pitch intervals that may contribute to speech and music processing. Therefore, we designed our discrimination tasks in a parametric fashion, such that the difference between intervals was augmented incrementally to identify brain regions that are increasingly recruited in the calculation and size-related comparison (larger versus smaller) of pitch intervals as a function of the interval difference. Here, we present the psychophysical data collected during the first of three stand-alone sessions; the functional images, neuromagnetic data, and their corresponding behavioral data gathered from two different sessions are discussed in separate publications.

We expected that both fixed- and roving-pitch interval discrimination would improve in both groups as the interval difference increased. We also hypothesized that all subjects would discriminate between fixed-pitch intervals better than roving-pitch intervals, as reported previously (Burns and Campbell, 1994); subjects may only need to compare the second and fourth pitches, rather than extracting and then comparing magnitudes from intervals presented at different frequencies. As expected, musicians perform better than non-musicians in basic pitch- and pitch-interval discrimination tasks by virtue of enhanced auditory skills, which stems presumably from their musical training (Kishon-Rabin et al., 2001; McDermott et al., 2010; Micheyl et al., 2006; Spiegel and Watson, 1984). Additionally, discrimination tasks with musically relevant frequencies and intervals can give musicians an extra advantage in task performance; musicians with relative pitch sometimes can identify or sing a particular note if they use it frequently. For instance, violinists can often sing an A4 or 440 Hz on command because they use it to tune their instruments. Therefore, we removed additional musical-training advantages by selecting base frequencies and interval differences that could not be easily assigned to Western musical conventions—such as A4 (440 Hz) or a perfect fourth (500 cents)—with the exception of the 100 cent interval difference, which is equivalent to a semitone in Western music. Most of the 16 intervals (75%) used to create interval differences could not be classified easily with Western musical conventions—such as A4 (440 Hz) or a perfect fourth (500 cents)—with the exception of the 100 cent interval difference, which is equivalent to a semitone in Western music. Most of the 16 intervals (75%) used to create interval differences could not be classified easily with Western musical standards. In addition, this experiment was designed to emphasize the underlying calculations and comparisons between intervals needed to perform the interval-discrimination task, and not the identification of the presented intervals per se. Despite the neutralization of some musical contexts, we still predicted that musicians would perform better than the non-musicians in these tasks due to their training-enhanced auditory skills.

II. METHODS

A. Subjects

A total of 25 subjects were recruited from the New York University (NYU) community and surrounding areas. We excluded four subjects from the study who exhibited pitch-discrimination thresholds (ranging from 115 to 438 cents) outside 2 standard deviations (SD) from the remaining subjects’ average pitch-discrimination threshold at 225 Hz, which indicated that they may have had difficulty with performing the fixed-pitch interval discrimination task.

The remaining 21 subjects (mean age = 24.1 years, SD = 5.6 years) were right-handed, had normal hearing, and had no neurological disorders. All testing was performed with the subjects’ informed consent and in accordance with procedures approved by the NYU Committee on Activities Involving Human Subjects. All subjects were categorized as non-musicians or musicians according to self-report of musical experience, as assessed with an in-house survey. Eight non-musicians (six female) had minimal musical experience (mostly exposure to music classes in school; mean = 0.8 years, SD = 1.5 years, range = 0–4.7 years) and did not play music regularly at the time of study. Thirteen musicians (seven female) had a minimum of 3 years of musical experience and/or training (via formal music lessons or self-training and musical performance; mean = 11.5 years, SD = 4.6 years, range = 4.2–20 years) and were practicing or performing music at the time of study. None of the subjects reported having absolute pitch.

B. Stimuli

We used MATLAB (The MathWorks, Natick, MA) and the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) to create and present all auditory stimuli. For all interval-discrimination stimuli, 200 ms sinusoidal tones were created at base frequencies of 225 and 475 Hz (16 bit depth, 44 100 Hz sampling frequency, 7 ms cosine-ramp onset and offset). Additional sinusoidal tones were created at particular pitch distances from both base frequencies (25–400 cents at 25 cent increments; 100 cents is equal to 1 semitone, the smallest tonal unit in Western music). Tones were then paired (50 ms gap of silence between tones) to create intervals ranging from 25 to 400 cents in each base frequency (225 or 475 Hz). Intervals were combined [interstimulus interval (ISI) = 0.8, 0.9, or 1 s] to create individual test trials, and we parametrically varied the differences between intervals within a trial from 25 to 175 cents (1/8 whole tone to 7/8 whole tone in Western music), in 25 cent increments. In the “fixed” task, both pairs in a given trial were presented at the same base frequency, such that the first and third tones were equal to 225 Hz, and only the second and fourth tones were altered to create different intervals for comparison. In the “roving” task, tone pairs were presented at different base frequencies (225 and 475 Hz), and subjects had to extract and compare the pitch intervals despite the change in base frequency. In both tasks, the intervals (ranging from 25 to 400 cents) used to create each interval difference (from 25 to 175 cents at 25 cent steps) were chosen at random for each participant; as a result, the number of trials during which interval sizes increased (i.e., second pair was larger than the first pair) or decreased were not fixed.

C. Experimental procedure

Participants sat in at a computer and wore headphones (Sennheiser HD 380 Professional, Sennheiser Electronic Corporation, Wedemark, Germany), through which all auditory stimuli were delivered at a comfortable level (mean 65 dB, SPL A). Prior to interval discrimination, pitch-discrimination
thresholds were measured across three testing blocks at each base frequency (225 and 475 Hz) in a “2 down–1 up” adaptive staircase procedure (Levitt, 1971) implemented as part of the MLP toolbox for MATLAB tool for auditory psychophysical testing (Grassi and Soranzo, 2009). After discrimination-threshold testing, subjects were presented with intervals in a two-alternative, forced-choice design, and they indicated the larger interval with a mouse-button press in speeded trials (i.e., reaction times were measured). The fixed and roving conditions were presented in separate blocks (one block each) in a pseudo-randomized order. Each of the seven interval differences was presented a total of 20 times in a pseudo-randomized order. Each of the seven interval differences was presented across three blocks of threshold testing, subjects were presented with intervals in a two-alternative, forced-choice design, and they indicated the larger interval with a mouse-button press in speeded trials (i.e., reaction times were measured). The fixed and roving conditions were presented in separate blocks (one block each) in a pseudo-randomized order. Each of the seven interval differences was presented a total of 20 times in a pseudo-randomized order within each block, resulting in a total of 140 trials per block.

D. Analyses

Given that our parametric manipulation of interest focused on the interval difference, and not on the interval magnitudes presented in each trial, we analyzed the data by interval difference rather than by individual interval magnitudes. In order to query our data for individual interval-magnitude effects, we would require an unwieldy examination of pairs of interval magnitudes, as there were 14 possible interval combinations to create a 25 cent interval difference (e.g., 25 plus 50 cents, 50 plus 75 cents, 375 plus 400 cents), 13 possible pairings for a 50 cent interval difference, 12 possible combinations for an interval difference of 75 cents, and so on. For each interval difference, these pairings were chosen at random and may not be equally presented in each subject or even across subjects. As our design is not optimal for assessing interval-magnitude effects, we therefore collapsed our data across all interval sizes (range: 25–400 cents) to target our parametrically varied interval differences (25–175 cents). This also makes our interval-discrimination analyses more similar to pitch-discrimination analyses, in which discrimination is measured by the difference between individual pitches, not by the frequency of the pitches themselves. However, the potential for interval-magnitude influences on interval discrimination should be examined in future research.

Subjects’ performances were measured as percent-correct scores for each of the interval differences within each condition (fixed and roving). Using signal detection theoretic analyses, we also calculated $d'$ and $\beta_{\text{normalized}}$ values to measure sensitivity and response bias, respectively, (Dorfman and Alf, 1968; Rosenblith and Stevens, 1953; Swets, 1982). The hit and false alarm (FA) rates and $d'$ and $\beta_{\text{normalized}}$ values were calculated in the following manner:

<table>
<thead>
<tr>
<th>Formula</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hit = $H(\times$ times 1st pair was chosen/$\times$ trials with larger 1st pair) = score for larger 1st pair trials,</td>
<td></td>
</tr>
<tr>
<td>FA = $1–H(\times$ times 2nd pair was chosen/$\times$ trials with larger 2nd pair) = 1–score for larger 2nd pair trials,</td>
<td></td>
</tr>
<tr>
<td>$d' = Z\text{score}(\text{Hit})–Z\text{score}(\text{FA})$,</td>
<td></td>
</tr>
<tr>
<td>$\beta = -0.5(Z\text{score}(\text{Hit}) + Z\text{score}(\text{FA}))$,</td>
<td></td>
</tr>
<tr>
<td>$\beta_{\text{normalized}} = \beta/d'$; 0 = no bias; negative values = bias toward selecting 1st pair, positive values = bias toward selecting 2nd pair.</td>
<td></td>
</tr>
</tbody>
</table>

Pitch-discrimination thresholds (in cents), percent-correct scores, $d'$ values, $\beta_{\text{normalized}}$ values, mean reaction time, and SD of reaction times were analyzed using repeated-measures analyses of variance (ANOVA). Planned comparisons were used to further analyze significant interactions. We also examined correlations between the mean percent-correct scores across all interval differences and (1) mean pitch-discrimination thresholds (across three blocks of threshold testing) at 225 and 475 Hz and (2) variables from the in-house musical experience survey (i.e., total musical experience measured in months, number of hours per week spent listening to music, rating of ability to play music by ear, and rating of sight-singing ability).

Although we performed correlation analyses with total musical experience as a continuous variable to assess the overall effect of musical experience on interval discrimination, we performed ANOVAs with musical experience as a grouping variable; the final group classification criterion was based on whether subjects had been performing music at the time of study, which makes the grouping variable categorical in nature. Additionally, the type of musical experience varied greatly within the musician group (e.g., self-taught versus formal training in one or more different instruments, etc.), and thus the ANOVA results from analyses with musical experience as a continuous variable would be ambiguous.

III. RESULTS

A. Effects of musical experience on pitch discrimination and interval-discrimination accuracy

A two-way repeated-measures ANOVA (group by base frequency) performed on the average pitch-discrimination thresholds resulted in significant main effects of group $[F(1,19) = 6.08, \ p < 0.05]$ and base frequency $[F(1,19) = 16.90, \ p < 0.001]$ and a marginally significant interaction between group and base frequency $[F(1,19) = 3.38, \ p < 0.09]$. Planned comparisons on the two-way interaction determined that non-musicians’ discrimination thresholds were significantly higher than those of musicians only at 475 Hz (Table I; $p < 0.01$). Within the non-musician group,
TABLE I. Average ± SEM pitch-discrimination thresholds (in cents) at frequencies of 225 and 475 Hz for non-musicians and musicians.

<table>
<thead>
<tr>
<th></th>
<th>Pitch-discrimination thresholds (cents)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>225 Hz</td>
</tr>
<tr>
<td>Non-musicians</td>
<td>23 ± 4</td>
</tr>
<tr>
<td>Musicians</td>
<td>14 ± 3</td>
</tr>
</tbody>
</table>

*Lower thresholds reflect better pitch discrimination. The discrimination threshold of the non-musicians was significantly higher than that of musicians at 475 Hz (denoted by **; \( p < 0.01 \)). Thresholds within the non-musician group were significantly different between the two frequencies (marked by **; \( p < 0.01 \)), while thresholds within the musician group were only marginally different (denoted by *, \( p < 0.09 \)).

thresholds at 475 Hz were significantly higher than those at 225 Hz (Table I; \( p < 0.01 \)), whereas thresholds at 475 Hz were only marginally higher than thresholds at 225 Hz within the musician group (\( p < 0.09 \)).

Figure 1 illustrates the significant results of a repeated-measures, three-way ANOVA performed on percent-correct scores—with group as the between-subject factor, and condition (fixed or roving) and interval difference as the repeated within-subject factors—that was performed to assess musical training effects on fixed- and roving-pitch interval discrimination across parametric increases of interval difference. The ANOVA revealed significant main effects of group \( [F(1,19) = 6.10, \quad p < 0.05] \), condition \( [F(1,19) = 109.49, \quad p < 0.001] \), and interval difference \( [F(6,114) = 38.91, \quad p < 0.001] \), a significant two-way interaction between condition and interval difference \( [F(6,114) = 4.08, \quad p < 0.001] \), and a significant three-way interaction between all three factors (Fig. 1; \( F(6,114) = 2.30, \quad p < 0.05 \)). Planned comparisons on the three-way interaction established that musicians were better at discriminating between fixed-pitch intervals with 25, 100, (\( p < 0.05 \)) 125, and 150 cent differences (\( p < 0.08 \)) and roving-pitch intervals with 50, 100, 150, and 175 cent differences than non-musicians (\( p < 0.05 \)). Within each group, performances were significantly poorer during the roving task when compared to the fixed task, for all interval differences except 25 cents (\( p < 0.01 \) among musicians; \( p < 0.05 \) among non-musicians). This suggests that interval discrimination at 25 cents was already quite difficult in the fixed condition, and transposing one of the intervals to another base frequency (as in the roving condition) did not significantly alter performance in either group at the 25 cent interval difference. As expected, musicians’ discriminations improved as interval differences increased, with interval discrimination at a 25 cent difference being significantly poorer than all other interval differences in the fixed (\( p < 0.05 \)) and roving (\( p < 0.1 \)) conditions. Additionally, musicians performed significantly worse on all interval differences of 25–75 cents, compared to performance at 100 cents and larger, both in the fixed (\( p < 0.01 \)) and roving (\( p < 0.05 \)) conditions. Further, musician interval discrimination is significantly better at a 100 cent interval difference than at an interval difference of 75 cents for both fixed and roving conditions (\( p < 0.05 \)). No other significant differences were found between other adjacent interval differences (e.g., 50–75, 100–125, 125–150 cents, etc.), with the exception of significantly different performance between 25 and 50 cent interval differences in the fixed condition (\( p < 0.05 \)); this reiterates that performance at the smallest interval difference was worse than at all other levels. Non-musicians, on the other hand, showed a different performance profile. Their scores improved as interval differences increased only in the fixed task, with the exception of no significant improvement between differences of 125 and 150 cents (\( p > 0.1 \)); as also seen in the musicians, non-musicians had the worst scores at 25 cents (\( p < 0.07 \)). However, non-musicians’ performances during the roving task did not show the expected improvement as interval differences

![Figure 1](image-url)
increased [Fig. 1(B)]. Although the score at the 50 cent difference was significantly worse than scores at the 100 cent difference and larger (marked by *, \( p < 0.07 \)) as expected, surprisingly their performance at 50 cents difference also was worse than at a 25 cent interval difference (\( p < 0.05 \)). Additionally, there were no significant differences between scores at 75 and 100 cent, at 100 and 125 cent, and at 125 and 150 cent differences (\( p > 0.1 \)).

B. Effects of musical experience on interval-discrimination sensitivity and response bias

Figure 2(A) depicts the \( d' \) values for both groups during interval discrimination. We performed a three-way repeated-measures ANOVA (group by condition by interval difference) on \( d' \) values to assess how musical training influenced
discrimination sensitivity in both conditions across varying interval differences. The ANOVA revealed significant main effects of group \(F(1,19) = 6.59, p < 0.05\), condition \(F(1,19) = 48.93, p < 0.001\), and interval difference \(F(6,114) = 29.16, p < 0.001\), and significant two-way interactions between interval difference and group \(F(6,114) = 2.87, p < 0.05\) and condition and interval difference \(F(6,114) = 2.87, p < 0.05\). Planned comparisons on the interaction between group and interval difference clarified that musicians were more sensitive to interval differences of 100 and 125 cents than non-musicians [Fig. 2(A, left); \(p < 0.05\)]. As seen with percent-correct scores within the musician group, interval-discrimination sensitivity was much poorer at interval differences of 25–75 cents than interval differences of 100 cents and larger (\(p < 0.01\)). Musician discrimination sensitivity at the 25 cent interval difference was the poorest, compared to all other levels (\(p < 0.05\)). Among the larger interval differences, there was only a significant improvement in sensitivity between 100 and 175 cent differences; all other values were not significantly different from each other (\(p > 0.1\)). In the non-musician group, discrimination sensitivity at interval differences from 25 to 125 cents was significantly poorer than at 150 and 175 cents [Fig. 2(A, left); \(p < 0.05\)]; sensitivity at 25, 50, and 100 cent differences was also worse than at 125 cents (\(p < 0.05\)). No other significant differences were found. All together, this significant interaction suggests that musicians have a threshold for interval discrimination at the 100 cent interval difference—the only interval size with musical relevance in Western music—whereas non-musicians require interval differences larger than 100 cents to discriminate between tone-pairs successfully [see Fig. 2(A, left)]. Planned comparisons performed on the condition-by-interval-difference interaction determined that all subjects were significantly better at detecting differences between fixed-pitch intervals at all interval differences larger than 25 cents, compared to roving-pitch intervals [Fig. 2(A, right); \(p < 0.05\)]. During the fixed task, there was a steady, significant increase in the ability to discriminate between intervals as the interval difference grew larger, with values at differences of 25–125 cents being significantly smaller than the \(d'\) value at 175 cents [Fig. 2(A, right); all \(p < 0.05\)]. During the roving task, there was no significant improvement in interval discrimination until the interval difference was equal to 100 cents or higher [Fig. 2(A, right); \(p < 0.1\)], again suggesting that intervals are better discerned once they are separated by 100 cents or larger, even if they are presented at different base frequencies.

Three-way repeated-measures ANOVA (group by condition by interval difference) on \(\beta_{normalized}\) values [measure of response bias; Fig. 2(B)] to examine the effects of musical training on response bias across different interval differences in both discrimination tasks (fixed and roving pitch). This analysis resulted in significant main effects of group \(F(1,19) = 7.10, p < 0.05\), condition \(F(1,19) = 53.03, p < 0.001\), and interval difference \(F(6,114) = 31.05, p < 0.001\), and significant interactions between group and interval difference \(F(6,114) = 2.76, p < 0.05\) and condition and interval difference \(F(6,114) = 2.30, p < 0.05\). Planned comparisons on the interaction between the group and interval-difference factors determined that non-musicians had a stronger bias than musicians toward selecting the second interval as the larger interval at differences of 100, 125, and 175 cents [Fig. 2(B, left); \(p < 0.05\) for 100–125 cents, \(p < 0.09\) for 175 cents]. Within each group, this bias was reduced as interval differences increased in size, which reflects an expected improvement in discriminating between intervals as interval differences increased (see results for \(d'\) values). Musicians showed larger bias values at interval differences from 25 to 125 cents than the bias at a 175 cent difference, with a 25 cent difference evoking the highest bias toward selecting the second tone pair as the largest (\(p < 0.06\)). Similarly, non-musicians exhibited greater biases at interval differences from 25 to 125 cents than at 150–175 cents (\(p < 0.06\)). More importantly, musicians’ biases seen at interval differences of 25–75 cents were significantly stronger than those observed at 100 cent differences and larger (\(p < 0.001\)), and non-musicians’ \(\beta_{normalized}\) values were greater at 25–100 cent differences than at interval differences of 125 cents and larger (\(p < 0.06\), with the exception of no significant difference between values at interval differences of 75 and 125 cents). Planned comparisons on the condition-by-interval-difference interaction found that all subjects—regardless of group—showed a stronger bias toward picking the second pair as the larger interval during the roving task across all interval differences except at the 25 cent difference, compared to the fixed task [Fig. 2(B, right); \(p < 0.01\)]. Within each condition, this bias toward the second pair decreased as the interval differences increased in size (\(p < 0.09\), most likely due to decreased task difficulty and better discrimination sensitivity.

Figure 3 displays significant correlations between all subjects’ mean percent-correct scores (averaged across all interval differences) for each condition and other variables, such as pitch-discrimination thresholds and musical-experience variables collected from an in-house survey (i.e., total musical experience measured in months, number of hours per week spent listening to music, self-rating of ability to play music by ear, and self-rating of sight-singing ability). Subjects’ average pitch-discrimination thresholds at 225 Hz were negatively correlated with mean percent-correct scores in the fixed-pitch task \(r(19) = -0.59, p < 0.01\), which suggests that subjects with better pitch discrimination at 225 Hz (indicated by lower thresholds) were more accurate at fixed-pitch interval discrimination [Fig. 3(A, top)]. Performance at the roving-pitch task was also negatively correlated with the pitch-discrimination thresholds at both of the base frequencies within a given trial [225 and 475 Hz; \(r(19) = -0.56, p < 0.01\)], again suggesting that listeners with better discrimination abilities perform more accurately in the roving interval-discrimination task [Figs. 3(A) and 3(B)]. Notably, these negative correlations are seen with the entire study sample, regardless of group designation; i.e., non-musicians who performed better at pitch discrimination than some musicians also performed more accurately at interval discrimination. Fixed- and roving-pitch interval discrimination were positively correlated with the total amount of months of musical training or experience; subjects with more musical training or experience performed better at both tasks than those with less
musical experience [Fig. 3(C); fixed: $r(19) = 0.64$, $p < 0.01$; roving: $r(19) = 0.58$, $p < 0.01$]. Additionally, subjects’ self-ratings of their ability to play music back by simply listening to it (“playing-by-ear”) were positively correlated with performance in the roving task—as subjects were more confident with their ability to play-by-ear, perhaps due in part to their musical experience, they were also more accurate in discriminating between intervals that were presented at different base frequencies [Fig. 3(D); $r(19) = 0.45$, $p < 0.05$]. Remarkably, almost all of the subjects who were the least confident with playing-by-ear were non-musicians; the only musician who was the least confident with this ability also performed the worst on the roving-pitch task, compared to other musicians [Fig. 3(D)]. Interval-discrimination performance in either condition did not significantly correlate with other musical-experience variables.
C. Reaction time

Analysis of the mean reaction times of each group for both fixed and roving tasks at each interval difference revealed a marginally significant main effect of condition \( [F(1,19) = 3.82, p < 0.07] \) and a significant main effect of interval difference \( [F(6,114) = 13.80, p < 0.001] \). A significant two-way interaction between condition and interval difference was also found \( [F(6,114) = 5.04, p < 0.001] \), but no other significant main effects or interactions were revealed. Planned comparisons performed on the two-way interaction determined that reaction times were significantly longer during the roving task than during the fixed task at interval differences of 75–150 cents (Fig. 4; \( p < 0.05 \)). During the fixed task, reaction times decreased as the interval differences increased in magnitude, which probably reflected easier interval discrimination \( (p < 0.05) \). In the roving task, reaction times for larger interval differences \( (125–175 \text{ cents}) \) were significantly shorter than for smaller differences, with the reaction time at 175 cents being the shortest \( (p < 0.05) \). Analyses on the variability of reaction times (measured as SD of reaction time) revealed no significant main effects or interactions.

IV. DISCUSSION

A. Summary: Interval discrimination across different frequencies and interval differences

Interval discrimination significantly improved as a function of increasing interval-difference magnitude as predicted. Musicians (by virtue of their extensive auditory training) exhibited better accuracy and sensitivity as interval differences increased in both fixed and roving conditions, whereas non-musicians were more accurate only with increasingly larger interval differences in the fixed condition. The absence of performance improvement (as a function of interval difference) in non-musicians during the roving condition may be due to the inherent musical nature of this particular task. The presentation of intervals at different base frequencies in the roving condition is similar essentially to a musical transposition, where melody intervals are preserved despite a change in key (i.e., central base frequency). In our roving task, subjects are asked to extract and compare interval magnitudes across a “transposition,” which should be easier for musicians; in practice, to reproduce the same melody or chord progression in different keys, musicians need to compare intervals after different transpositions. Alternatively, it is possible that people with a natural propensity for perceiving intervalic relationships across transpositions may be inclined to study and/or perform music, which then reinforces interval comparisons on a regular basis. Thus, even though the task difficulty decreased as interval differences increased, comparing interval magnitudes after a transposition may too difficult for non-musicians. As such, both groups performed better overall at the fixed-pitch task than the roving-pitch task across all interval differences (as expected), except at the 25 cent difference between intervals, which seemed to be the most difficult comparison for all subjects, regardless of the base frequency of the intervals. It should be noted, however, that both groups appear to be performing interval discrimination during the fixed-pitch task, rather than just comparing the second and fourth pitches. If participants were performing only pitch discrimination between these pitches, then musicians should be more accurate at interval differences of 25–75 cents, and non-musicians should perform closer to chance at the 25 cent interval difference and more accurately at all other interval differences, given their measured pitch-discrimination thresholds (Table I). Relative to the fixed-pitch task, the poorer performance observed in our roving-pitch task is similar to an overall decrease in interval-discrimination sensitivity in roving-pitch interval discrimination reported by Burns and Campbell (1994). This common performance profile was concomitant with generally faster reaction times during the fixed task than in the roving task, except at the more difficult interval comparisons with only a 25 or 50 cent difference or relatively easy interval difference of 175 cents.

B. The semitone as a training-induced perceptual boundary

As expected, musicians performed more accurately than non-musicians across the majority of interval differences in fixed- and roving-pitch interval discrimination (outside of a musical context). Analyses of discrimination sensitivity \( (d') \) determined that musicians were more sensitive to 100 and 125 cent differences between intervals in both tasks than non-musicians. Overall, musicians showed greater sensitivity and accuracy at interval differences of 100 cents or greater in both tasks, which is further corroborated by the near absence of a significant change in discrimination at interval differences less than 100 cents; non-musicians exhibited enhanced discrimination sensitivity only at interval differences larger than 125 cents. These results are consistent with the data of McDermott and colleagues (2010), which demonstrated that music-degree students and amateur musicians have interval-discrimination thresholds around 100 cents.
whereas non-musicians exhibit thresholds larger than 100 cents. Despite neutralizing some musical contexts from our discrimination paradigm [compared to Burns and Ward (1978)], we still found evidence of musically relevant sensitivity in musicians, suggesting that musical training or experience may influence interval discrimination. Indeed, our correlation analyses determined that subjects’ total amount of musical experience positively correlated with their overall performance in both fixed- and roving-pitch interval discrimination.

Interestingly, despite the fact that both groups exhibited basic pitch-discrimination thresholds much smaller than a semitone, musicians’ and non-musicians’ interval-discrimination thresholds were found only at 100 cents and higher. Non-musicians’ thresholds for enhanced interval discrimination are ~150 cents, which implies that processing everyday auditory events may require sensitivity to intervals between 100 and 200 cents. Music and speech processing may contribute to establishing this basic interval-discrimination threshold, as a recent study determined that non-tonal languages (including English) and music have a higher occurrence of intervals of 200 cents or smaller than tonal languages (Han et al., 2011). Extensive musical training appears to reduce this interval-discrimination threshold down to a musically meaningful interval of 100 cents, but not further to even smaller magnitudes. Two arguments can arise from this line of reasoning: (1) this threshold may be a by-product of training in the Western musical system, which utilizes various scales constructed with derivatives of the semitone (i.e., 100 cents = minor second, 200 cents = major second, 300 cents = minor third, etc.) or (2) the semitone may be a principal intervarcial limit to acoustic processing, which in turn has influenced the structure of Western musical scales and speech. Our data, along with the results of McDermott and colleagues (2010), currently support the first argument—training in Western music may limit the perceptual boundary of interval discrimination to a semitone, an interval that is predominantly used in Western music and speech.

C. Identifying putative neural substrates underlying interval discrimination

Among the extensive neural networks involved in speech and music processing, posterior auditory cortex and the intraparietal sulcus (IPS) may be involved specifically in processing intervals, as previous studies have shown that the planum temporale and IPS displayed more cortical activity with larger pitch changes (Hyde et al., 2008; Rinne et al., 2007; Zarate et al., 2010). Thus, we hypothesize that parametric increases of interval differences will incrementally recruit cortical activity within these areas. Compared to a control condition with passive listening to intervals, we expect that the calculation and comparison of interval magnitudes required for our interval discrimination tasks may also recruit additional activity within the IPS, as this region has been previously associated with mathematical calculations and comparisons (Dehaene, 2009). Due to its similarity to musical transposition, we predict that the roving-pitch interval discrimination task may recruit more IPS activity than the fixed-pitch task, similar to enhanced IPS activity observed during a melodic-contour comparison task, in which one melody’s base frequency was transposed (Foster and Zatorre, 2010). We also hypothesize that activity within auditory cortex and IPS will be modulated as a function of musical experience, as seen in earlier studies of auditory processing in musicians and non-musicians (Margulis et al., 2009; Pantev et al., 2001; Zarate et al., 2010; Zarate and Zatorre, 2008). Given our findings of training-dependent interval-discrimination thresholds, we propose that cortical activity within auditory cortex and IPS may be significantly different in both groups at suprathreshold interval differences than at smaller magnitudes.

In general, the possible involvement of auditory cortex and IPS in pitch-interval processing and discrimination supports the notion that the dorsal auditory stream, which includes these regions, is involved in processing speech- and music-related functions (Belin and Zatorre, 2000; Hickok and Poeppel, 2000, 2004, 2007). The hemodynamic, neuromagnetic, and behavioral data collected with these interval-discrimination tasks in separate sessions are reported and discussed in this context in separate publications.

ACKNOWLEDGMENTS

The research was funded in part by a grant from the GRAMMY Foundation® to J.M.Z. and by a grant from the National Institutes of Health (NIH R01 05660) to D.P. The authors gratefully acknowledge Luc Arnal, Ph.D., and Xing Tian, Ph.D., for their insightful feedback on this manuscript.


