Symmetric interactions and interference between pitch and timbre

Emily J. Allen[a] and Andrew J. Oxenham
Department of Psychology, University of Minnesota, Minneapolis, Minnesota 55455

(Received 17 July 2013; revised 2 January 2014; accepted 9 January 2014)

Variations in the spectral shape of harmonic tone complexes are perceived as timbre changes and can lead to poorer fundamental frequency (F0) or pitch discrimination. Less is known about the effects of F0 variations on spectral shape discrimination. The aims of the study were to determine whether the interactions between pitch and timbre are symmetric, and to test whether musical training affects listeners' ability to ignore variations in irrelevant perceptual dimensions. Difference limens (DLs) for F0 were measured with and without random, concurrent, variations in spectral centroid, and vice versa. Additionally, sensitivity was measured as the target parameter and the interfering parameter varied by the same amount, in terms of individual DLs. Results showed significant and similar interference between pitch (F0) and timbre (spectral centroid) dimensions, with upward spectral motion often confused for upward F0 motion, and vice versa. Musicians had better F0DLs than non-musicians on average, but similar spectral centroid DLs. Both groups showed similar interference effects, in terms of decreased sensitivity, in both dimensions. Results reveal symmetry in the interference effects between pitch and timbre, once differences in sensitivity between dimensions and subjects are controlled. Musical training does not reliably help to overcome these effects.
[http://dx.doi.org/10.1121/1.4863269]

PACS number(s): 43.66.Hg, 43.66.Fe, 43.66.Jh [FJG] Pages: 1371–1379

I. INTRODUCTION

The sounds we hear can be described in terms of multiple perceptual attributes, including pitch, timbre, and loudness. The present study focuses on pitch and timbre. Although several researchers have suggested that it is multidimensional (e.g., Shepard, 1982), pitch has been defined as that perceptual attribute of sound that can be ordered on a scale from low to high (ANSI, 1994), with the two most commonly cited dimensions being pitch height and pitch chroma, corresponding roughly to the physical attributes of fundamental frequency (F0) and position within an octave, respectively. The present study focuses on the dimension of pitch height. Timbre is associated with multiple acoustical and perceptual attributes (Grey, 1977). Its technical definition includes everything by which a listener can distinguish between sounds with the same loudness and pitch (ANSI, 1994), although duration (Plomp, 1970) and spatial location are also attributes that are not normally considered part of timbre. A primary determinant of timbre is the spectral centroid of a sound (Caclin *et al.*, 2005). In general, a low-frequency emphasis in the spectral envelope leads to a "duller" sound, whereas more high-frequency emphasis leads to a "brighter," "tinnier," or "sharper" sound (e.g., Fastl and Zwicker, 2007).

Although some previous studies have shown pitch and timbre to be perceived independently (e.g., Marozeau *et al.*, 2003), there are several examples of interference between them (e.g., Marozeau and de Cheveigné, 2007; Melara and Marks, 1990). Notably, variations in timbre are known to interfere with subjects' ability to discriminate small changes in pitch. There are different hypotheses regarding how this interference occurs (e.g., Faulkner, 1985; Moore and Glasberg, 1990), but a prevailing view is that changes in spectral timbre (on the dull–bright continuum) either produce a general distraction effect or are confused with changes in pitch height, based on F0 (e.g., Moore and Glasberg, 1990; Singh and Hirsh, 1992; Warrier and Zatorre, 2002; Borchert *et al.*, 2011).

Although many studies have examined the effect of spectral changes on F0 perception and discrimination, fewer have investigated the effects of F0 variation on spectral-shape discrimination. Beal (1985) conducted a study in which both the effect of timbre variation on pitch discrimination and the effect of pitch variation on timbre discrimination was observed. When listening to chord changes on different musical instruments, subjects found it challenging to ignore changes in timbre, i.e., switching between instruments, when attempting to focus exclusively on the pitches in musical chords. They had less difficulty ignoring chord changes when attempting to judge whether the two timbres were the same, suggesting an asymmetry between the dimensions of pitch and timbre. However, the salience or discriminability of the changes in the different dimensions was not controlled, and the timbres were limited to three distinctly different instruments (acoustic guitar, piano, and harpsichord). Beal (1985) also found differences in performance between musicians and non-musicians. Musicians were better at recognizing when the same chord was played on two different instruments, although the benefit of musicianship was only found when the chords were diatonic, suggesting that the successful referencing of familiar musical structures was the defining difference between musicians and non-musicians.

---

[a] Author to whom correspondence should be addressed. Electronic mail: prac0010@umn.edu

Pitt (1994) also compared musicians and non-musicians on pitch and timbre discrimination. In a categorization task, as subjects listened to different tones, they were asked to determine whether there was a pitch change, an instrument change (timbre change), or both. Subjects were not required to report the direction of change, however. Non-musicians were more strongly affected than musicians by variations in timbre when discriminating pitch, suggesting that non-musicians experienced greater difficulty processing the two dimensions independently. However, the number of stimuli used was again limited (two different timbres: recordings of a trumpet and a piano, and two different pitches: 294 Hz and 417 Hz), and no attempt was made to equate the perceptual salience across the two dimensions, making direct comparisons difficult.

Melara and Marks (1990) found interactions between pitch and timbre for individual tones on speeded classification tasks. They attributed these interactions to failure in selective attention, or Garner interference. In one experiment, subjects were instructed to attend to either timbre changes or pitch changes, while both dimensions varied. Like Beal (1985) and Pitt (1994), however, a limited number of stimuli were used: a combination of two different duty cycles of square waves (0.1878 and 0.3128, labeled "twangy" and "hollow," respectively) were combined with two different F0s (900 Hz and 920 Hz). Krumhansl and Iverson (1992) also found interactions between pitch and timbre for individual tones on speeded classification tasks, but used more musical sounds (notes F4 and C5 for the pitches, and a synthesized trumpet and piano for the timbres). They found that variation in the non-target dimension interfered with classification for both pitch and timbre symmetrically. Again, however, a limitation of the study lies in the small number of stimuli used, and the fact that the differences in pitch and timbre were not equated for discriminability or perceptual salience. The importance of equating the dimensions of interest in terms of perceptual salience has been noted in both auditory and visual research by Melara and Mounts (1993, 1994).

More recently, Silbert *et al.* (2009) explored a general framework for understanding interactions between perceptual dimensions based on signal detection theory (Green and Swets, 1966). They used concurrent changes in spectral centroid and F0 as an example of dimensional interactions and concluded that, for most of their seven listeners, the two dimensions were not processed independently. However, because they did not test identification performance for either dimension in isolation and only tested two values of each dimension, it is not clear how much interference each dimension produced on the other, or whether the effects were symmetric. It is also not clear what accounted for the relatively large individual differences observed in that study.

The present study explored the effects of spectral shape variation on F0 discrimination and vice versa. The two aims of the study were (i) to determine whether the interference and interactions between pitch and timbre are symmetric, and (ii) to assess the effects of musical training on subjects' ability to ignore variations in irrelevant perceptual dimensions when performing a discrimination task. The first aim addresses the more general question of whether pitch has a privileged role in auditory perception. For instance, it is known that sensitivity to small changes in pitch is generally much greater than to changes in other dimensions (McDermott *et al.*, 2010), and pitch has been cited as an exception to Miller's "seven plus-or-minus two" rule, in that musicians are able to perfectly identify more than just nine pitch intervals (Burns, 1999). On the other hand, more recent work has suggested that some of the properties that were thought to make pitch "special" can also be found in other dimensions (such as timbre and loudness), when differences in basic sensitivity are equated (e.g., McDermott *et al.*, 2008, 2010).

The second aim tackles the question of differences in basic perceptual skills between musicians and non-musicians. As mentioned above, Silbert *et al.* (2009) observed relatively large individual differences that were not accounted for. One factor may be the amount of prior musical training. There are some studies that have found better performance in musicians than non-musicians in tasks involving both pitch perception (e.g., Micheyl *et al.*, 2006) and analytic listening in an informational masking context (Oxenham *et al.*, 2003). Attending to one dimension and ignoring another could be considered a form of analytic listening, so it may be that musicians are less susceptible to interference effects. In contrast to this expectation, Borchert *et al.* (2011) found no significant benefit of musical training in a task that involved pitch discrimination between two sounds that varied widely in spectral shape. Little is known about differences between musicians and non-musicians in their ability to discriminate spectral shape, with or without the presence of F0 changes. On one hand, some benefit of musicianship in attending selectively to separate auditory dimensions beyond pitch might be expected; on the other hand, timbre discrimination may not be as highly trained in musicians as pitch discrimination because discriminating between very subtle spectral differences is not part of a typical ear-training program.

Experiment 1 measured basic sensitivity to small changes in either F0 or spectral centroid, in the absence of variation in the non-target dimension. Experiment 2 used the individual difference limens (DLs) from Experiment 1 to examine the effects of random variations in either F0 or spectral centroid on listeners' ability to discriminate small changes in the other dimension. Finally, Experiment 3 provided a direct test of perceptual symmetry of the two dimensions by measuring performance in both dimensions using stimuli that varied by the same amount in terms of DLs obtained from the individual subjects.

II. EXPERIMENT 1: BASIC PITCH AND TIMBRE DISCRIMINATION

A. Rationale

The goal of Experiment 1 was to find thresholds for each subject on basic pitch and timbre discrimination tasks. We did this by separately measuring DLs for F0 and spectral centroid of a bandpass-filtered harmonic tone complex. These DLs were then used in subsequent experiments to equate changes in F0 and spectral centroid in terms of basic sensitivity for each subject individually.

B. Methods

1. Stimuli

The stimuli were harmonic complex tones, 500 ms in duration with 20-ms raised cosine onset and offset ramps, and an overall level of 70 dB sound pressure level. The components were added in sine phase. All harmonics of the complex tone up to 10 000 Hz were generated and then individually scaled to produce slopes of 24 dB/octave around the center frequency (CF), or spectral centroid, with no flat bandpass region. Thus, the 3-dB bandwidth of the filter was 1/4 octave. MATLAB (Mathworks, Natick, MA) was used to generate the stimuli and control the experimental procedures. All stimuli were generated via an L22 soundcard (LynxStudio, Costa Mesa, CA) with 24-bit resolution at a sampling rate of 44 100 Hz, and were presented diotically through HD580 headphones (Sennheiser, Old Lyme, CT).

In the pitch discrimination task, the CF of the filter was held constant at 1200 Hz. The nominal F0 value of 200 Hz was roved across trials by ±10% with uniform distribution. Each trial consisted of two presentation intervals, each containing a complex harmonic tone with the F0s differing by Δ_{F0}, expressed as a percentage of the F0 of the lower tone. The F0s of the two tones in each trial were geometrically centered around the nominal F0 value after roving.

In the timbre-discrimination task, the F0 of the complex tone was held constant at 200 Hz, and the nominal CF of the bandpass filter was roved between trials by ±10% around 1200 Hz, with uniform distribution. Within each trial, the CF of the filter differed across the two presentation intervals by Δ_{CF}, again expressed as a percentage of the lower CF, and the two CFs were geometrically centered around the nominal CF after roving. See Fig. 1 for a schematic diagram of changes in stimuli.

2. Procedure

Prior to running the experiment, subjects were given basic definitions of pitch and timbre; pitch was related to notes on a musical scale, and timbre was related to sound quality differences between different musical instruments using adjectives such as bright or dull. For comparison, they were told that a saxophone has a brighter timbre than a grand piano. Not surprisingly, subjects often had more difficulty grasping the concept of timbre, but were encouraged to use the practice runs and feedback to get a sense for what a brighter timbre sounded like, relative to a duller timbre. Subjects were tested individually in double-walled sound-attenuating chambers. The subjects' preliminary tasks were to compare tone pairs differing in either F0 or spectral centroid (i.e., "pitch" or "timbre"). In each trial, subjects were played two complex harmonic tones separated by a silent interstimulus interval of 300 ms. The task was to determine which of the two tones had the higher pitch or brighter timbre. The order of the tone presentations was random, with the higher pitch (or timbre) being equally likely to be presented in the first or second interval. Two virtual buttons were displayed on a computer screen, which lit up with each corresponding tone. Subjects could select a button with the computer mouse or by pressing "1" or "2" on the keyboard, corresponding to the "1" and "2" displayed on the virtual buttons. Immediate feedback was provided after each trial, stating if the selection was "correct" or "wrong."

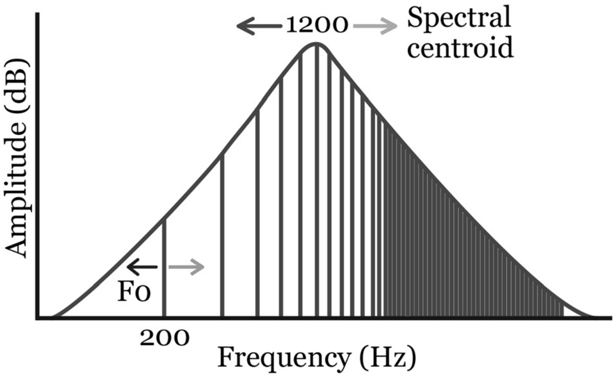

FIG. 1. Schematic diagram of the stimuli used in this study (plotted on log–log axes). Changing the F0 results in changes in the frequencies of the harmonics (represented by the vertical lines). Changing the center frequency of the filter results in changes in the spectral envelope of the sound and hence changes in the amplitudes (but not frequencies) of the harmonics.

Each participant's DLs for F0 and spectral centroid were obtained using a standard two-alternative forced-choice procedure with a two-down one-up adaptive tracking rule that tracks the 70.7% correct point on the psychometric function (Levitt, 1971). The starting value of Δ_{F0} or Δ_{CF} was 200%. Initially, Δ_{F0} or Δ_{CF} was increased or decreased by a factor of 2. After the first reversal in the direction of the change in the tracking variable from "up" to "down," the factor was decreased to 1.26. After two more reversals, the factor was decreased to 1.12, which was the final step size. The run was terminated after six reversals at the final step size, and the DL in each run was the geometric mean of the value of Δ at those last six reversal points.

The first six runs performed by each subject in each condition were treated as practice. The next six runs in each condition were geometrically averaged to obtain the estimated DL for each subject. Each subject completed all testing in one dimension before proceeding to the other dimension, and the F0 and spectral centroid conditions were completed in counterbalanced order across subjects. Subjects were able to complete Experiment 1 in about 45 min on average, but the time varied for each participant, depending on the number and duration of breaks taken and the amount of time subjects took to make their responses.

3. Subjects

To avoid including subjects with severe F0 discrimination difficulties (Peretz *et al.*, 2009; Semal and Demany, 2006), only subjects whose F0DLs were 6% (about 1 semitone) or better were included in the study. Since we have no estimate of an appropriate cutoff for "poor" spectral centroid discrimination, we did not exclude subjects based on exceeding a specific spectral centroid DL. After several subjects failed to reach the F0DL cutoff in the initial training phase, an additional training protocol was added in which the between-trial roving of F0 or spectral centroid was eliminated. A total of 25 of the 57 subjects tested were given the non-roving practice trials. This appeared to make the task easier, and helped some subjects to subsequently improve their performance in the tasks with between-trial roving. Nevertheless a total of 12 subjects (7 of whom were given the non-roving practice) failed to achieve DLs of 6% or less. Eleven of the 12 disqualified subjects were non-musicians. The remaining 45 subjects (21 musicians and 24 non-musicians) took part in the experiment.

All 45 subjects had normal hearing, defined as audiometric pure-tone thresholds of 20 dB hearing level or better at octave frequencies between 500 Hz and 8 kHz, and were recruited from the University of Minnesota community. Ages ranged from 19 to 59 years (mean 25.3 yr). Twenty-one subjects were categorized as musicians (12 females, 9 males, age range 19–59 years, mean 26.3 yr) with at least 8 years of formal musical training, and 24 were categorized as non-musicians (13 females, 11 males, age range 19–34 years, mean 24.4 yr), with 2 or less years of formal musical training. All protocols were approved by the University of Minnesota Institutional Review Board, and all subjects provided written informed consent.

C. Results

The results for musicians and non-musicians are shown in Fig. 2. The average F0DL for musicians was 0.8%, whereas the non-musicians had an average F0DL of 1.9%. Musicians had an average spectral-centroid DL of 4.0%, while the non-musicians had an average DL of 5.0%. Mixed-model analyses of variance (ANOVAs) on the log-transformed DLs were used here and throughout this study, with a Greenhouse-Geisser correction for lack of sphericity included where appropriate. A mixed-model ANOVA with a within-subject factor of dimension (F0 vs spectral centroid) and a between-subjects factor of musicianship showed a main effect of dimension $[F(1,43) = 226.72,\ p < 0.0001$, partial $\eta^2 = 0.84]$, a main effect of musicianship $[F(1,43) = 10.91,\ p = 0.002$, partial $\eta^2 = 0.20]$, and an interaction between dimension and musicianship $[F(1,43) = 0.87,\ p < 0.0001$, partial $\eta^2 = 0.26]$.

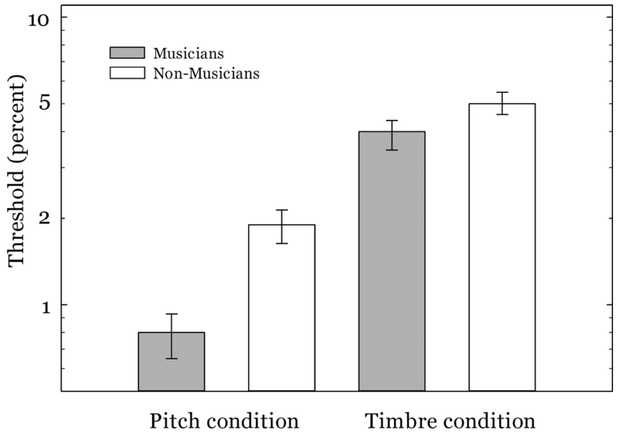

FIG. 2. Results from Experiment 1. Average DLs of musicians and non-musicians on basic pitch and timbre discrimination tasks. Error bars represent +/− one standard error of the mean.

A planned comparison revealed that musicians had significantly better F0DLs compared to non-musicians $[t(43) = 4.05,\ p < 0.0001,\ r = 0.53]$, but no significant difference was found between the groups' spectral centroid DLs $[t(33.7) = 1.36,\ p = 0.183,\ r = 0.23]$. Levene's test indicated unequal variances for the timbre condition $[F = 4.47,\ p = 0.04]$, so degrees of freedom were adjusted from 43 to 33.7 in SPSS (SPSS IBM, New York, NY), using the Welch-Satterthwaite approximation.

D. Discussion

Musicians and non-musicians differed in their F0DLs, but had similar spectral centroid DLs. The differences in basic F0 discrimination with musical training are consistent with previous research that also used subjects with no extensive training (Micheyl *et al.*, 2006). Based on earlier studies, however, we would expect the F0DLs from the non-musicians to converge with those of the musicians after more extensive practice. For instance, Micheyl *et al.* (2006) found that F0DLs from non-musicians reached the levels obtained by professional musicians after about 6 to 8 h of practice, whereas our subjects typically had only around 20 min of practice before data were collected.

The lack of difference between musicians and non-musicians in sensitivity to spectral centroid is also consistent with previous research involving dissimilarity ratings (Caclin *et al.*, 2005; McAdams *et al.*, 1995). The effect of musicianship on F0, but not spectral centroid, may be due to the fact that musicians regularly make fine judgments of pitch differences, for instance when tuning instruments, whereas fine timbre judgments tend to be less critical, since different musical instruments have rather distinct timbres. In addition, pitch changes define melodies, whereas the timbre of a particular instrument generally remains relatively constant. On the other hand, it could be argued that fine timbre discrimination is required when assessing the musical "color" of particular notes or a particular performance.

An alternative explanation as to why musicians did not have better spectral centroid DLs is that the stimuli in this experiment do not sound like musical instruments. These stimuli are synthesized and controlled exclusively by varying the location of the single spectral peak in the stimulus. Thus, it remains possible that musicians are more skilled at discriminating fine timbre differences in more natural musical sounds, perhaps even related to their own instrument. This idea is supported by previous research (Crummer *et al.*, 1994; Pantev *et al.*, 1998).

Finally, a potential limitation of excluding subjects with very poor F0 discrimination is that our population sample may be skewed toward better performance. Had we not excluded these subjects, based on the 6% F0DL cutoff, we would have likely seen a larger difference in F0DLs between the musician and non-musician groups, since 11 of the 12 subjects who were excluded were non-musicians.

III. EXPERIMENT 2: THRESHOLDS AS A FUNCTION OF AMOUNT OF INTERFERENCE

A. Rationale

The aim of Experiment 2 was to investigate the effects of variations in a non-target dimension on discrimination performance in the target dimension. This experiment involved similar stimuli and tasks to those used in Experiment 1, with the addition of random variation in the non-target dimension. Subjects were asked to attend to one dimension while ignoring the other. Shifts in F0 were paired with shifts in spectral centroid, in order to determine the effect of variations in one dimension on subjects' ability to discriminate changes in the other.

B. Methods

1. Stimuli

The stimuli were generated and presented in the same way as in Experiment 1. A standard adaptive two-alternative forced-choice procedure was again used. For this experiment, however, variations in the non-target dimension were introduced in each trial. The amount of variation in the non-target dimension was based on multiples of the DL with no non-target variations, as measured in Experiment 1 for each subject individually (DL_0). Values tested were 0, 2, 5, 10, 25, 50, and $100\mathrm{DL}_0$, where zero indicates a lack of variation (i.e., a repeat of the conditions tested in Experiment 1). As in Experiment 1, the nominal F0 of the harmonic complex was 200 Hz and the nominal CF (spectral centroid) was 1200 Hz. In each trial, both the nominal F0 and the nominal spectral centroid were roved independently by $\pm 10\%$.

2. Procedure

In runs where the F0DL was adaptively tracked, the spectral centroid in each trial differed between the two intervals by a multiple of the centroid DL, as measured individually for each subject in Experiment 1, geometrically centered around the nominal centroid. The interval containing the higher centroid was selected randomly and independently from the F0 in each trial. In runs where the spectral centroid DL was adaptively tracked, the F0 between the two intervals also varied independently in multiples of the individual F0DL around the nominal F0 of 200 Hz, as described in Experiment 1 for the spectral centroid variations. Thus, the random variation in the non-target dimension was uninformative for the subjects' task.

The two parts of the experiment (the F0 task and the spectral centroid task) each contained seven conditions repeated 3 times, totaling 21 runs. The pitch and timbre tasks were performed in counterbalanced order across subjects and all measurements of one dimension were completed before beginning measurements in the other dimension. No practice was given beyond the practice in basic discrimination received in Experiment 1. The basic discrimination tasks in Experiment 1 were performed just prior to starting Experiment 2. Completion of both experiments generally required two sessions, with the first session lasting two hours and the second session (which generally took place within a week of the first session) lasting between one and two hours. Participants were encouraged to take breaks when needed to avoid fatigue effects.

3. Subjects

Thirty listeners took part in this experiment, all of whom had also participated in Experiment 1. Ages ranged from 19 to 59 years (mean 28.0 yr). Fifteen subjects were categorized as musicians (8 females, 7 males, age range 19–59 years, mean 28.5 yr) with at least 8 years of formal musical training, and 15 were non-musicians (9 females, 6 males, age range 19–34 years, mean 24.3 yr), with 2 or less years of formal musical training.

C. Results

The results of Experiment 2 are shown in Fig. 3. A mixed-model repeated-measures ANOVA on the log-transformed DLs was used to analyze the data. Within-subject factors were target dimension (F0 vs spectral centroid) and amount of variation in the non-target dimension. The between-subjects factor was musicianship (musician vs non-musician). Results showed a main effect of target dimension [$F(1,27) = 13.4$, $p = 0.001$, partial $\eta^2 = 0.33$], a main effect of variation in the non-target dimension [$F(6,22) = 18.5$, $p < 0.0001$, partial $\eta^2 = 0.39$], and a main effect of musicianship [$F(1,27) = 5.17$, $p = 0.031$, partial $\eta^2 = 0.16$]. The interaction between musicianship and dimension just failed to reach significance [$F(1,27) = 4.07$, $p = 0.054$, partial $\eta^2 = 0.13$], presumably reflecting the trend for musicians to perform better than non-musicians in the F0 dimension, but not in the spectral centroid dimension. Indeed, separate ANOVAs revealed that musicians were significantly better than non-musicians on the F0 dimension [$F(1,27) = 6.41$, $p = 0.017$, partial $\eta^2 = 0.19$], while they were not significantly better than non-musicians on the spectral centroid dimension [$F(1,27) = 1.82$, $p = 0.188$, partial $\eta^2 = 0.06$]. No other interactions reached significance.

The amount of interference was assessed using the ratio of the DLs between the conditions with variation and the conditions with no variation in the non-target dimension; this measure is referred to as the "interference ratio." The interference ratio at the largest variation level ($100\mathrm{DL}_0$) was 2.8 (i.e., 2.1% divided by 0.76%) and 4.1 (i.e., 14.7% divided by 3.6%) for the musicians and non-musicians, respectively, in the pitch target dimension. The same interference ratios in the

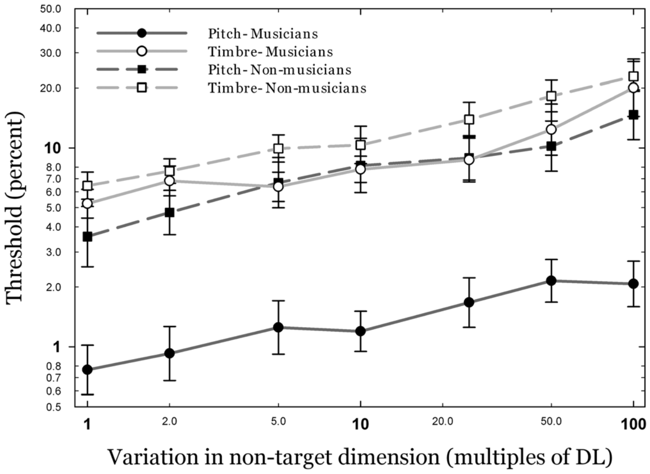

FIG. 3. Results from Experiment 2. Average DLs for musicians and non-musicians on pitch and timbre discrimination tasks are shown as a function of variation in the non-target dimension (in multiples of DL). Error bars represent +/− one standard error of the mean.

timbre target dimension were 3.8 and 3.5 for the musicians and non-musicians, respectively. All four of these represented highly significant increases in DLs, based on paired-samples *t*-tests for F0 [$t(14) = 7.41$, $p < 0.0001$, $r = 0.89$] and spectral centroid [$t(14) = 5.96$, $p < 0.0001$, $r = 0.85$] for musicians, and for F0 [$t(14) = 5.04$, $p < 0.0001$, $r = 0.80$] and spectral centroid [$t(14) = 10.8$, $p < 0.0001$, $r = 0.95$] for non-musicians.

The fact that the original ANOVA found no significant interaction between musicianship and amount of variation in the non-target dimension suggests that the effect of interference was similar for both groups. This was confirmed in a new mixed-model ANOVA with the interference ratio as the dependent variable, target dimension and amount of non-target variation as the within-subject factors, and musicianship as the between-subjects factor. The results showed a significant main effect of non-target variation [$F(5,92.3) = 39.8$, $p < 0.0001$, partial $\eta^2 = 0.59$], but no main effect of the target dimension [$F(1,28) = 0.63$, $p = 0.434$, partial $\eta^2 = 0.02$], no main effect of musicianship [$F(1,28) = 1.24$, $p = 0.274$, partial $\eta^2 = 0.04$], and no significant interactions ($p > 0.24$ in all cases). This outcome confirms that the interference was similar for both pitch and timbre target dimensions, and that both musicians and non-musicians experienced similar amounts of interference in both dimensions.

D. Discussion

Variations in the non-target dimension led to increased (poorer) DLs in the target dimension for both F0 and spectral centroid, and for both musicians and non-musicians. The amount of interference (defined as the ratio between DLs with and without non-target variation) increased with increasing amount of variation, up to the maximum tested ($100\mathrm{DL}_0$).

Although musicians had generally lower F0DLs, their spectral centroid DLs were similar to those of non-musicians, as found in Experiment 1. The effect of variations in both non-target dimensions was not significantly different for musicians and non-musicians, suggesting that musicians are as susceptible to interference due to random stimulus variations as non-musicians. For both groups, when the variations were equated in terms of DL_0, the effects of F0 variation on spectral centroid discrimination and the effects of spectral centroid variation on F0 discrimination were symmetric—random variations in both dimensions produced substantial and similar interference. Thus, our results provide further support for the idea that pitch does not occupy a privileged position in auditory perception, once differences in basic discrimination are equated (McDermott and Oxenham, 2008; McDermott *et al.*, 2010).

IV. EXPERIMENT 3: CONGRUENT AND INCONGRUENT INTERFERENCE

A. Rationale

In Experiment 2, the direction of the variation in the non-target dimension was randomly selected on each trial and was independent of the direction of the change in the target dimension. Thresholds were determined using an adaptive procedure and no attempt was made to separate trials with "congruent" motion (i.e., F0 and spectral centroid changed in the same direction) from trials with "incongruent" motion (i.e., F0 and spectral centroid changed in opposite directions). The interference produced by changes in the non-target dimension may reflect a "distraction" effect (Moore and Glasberg, 1990) produced by any task-irrelevant change, or it may reflect a partial inability on the part of subjects to distinguish between a change in timbre (i.e., higher brightness with increasing spectral centroid) from a change in pitch (i.e., higher pitch with increasing F0; e.g., Russo and Thompson, 2005). It is also possible, in instances with large timbre variation, that an upward shift in spectral centroid induces an "octave error" (e.g., Robinson, 1993), causing subjects to perceive the pitches an octave higher than the stimulus F0.

For this experiment, a method of constant stimuli was used. Congruent trials were randomly interleaved with incongruent trials, but the two categories were analyzed separately to determine whether changes in the non-target dimension produced systematic biases in responses to the target dimension. Only relatively small variations in the dimensions were tested, making octave errors due to large spectral shifts less likely.

A second open question from Experiment 2 is whether multiples of DL_0 provide an appropriate scale along which to equate the perceptual salience of larger changes. If equal changes in terms of DL_0 result in equal salience, then presenting changes in both dimensions that are equal in terms of DL_0 should result in equal performance in both dimensions. The current experiment tested this hypothesis by presenting pairs of tones that varied in F0 and spectral centroid by the same amount, in terms of the individual DL_0s; the task varied (subjects were asked to judge either the pitch or timbre), but the stimuli were identical in the two conditions.

B. Methods

1. Stimuli and procedure

The method in which the stimuli were generated and presented was the same as that used in Experiments 1 and 2. However, this experiment used a method of constant stimuli rather than an adaptive procedure. The subjects were presented with pairs of tones that varied in both F0 and spectral centroid by the same amount, in terms of the individual DL_0s, which had been determined in Experiment 1. The following five multiples of DL_0 were tested: 0.5, 1, 2, 3, and 5. Each trial had a pair of stimuli, as described in Experiment 2, in which both the F0 and spectral centroid varied by one of the multiples of DL_0. In each block of 50 trials, half the trials had congruent pairings (F0 and spectral centroid changed in the same direction) and the other half had incongruent pairings (F0 and spectral centroid varied in opposite directions). Thus, each block included five repetitions of each condition and pairing type. The trials were evenly divided into separate blocks in which either pitch or timbre discrimination was measured.

As in the previous two experiments, subjects were instructed to select which pitch was higher or which timbre was brighter in the tone pair, depending on which task they were performing, and were instructed to ignore the other dimension. A total of 10 blocks were run for each dimension, meaning the estimate of performance for each subject on each dimension was based on a total of 500 trials (100 trials per DL_0 multiple). Feedback was provided after each trial. Each subject completed all the measurements in one dimension before the other dimension was tested, and the order of presentation was counterbalanced across subjects. The experiment took around an hour to complete, but the time varied for each participant, depending on the number and duration of breaks taken and the amount of time subjects took to make their responses.

2. Subjects

A total of 20 subjects participated; all of whom also took part in Experiment 1. Five of these 20 subjects (4 musicians, 1 non-musician) also participated in Experiment 2. The ages of the subjects ranged from 20 to 59 years (mean 25.9 yr). Ten subjects were categorized as musicians (six females, four males, age range 20–59 years, mean 27.2 yr) with at least eight years of formal musical training, and ten were non-musicians (four females, six males, age range 21–34 years, mean 24.6 yr), with two or less years of formal musical training.

C. Results

The mean results in the different conditions for congruent and incongruent trials are shown in terms of proportion correct for musicians and non-musicians in the right and left panels of Fig. 4, respectively. Statistical analysis was performed on values of d', converted from proportion correct by assuming unbiased responding to the first and second intervals in each trial (Hacker *et al.*, 1979). To avoid infinite values of d' when 100% correct performance was achieved, a small correction factor was included that effectively limited the maximum value of d' to 4.65 corresponding to a proportion correct of about 99.95%.

A mixed-model ANOVA was performed on the d' values with within-subject factors of target dimension (F0 or spectral centroid), congruence (congruent or incongruent changes between intervals), amount of variation (0.5 through $5DL_0$), and a between-subjects factor of musicianship. A significant main effect of congruence was found [$F(1,18) = 66.9$, $p < 0.0001$, partial $\eta^2 = 0.79$], reflecting the observation that performance was generally better in congruent than in incongruent trials. The main effect of amount of variation was also significant [$F(2.23,40.2) = 108$, $p < 0.0001$, partial $\eta^2 = 0.86$, $\varepsilon = 0.56$], reflecting the observation that performance improved as the size of the F0 or spectral-centroid difference increased. Finally, the main effect of target dimension (F0 or spectral centroid) was not significant [$F(1,18) = 0.04$, $p = 0.847$, partial $\eta^2 = 0.002$], showing that overall levels of performance were similar in the two dimensions.

A significant interaction between the amount of variation and congruence was also found [$F(2.35,42.3) = 7.78$, $p < 0.0001$, partial $\eta^2 = 0.302$, $\varepsilon = 0.59$], possibly reflecting the widening gap between the congruent and incongruent performance with increasing amount of variation. Additionally, a significant interaction was found between dimension and congruence [$F(1,18) = 6.77$, $p = 0.018$, partial $\eta^2 = 0.273$], indicating that congruence differentially affected F0 and spectral centroid, with the congruence effect being larger when the target dimension was pitch than when it was timbre. However, performance in congruent trials was significantly higher than performance in incongruent trials for both F0 [$F(1,18) = 43.7$, $p < 0.0001$, partial $\eta^2 = 0.71$] and spectral centroid [$F(1,18) = 49.8$, $p < 0.0001$, partial $\eta^2 = 0.73$].

There was a significant effect of musicianship [$F(1,18) = 8.03$, $p = 0.011$, partial $\eta^2 = 0.309$], and a significant interaction between amount of variation and musicianship [$F(4,72) = 4.44$, $p = 0.003$, partial $\eta^2 = 0.198$]. These effects seem to reflect the somewhat worse performance of non-musicians, particularly at larger levels of variation. No significant interaction was found between dimension and musicianship [$F(1,18) = 2.28$, $p = 0.148$, partial $\eta^2 = 0.112$], indicating that the two groups performed similarly across the F0 and spectral centroid conditions. Additionally, no significant interaction was found between congruence and musicianship [$F(1,18) = 0.30$, $p = 0.591$, partial $\eta^2 = 0.016$], suggesting that these groups were similarly affected by

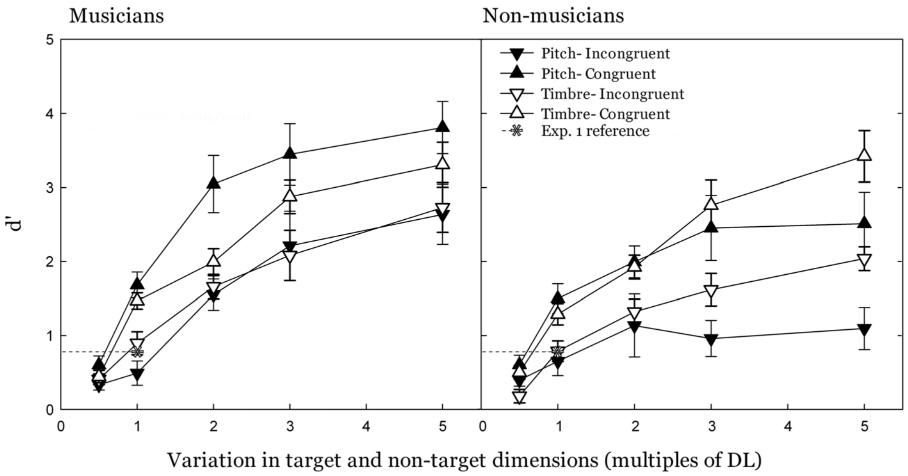

FIG. 4. Experiment 3: Values of d' are shown for congruent and incongruent stimulus pairings for pitch and timbre tasks as a function of amount of variation in target and non-target dimensions (in multiples of DL). Musicians' scores are shown in the left panel, and non-musicians' scores are shown in the right panel. The asterisk in each panel is shown at the point corresponding to the DL in Experiment 1. Error bars represent +/− one standard error of the mean.

whether the dimensions were congruent or incongruent. There was one significant three-way interaction for dimension, variation, and musicianship [$F(1,18) = 0.30$, $p = 0.024$, partial $\eta^2 = 0.143$], suggesting that the groups may be differentially affected by the amount of variation across dimensions. However, the three-way interaction for congruence, dimension, and musicianship was not significant ($p = 0.167$), suggesting the two groups were similarly affected by congruence across the dimensions. There was no significant four-way interaction.

An asterisk connected to a horizontal dashed line in each panel of Fig. 4 is shown at the point corresponding to the DL in Experiment 1. By definition, based on our tracking procedure, the DL was 70.7%, which in a two-interval two-alternative forced-choice task corresponds to a d' of about 0.77 (Hacker *et al.*, 1979). The asterisks fall closer to the downward than to the upward triangles, suggesting at face value that performance was enhanced in the congruent trials, but not degraded in the incongruent trials, relative to no variation. However, this outcome may be related to improvements with practice as all the subjects through necessity participated in Experiment 1 (asterisks) before embarking on Experiment 3. Thus, without this potential confound, it may be expected that congruence would lead to improved performance, whereas incongruence would lead to poorer performance, relative to no irrelevant changes.

D. Discussion

The first important finding from this experiment is that performance in the congruent trials (where the variation in the non-target dimension was in the same direction as that in the target dimension) was better than performance in incongruent trials. This outcome suggests that variations in the non-target dimension did not just provide a distraction, but were confused to some extent with changes in the target dimension. This confusion could be of at least two types. The first possibility is that the two dimensions are not perceptually separable, and that a change in spectral centroid may induce a change in the pitch percept (and vice versa). This seems unlikely, as pitch-matching experiments using harmonic stimuli with widely different spectral content have not shown large or systematic biases in pitch away from the underlying F0 (Oxenham *et al.*, 2011; Walliser, 1969). The second, and more plausible, possibility is that changes in F0 and spectral centroid elicit changes in pitch and timbre, respectively, but that subjects sometimes confuse the two, and therefore respond to the inappropriate dimension. When the dimensions change in a congruent manner, an inappropriate response will still be correct, thereby leading to higher performance in the congruent than in the incongruent trials. This would suggest the confusion is more post-sensory, which aligns with the conclusions of Silbert *et al.* (2009). Nevertheless, as variations in both F0 and spectral centroid elicit changes along the tonotopic dimension in the auditory periphery, there remains a possible basis for sensory confusion.

The second important finding is that overall performance in the F0 and spectral centroid discrimination tasks (averaged across congruent and incongruent conditions) was similar when variations in the two dimensions were equated in terms of multiples of DL_0 for each dimension separately. This finding suggests that salience (and coding accuracy) in the two dimensions may be equated using basic discrimination thresholds, at least for differences up to multiples of $5DL_0$. However, performance was not identical, as indicated by the significant interaction of dimension and congruence, suggesting that equivalence only holds when both congruent and incongruent trials are employed in roughly equal measure. In addition, we cannot rule out the possibility that more differences might be revealed through the use of other measures, such as reaction time.

The third important finding is that musicians and non-musicians showed similarities in terms of overall performance on the pitch and timbre tasks, as well as similarities in how they were affected by congruence. The main effect of musicianship and the interaction with amount of variation reflect some differences between the groups, but the general pattern of results was quite similar. Taken together with the results from Experiment 2, where no significant effect of musicianship was found on the amount of interference, the outcome suggests that musicians' superior analytic listening ability, as demonstrated in an informational masking task that involves attending to one frequency while ignoring others (Oxenham *et al.*, 2003), does not extend to attending to one perceptual dimension while ignoring another.

Finally, it is worth noting that any differences observed between groups may depend to some extent on how the groups are defined. Although many studies have compared the performance of musicians and non-musicians, there are no uniform criteria that are used to distinguish between the two groups. We defined musicians as those with at least eight years of formal musical training; however, no ear-training test was used to verify musical ability (e.g., Oxenham *et al.*, 2003), no requirement was made that they were currently active musicians, and there was no maximum age allowed by which musical training should have commenced. Similarly, although non-musicians were defined as those with two years or less of formal training, it is possible that at least some members of this group had informal experience with listening to or performing music. Thus, as with any study comparing these two groups, the conclusions are qualified by the specific definitions of musical training used here.

V. CONCLUSIONS

DLs for F0 and spectral centroid (perceptually, pitch and timbre) were measured in groups of listeners with and without musical training in a two-alternative forced-choice paradigm. The following results were obtained:

(1) In line with earlier studies, F0DLs were better in musicians than in untrained listeners without musical training. However, DLs for spectral centroid were not significantly different between the two groups.

(2) Discrimination thresholds in either F0 or spectral centroid were impaired by random variations in the non-target dimension. The amount of interference was similar for the

two dimensions and for both musicians and non-musicians.

(3) Performance was better when the interference varied coherently with the target (i.e., both F0 and spectral centroid increased from the first to the second interval) than when it varied in the opposite direction. This outcome suggests that listeners sometimes confused changes across the two dimensions. Musicians were no less susceptible to this "confusion" than non-musicians.

Overall the results provide evidence that judgments in pitch and timbre (in terms of F0 and spectral centroid, respectively) are similarly affected by random variations in the other dimension, suggesting relatively symmetric processes. In addition, musical training does not appear to provide strong immunity from interference effects in either dimension.

ACKNOWLEDGMENT

The work was supported by National Institutes of Health Grant No. R01 DC005216.

ANSI (**1994**). S1.1-1994, *American Standard Acoustical Terminology (R2004)* (American National Standards Institute, New York, 1960).

Beal, A. L. (**1985**). "The skill of recognizing musical structures," Mem. Cognit. **13**, 405–412.

Borchert, E. M. O., Micheyl, C., and Oxenham, A. J. (**2011**). "Perceptual grouping affects pitch judgments across time and frequency," J. Exp. Psychol. Hum. Percept. Perform. **37**, 257–269.

Burns, E. M. (**1999**). "Intervals, scales, and tuning," in *Psychology of Music*, 2nd ed. (Academic, San Diego), pp. 215–264.

Caclin, A., McAdams, S., Smith, B. K., and Winsberg, S. (**2005**). "Acoustic correlates of timbre space dimensions: A confirmatory study using synthetic tones," J. Acoust. Soc. Am. **118**, 471–482.

Crummer, G. C., Walton, J. P., Wayman, J. W., Hantz, E. C., and Frisina, R. D. (**1994**). "Neural processing of musical timbre by musicians, nonmusicians, and musicians possessing absolute pitch," J. Acoust. Soc. Am. **95**, 2720–2727.

Fastl, H., and Zwicker, E. (**2007**). *Psychoacoustics: Facts and Models* (Springer, Berlin), pp. 239–246.

Faulkner, A. (**1985**). "Pitch discrimination of harmonic complex signals: Residue pitch or multiple component discriminations?," J. Acoust. Soc. Am. **78**, 1993–2004.

Green, D. M., and Swets, J. A. (**1966**). *Signal Detection Theory and Psychophysics* (Wiley, New York), Vol. 1, pp. 1–505.

Grey, J. M. (**1977**). "Multidimensional perceptual scaling of musical timbres," J. Acoust. Soc. Am. **61**, 1270–1277.

Hacker, M. J., Ratcliff, R., Tables, P. B., and Ed, J. A. S. (**1979**). "A revised table of d' for M-alternative forced choice," Percept. Psychophys. **26**, 168–170.

Krumhansl, C. L., and Iverson, P. (**1992**). "Perceptual interactions between musical pitch and timbre," J. Exp. Psychol. Hum. Percept. Perform. **18**, 739–751.

Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.

Marozeau, J., and de Cheveigné, A. (**2007**). "The effect of fundamental frequency on the brightness dimension of timbre," J. Acoust. Soc. Am. **121**, 383–387.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.

McDermott, J. H., Keebler, M. V., Micheyl, C., and Oxenham, A. J. (**2010**). "Musical intervals and relative pitch: Frequency resolution, not interval resolution, is special," J. Acoust. Soc. Am. **128**, 1943–1951.

McDermott, J. H., Lehr, A. J., and Oxenham, A. J. (**2008**). "Is relative pitch specific to pitch?," Psychol. Sci. **19**, 1263–1271.

McDermott, J., and Oxenham, A. (**2008**). "Music perception, pitch, and the auditory system," Curr. Opin. Neurobiol. **18**, 452–463.

Melara, R. D., and Marks, L. E. (**1990**). "Interaction among auditory dimensions: Timbre, pitch, and loudness," Percept. Psychophys. **48**, 169–178.

Melara, R. D., and Mounts, J. R. (**1993**). "Selective attention to Stroop dimensions: Effects of baseline discriminability, response mode, and practice," Mem. Cognit. **21**, 627–645.

Melara, R. D., and Mounts, J. R. (**1994**). "Contextual influences on interactive processing: Effects of discriminability, quantity, and uncertainty," Percept. Psychophys. **56**, 73–90.

Micheyl, C., Delhommeau, K., Perrot, X., and Oxenham, A. J. (**2006**). "Influence of musical and psychoacoustical training on pitch discrimination," Hear. Res. **219**, 36–47.

Moore, B. C. J., and Glasberg, B. R. (**1990**). "Frequency discrimination of complex tones with overlapping and non-overlapping harmonics," J. Acoust. Soc. Am. **87**, 2163–2177.

Oxenham, A. J., Fligor, B. J., Mason, C. R., and Kidd, G. (**2003**). "Informational masking and musical training," J. Acoust. Soc. Am. **114**, 1543–1549.

Oxenham, A. J., Micheyl, C., Keebler, M. V., Loper, A., and Santurette, S. (**2011**). "Pitch perception beyond the traditional existence region of pitch," Proc. Natl. Acad. Sci. U.S.A. **108**, 7629–7634.

Pantev, C., Oostenveld, R., Engelien, A., Ross, B., Roberts, L. E., and Hoke, M. (**1998**). "Increased auditory cortical representation in musicians," Nature **392**, 811–814.

Peretz, I., Brattico, E., Järvenpää, M., and Tervaniemi, M. (**2009**). "The amusic brain: In tune, out of key, and unaware," Brain **132**, 1277–1286.

Pitt, M. (**1994**). "Perception of pitch and timbre by musically trained and untrained listeners," J. Exp. Psychol. Hum. Percept. Perform. **20**, 976–986.

Plomp, R. (**1970**). "Timbre as a multidimensional attribute of complex tones," in *Frequency Analysis and Periodicity Detection in Hearing*, edited by R. Plomp, G. F. Plomp, and G. F. Smoorenburg (Sijthoff, Leiden), pp. 397–411.

Robinson, K. (**1993**). "Brightness and octave position: Are changes in spectral envelope and in tone height perceptually equivalent?," Contemp. Music Rev. **9**, 83–95.

Russo, F. A., and Thompson, W. F. (**2005**). "An interval size illusion: The influence of timbre on the perceived size of melodic intervals," Percept. Psychophys. **67**, 559–568.

Semal, C., and Demany, L. (**2006**). "Individual differences in the sensitivity to pitch direction," J. Acoust. Soc. Am. **120**, 3907–3915.

Shepard, R. N. (**1982**). "Geometrical approximations to the structure of musical pitch," Psychol. Rev. **89**, 305–333.

Silbert, N. H., Townsend, J. T., and Lentz, J. J. (**2009**). "Independence and separability in the perception of complex nonspeech sounds," Atten. Percept. Psychophys. **71**, 1900–1915.

Singh, P. G., and Hirsh, I. J. (**1992**). "Influence of spectral locus and F0 changes on the pitch and timbre of complex tones," J. Acoust. Soc. Am. **92**, 2650–2661.

Walliser, K. (**1969**). "Zusammenhänge zwischen dem Schallreiz und der Periodentonhöhe" ("Relationships between the sound stimulus and the pitch"), Acustica **21**, 319–328.

Warrier, C. M., and Zatorre, R. J. (**2002**). "Influence of tonal context and timbral variation on perception of pitch," Percept. Psychophys. **64**, 198–207.