Perception & Psychophysics
2006, 68 (5), 829-844

Acuity of auditory images in pitch and time

PETR JANATA
University of California, Davis, California
and Dartmouth College, Hanover, New Hampshire

and

KAIVON PAROO
Dartmouth College, Hanover, New Hampshire

We examined the pitch and temporal acuity of auditory expectations/images formed under attentional-cuing and imagery task conditions, in order to address whether auditory expectations and auditory images are functionally equivalent. Across three experiments, we observed that pitch acuity was comparable between the two task conditions, whereas temporal acuity deteriorated in the imagery task. A fourth experiment indicated that the observed pitch acuity could not be attributed to implicit influences of the primed context alone. Across the experiments, image acuity in both pitch and time was better in listeners with more musical training. The results support a view that auditory images are multifaceted and that their acuity along any given dimension depends partially on the context in which they are formed.

Although mental images are likely to play a central role in perception (Neisser, 1976), the issue of mental imagery is usually ignored in investigations of selective attention, perceptual acuity, or various forms of memory. However, most experiments on cognitive and perceptual processes rely on tasks in which reference mental images are formed and maintained in working memory over some period of time before they are compared with sensory input. For example, selective attention tasks often require the detection of infrequent deviants that occur in an attended stream. Mental images may be formed of either the deviant or the standard stimulus (or both) so that incoming stimuli can be categorized accordingly. Similarly, tasks that are used to examine the ability to maintain features of a reference item in working memory across a delay period or a series of distractor items also rely on the maintenance of a mental image of the reference item until the time at which a probe item is presented that either matches or differs from the reference item. Despite parsimonious unifying viewpoints (e.g., Neisser, 1976), the interplay between imagery, attention, perception, and memory remains to be clarified.

Perhaps the differences in the dominant paradigms and tasks have impeded the search for a common neural basis of mental image formation. Studies that directly address the issue of mental imagery generally seek to answer some combination of two questions. First, does the *process* of forming and maintaining mental images depend on resources associated with other cognitive functions, such as attention or working memory? Second, how well does the *form* of mental images match corresponding percepts? Both process and form are addressed in studies that examine whether imagery is modality specific and depends on the same neural substrates as perception (for a review, see Kosslyn & Thompson, 2000). Behavioral, neuropsychological, and neuroimaging findings across many visual imagery experiments overwhelmingly suggest that the properties of mentally generated visual images in imagery tasks parallel the properties of perceived stimuli and recruit the same temporal and occipital brain areas that are necessary for visual perception (for reviews, see Farah, 2000; Kosslyn & Thompson, 2000).

The literature on auditory imagery is sparse in comparison with the visual imagery literature, although many similar conclusions have been reached. For example, the process of imagining notes in familiar and unfamiliar melodies recruits the secondary auditory cortex in the temporal lobe, as has been shown using PET (Zatorre, Halpern, Perry, Meyer, & Evans, 1996), fMRI (Halpern & Zatorre, 1999; Kraemer, Macrae, Green, & Kelley, 2005), and ERPs (Janata, 2001), suggesting shared neural substrates for auditory perception and imagery. Working memory experiments have shown that the maintenance of auditory images for tones in working memory is impaired by distractor items (Deutsch, 1970) and the amount of attention that is paid to the distractors, with the greatest deficits arising for tonal distractors, in comparison with verbal or visual distractors (Pechmann & Mohr, 1992). Explicit rehearsal of the tone across a period of distractors strengthens its memory trace, relative to trials in which attention and working memory are also directed toward a distractor task (Keller,

This research was supported by National Institutes of Health Grant R03 DC05146 to P.J. We thank Sonja Rakowski for help in collecting data and Mari Riess Jones, Sean Fannon, Amrita Puri, and two anonymous reviewers for helpful comments on earlier versions of the manuscript. Correspondence concerning this article should be addressed to P. Janata, Center for Mind and Brain, University of California, One Shields Avenue, Davis, CA 95616 (e-mail: pjanata@ucdavis.edu).

 Copyright 2006 Psychonomic Society, Inc.

Cowan, & Saults, 1995). Although the working memory experiments cited are not explicitly auditory imagery experiments, they suggest that attention, working memory, and imagery are intimately intertwined (Kalakoski, 2001).

Auditory images also appear to have "perceptual" properties, in that they influence the perceptual processing of auditory stimuli. The interactions between auditory mental images and percepts have been examined behaviorally in a variety of ways and to varying degrees of precision. Using a signal detection task, Farah and Smith (1983) found that the intensity threshold for detecting a pure tone in noise was lower when listeners formed an image of the target tone (e.g., 715 Hz) either before or during the observation interval than when they formed an image of a different tone (e.g., 1000 Hz). They concluded that forming the mental image drew attention to the appropriate frequency, thus facilitating perceptual processing. Similarly, forming an expectancy that a 20-msec target tone will occur at a specific time following a cue facilitates the tone's detection, in comparison with when the target unexpectedly occurs 150 msec before or 100 msec after the expected time (Wright & Fitzgerald, 2004). Auditory images are also spectrotemporally complex, as is illustrated by imagery of timbre (instrument quality). Imagining that a pitch cued with a sine tone has a specific timbre (e.g., guitar) facilitates the judgment that a subsequent target has the same pitch when the target is played with a matching timbre, in comparison with when it is played with a different timbre (Crowder, 1989).

Further evidence that auditory images behave like their perceptual counterparts has come from studies of auditory images in musical contexts. Hubbard and Stoeckig (1988) asked listeners to form a mental image of a chord or tone that was one scale degree above a chord or tone cue and, upon forming the image, to decide whether a target consisted of the same pitch(es) as their image. As in the studies cited above, the mental image facilitated the decision to respond *same* to a target that was the same as the image. Moreover, the response times and accuracy to different chords depended on their harmonic proximity to the imagined chord, with a closely related target chord showing more interference effects than a distantly related chord. Thus, an imagined chord primed other chords in harmonic space in the same way that an externally presented chord would have.

Coarse temporal aspects of a musical stimulus are also preserved in mental images. When subjects make judgments about the relative pitch height of two lyrics in familiar songs, the time required to make those judgments increases with increasing distance (measured in beats) between those lyrics in the melody, as though the subjects imagine the melody in order to make the judgment (Halpern, 1988). The interplay of attention, expectation, timing, and imagery is also evident in studies that manipulate how attention is aimed in pitch and time (e.g., Dowling, Lung, & Herrbold, 1987; Jones & Boltz, 1989). For example, the temporal evolution of imagined endings can be shaped by manipulating the rhythmic and tonal accent structure of the preceding melodic phrase (Jones & Boltz, 1989).

With the present set of four experiments, we extend previous behavioral studies of auditory imagery in two ways. First, we investigate the precision of auditory images by measuring their tuning in pitch and time. Second, we make within-subjects comparisons of the acuity of the images formed in three different task contexts. The first task can be thought of in the context of attentional cuing. In Experiments 1 and 2, attention is guided by the initial seven notes of an ascending major scale that collectively generate a strong expectation to hear the tonic, the most stable note of the key after which the scale is named (Dowling et al., 1987; Jones, 1981). In our case, the tonic serves as the reference point about which a tuning or timing judgment is made. In Experiment 3, the tuning and timing judgments are made for the leading tone. The leading tone is the seventh note in the scale and is less perceptually stable and characteristic of the key than is the tonic.

The second task is explicitly an imagery task. Subjects are instructed to listen to the initial three to five notes of the scale and imagine the remaining notes prior to making the same tuning/timing judgment about a target that occurs at or around the time that the last note in the sequence would have occurred (the tonic in Experiments 1 and 2 and the leading tone in Experiment 3). By pairing a task that examines how auditory attention is deployed to locations in pitch and time with an imagery task in which a greater distance must be traversed along a cognitive map in order to attain the same location in pitch and time, our aim was to establish whether the expectations and the images that exist at the moment that the probe is heard are comparable in terms of their content. Although comparable acuity is not proof that mental images formed in the context of attentional orienting and expectation are the same as those formed in the context of imagery (or recall of note sequences from memory), it seems that comparable acuity is a prerequisite for stipulating that the mental images arrived at via different task instructions are, indeed, identical at the point at which an internally derived mental image is compared with sensory input. It bears pointing out that from the standpoint of Neisser's (1976) perceptual cycle model, our two tasks are functionally identical, with the only difference being that the anticipated sensory event in the imagery task is at the end of a longer chain of images/anticipations.

Finally, in Experiment 4, we utilized a key membership judgment task to compare the image acuity for the tonic versus the leading tone within subjects. Because the task was structured as a discrimination task, as in Experiments 2 and 3, performance on the task would benefit from accurate image formation. However, no imagery instructions were given, and the stimuli were modified slightly to discourage the use of the imagery strategy that had been explicitly encouraged in Experiments 1–3. Thus, this task assessed the dependence of image acuity on the sense of key that was primed by the context notes and allowed us to make two predictions. First, if goal-directed mental imagery imparts no additional clarity to the mental representations of the notes that are compared with the probes,

we would expect that the acuity of the pitch images in Experiment 4 would match the acuity of the pitch images in Experiments 2 and 3. In other words, any acuity would arise from the primed context alone. Second, we predicted that the listeners' judgments would be worse for the leading tone than for the tonic, because the relationships of the tonic and leading tone to the key that is primed by the context are so different (Krumhansl, 1990). Subjectively, the leading tone is judged to fit less well than the tonic into a primed key, and identifying it as a member of the key requires more time (Janata & Reisberg, 1988). We expected that the difference between the leading tone and the tonic would be echoed in our task when the listeners used the primed key information alone but that this deficit would be ameliorated when the judgment occurred in the context of an explicit imagery task.

EXPERIMENT 1

In the first experiment, we compared the attentional-cuing and imagery conditions, using a method of constant stimuli, in order to obtain an initial estimate of the image tuning curves in pitch and time. The dependent variable of interest was the proportion of *congruous*—that is, *in-tune* and *on-time*—responses. The proportion of congruous responses was calculated for each level of deviation in *listen* and *imagine* trial types. In the listen trials, we expected that the proportion of congruous responses would decrease as the amount of deviation increased. We chose the deviation levels so that the proportions would vary across the full range of 0 to 1, but the shapes of the distributions in the listen conditions were not of tremendous interest. Of primary interest was a comparison, between the listen and the imagine conditions, of the shapes of the response distributions across different levels of mistuning. Specifically, we expected that the mental images might be less precise in the imagine conditions because of fewer exogenous cues to guide their formation. Increased blurriness of the mental images would manifest itself as a greater proportion of congruous (in-tune) responses at larger deviations. Similarly, we were interested whether the number of notes to be imagined would influence the shape of the distributions in the imagery conditions. Specifically, we suspected that the blurriness of a mental image might increase when more notes of the scale had to be imagined.

Method

Subjects. Twenty-four Dartmouth College undergraduate students (16 of them female) from the introductory psychology course served as listeners in the experiment in return for partial course credit. Their age was 20.7 ± 2.9 years (mean ± standard deviation). Three of them reported being left-handed. Two reported possessing absolute pitch. Twelve listeners had at least 1 year of formal musical training for voice or instrument, and among those, the amount of training was 7.1 ± 4.1 years. All the listeners provided informed consent in accordance with the guidelines of the Committee for the Protection of Human Subjects at Dartmouth.

Stimuli. Tone sequences were ascending diatonic scales in five major keys beginning on the following tonic notes: C_4 (261.63 Hz), $D\flat_4$ (277.18 Hz), D_4 (293.66 Hz), $E\flat_4$ (311.13 Hz), and E_4 (329.63 Hz). Each note was 250 msec in duration, and the standard stimulus onset asynchrony (SOA) was 600 msec. Each note was synthesized in MATLAB from its fundamental frequency and the next seven higher harmonics. The amplitude, A, of each harmonic was given by the equation

$$A(N) = \frac{e^{1/N^2} - 1}{e - 1},$$

where N is the harmonic number (1 = fundamental frequency). All harmonics had the same starting phase. A 5-msec linear ramp was applied to the envelope at the beginning and end of each note.

The pitch of a final probe note (an octave above the starting note) in the scale could assume one of seven values. It was either in tune or mistuned by −60, −40, −20, +20, +40, or +60 cents. Similarly, the SOA between the penultimate and the final notes in the scale was either the standard SOA or one of six deviant times: −90, −60, −30, +30, +60, or +90 msec. Deviance levels were chosen to fall around discrimination thresholds commonly encountered in the literature. The final note never deviated from the standard either in pitch or in time. A unique sound file, containing the ascending notes and the probe note, was synthesized for each degree of deviation in order to avoid any possibility of timing variability that might have been introduced by the stimulus presentation program when switching from prime to target sound files.

Procedure. The listeners were seated in an double-walled sound-attenuating chamber (Industrial Acoustics Corporation) in front of a computer monitor and made responses on a computer keyboard. Stimuli were presented at a comfortable level through Sony headphones. Presentation software (www.neurobs.com) running on a Dell Latitude (Model C840) under Windows 2000 was used for randomizing and presenting the stimuli and recording responses.

On each trial, two cue words appeared on the screen to indicate the task and trial types to the listener. The top word was either "Time" or "Pitch" and cued one of the two tasks described below. The bottom word cued the trial type and was either "Listen" or "Imagine." Note that throughout the article, we will refer to the dimension that listeners are making judgments on as the task type (time or pitch), and we refer to trial types (conditions) in terms of the mode of image formation (listen or imagine). The cues appeared on the screen 1 sec prior to the onset of the scale. The words remained on the screen throughout the trial. In listen trials, the listeners heard every note of the scale, whereas in imagine trials, they heard either three notes and imagined five (n = 14 listeners) or they heard five and imagined three (n = 10 listeners). A final note was presented in the imagine trials at the time it would have occurred had all the notes been presented as in the listen trials.

The listeners were instructed on the tasks and continued to receive short blocks of practice trials until they felt that they understood both the tasks. The listeners were also instructed to refrain from making any movements or vocalizations that would help them keep time during the imagery trials.

Listen and imagine trials alternated during blocks of trials in pitch and time tasks. Each task block consisted of 60 trials. Two pitch blocks were presented in alternation with two time blocks, and the order was counterbalanced across subjects. In the pitch task, the listeners pressed one of two keys upon hearing the final note in the scale, to indicate whether that note was congruous (in tune) or incongruous (out of tune). Similarly in the time task, the listeners judged whether the final note was congruous (on time) or incongruous (either early or late). In the imagine trials, the listeners indicated whether the final note in the scale was congruous with their mental image of either the pitch or the onset time of that note. In each task, the listeners received 60 listen and 60 imagine trials, of which 50% were congruous (deviations of 0 cents or 0 msec). Thus, each level of deviation away from zero was experienced once for each key in each task. No feedback was given regarding response accuracy.

Upon completing all the trials, the listeners filled out a questionnaire about their musical background and about their perception of the difficulty of the pitch and time tasks and a few general aspects of their mental images. The questions and rating scales are given in the Results section.

Data analysis. The data were analyzed using MATLAB and SAS. Mixed model analyses were implemented using PROC MIXED in SAS for the pitch and time tasks separately (Littell, Milliken, Stroup, & Wolfinger, 1996). The level of deviation, trial type (listen or imagine), and number of imagined notes (three or five) were treated as fixed effects. Level of deviation and trial type were treated as within-subjects repeated measures. The number of imagined notes was a between-subjects variable. The model was estimated using the restricted maximum likelihood method and utilized an unstructured covariance structure, since this covariance structure yielded better fit statistics (e.g., Akaike's information criterion) than did a variance components covariance structure. The Satterwaithe approximation was used to calculate degrees of freedom associated with fixed effects.

Results

Image tuning curves. Figures 1A and 1B show that the likelihood of judging a probe note to be in tune at any given degree of mistuning was comparable when the listeners heard all of the notes leading up to the probe note and when the listeners had to imagine either three or five notes (including the imagined note that was compared against the probe). The statistical analysis indicated that the only significant effect on the proportion of congruous responses was the main effect of level of deviation [$F(6,22) = 44.17$, $p < .0001$]. Not surprisingly, probe notes that were in tune were judged as such 88% and 78% of the time in the heard and imagine conditions, respectively. A probe note that was tuned flat by 20 cents was ambiguous, with ~50% in-tune responses. When the probe was tuned sharp by 20 cents, the proportion of in-tune responses approached 70%, suggesting that the listeners' expectations and images tended to be mistuned upward. The apparent presence of an upward asymmetry in both the Imagine 3 and the Imagine 5 groups of subjects was tested with a contrast of the positive mistunings with the negative mistunings. As a group, positive mistunings were judged congruous 16% more often than were the corresponding group of negative mistunings [$t(1,22) = 5.41$, $p < .0001$]. Of greatest importance was the absence of significant interactions of trial type with level of deviation and trial type with number of imagined notes and of a significant three-way interaction of deviation, trial type, and number of imagined notes, indicating that the images/expectations that the listeners formed for the probe notes were comparable in all the conditions. The only other

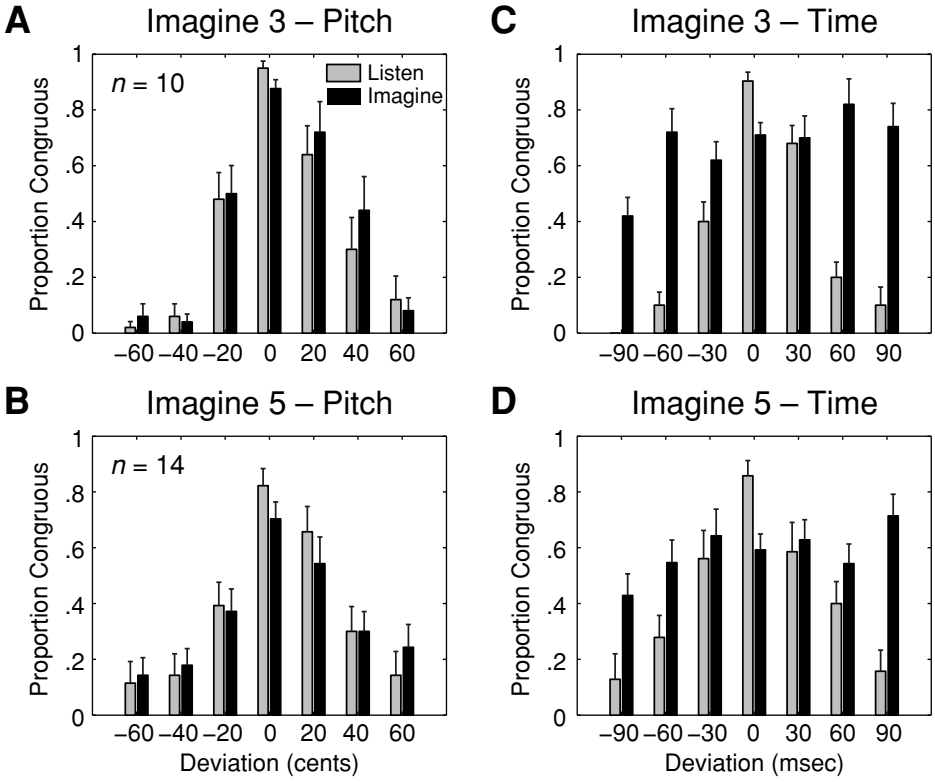

Figure 1. Tuning functions for mental images in pitch and time. The height of each bar in a panel gives the probability that a probe note at a particular deviation will be judged as *in tune* in the pitch task (panels A and B), or *on time* in the time task (panels C and D). Gray bars show the judgments of probe notes in the *listen* conditions when the listeners heard all of the notes leading up to the probe note. Black bars give probabilities for probe notes in the *imagine* conditions when the listeners either heard five notes and imagined three (panels A and C), or heard three notes and imagined five (panels B and D). Error bars indicate 1 *SEM*.

outcome of the statistical analysis was an interaction between level of deviation and number of notes imagined [$F(6,22) = 3.46$, $p = .015$], which was probably due to a greater proportion of congruous judgments in both the listen and the imagine conditions at the extremes of the deviation range (−60, −40, and +60 cents) for the listeners in the Imagine 5 group (Figure 1B). In other words, the listeners in the Imagine 5 condition tended to form less accurate images/expectations overall.

The results from the time task differed markedly from those in the pitch task, as can be seen in Figures 1C and 1D. The requirement to imagine either three or five notes considerably impaired a listener's ability to determine whether a probe note occurred early or late, when compared with having to determine whether the next note in a sequence of notes arrived on time. Whereas the listeners were reliably judging notes with 60-msec deviation as arriving too early or too late in the listen trials, they were uncertain about whether notes were on time throughout the range of −90 to +90 msec in the imagine trials, irrespective of the number of notes to be imagined. These effects manifest themselves as a significant main effect of trial type [$F(1,22) = 42.35$, $p < .0001$] and a significant trial type × level of deviation interaction [$F(6,22) = 24.05$, $p < .0001$]. The main effect of level of deviation was also significant [$F(6,22) = 32.54$, $p < .0001$], as was the interaction of trial type with number of imagined notes [$F(1,22) = 5.22$, $p = .0324$].

Assessments of task difficulty and image clarity. Mixed model analyses of the task difficulty ratings (see Table 1) confirmed that the listeners found trials in which they had to imagine notes in the scale more difficult than those in which they heard all of the notes [pitch, $F(1,21) = 25.46$, $p < .0001$; time, $F(1,21) = 23.16$, $p < .0001$]. Neither the main effect of number of imagined notes nor the interaction of number of imagined notes with trial type was significant for either the pitch or the time task.

Table 1
Subjective Assessments of Task Difficulty and Image Properties in Experiments 1–3

Question	Experiment 1		Experiment 2	Experiment 3
	Imagine 5	Imagine 3	Imagine 3	Imagine 3
How difficult did you find the PITCH task when you HEARD the entire scale (1 = *very easy*; 7 = *very difficult*)?	3.0 ± 1.2	2.3 ± 1.2	3.9 ± 1.3	3.3 ± 1.6
How difficult did you find the PITCH task when you IMAGINED the scale (1 = *very easy*; 7 = *very difficult*)?	4.8 ± 1.4	4.6 ± 1.5	5.0 ± 1.4	4.6 ± 1.8
How difficult did you find the TIME task when you HEARD the entire scale (1 = *very easy*; 7 = *very difficult*)?	3.3 ± 1.5	2.8 ± 1.2	2.9 ± 1.8	3.2 ± 1.4
How difficult did you find the TIME task when you IMAGINED the scale (1 = *very easy*; 7 = *very difficult*)?	5.1 ± 1.5	5.2 ± 1.5	4.2 ± 1.7	5.1 ± 1.9
How often did you find yourself singing along in your mind with the heard scales (1 = *often*; 7 = *not often*)?	1.7 ± 1.0	2.2 ± 2.0	2.8 ± 2.0	2.4 ± 1.2
How vivid were your images of the notes during the imagined scales (1 = *very vivid*; 7 = *had no image*)?	3.6 ± 1.5	3.8 ± 2.0	4.6 ± 1.5	4.2 ± 2.5
Rate the ease with which you could maintain the tempo/rhythm when you were imagining the notes (1 = *very easy*; 7 = *very difficult*).	3.5 ± 1.5	3.8 ± 1.3	3.0 ± 1.2	4.2 ± 1.5

Note—All values are given as mean ± standard deviation.

Discussion

As was expected, the listeners readily detected mistuned or mistimed targets when the target was the next expected note in the scale. Interestingly, when having to image a sequence of notes, their acuity differed dramatically with the type of judgment they were asked to make. Intonation judgments were rendered with an accuracy comparable to that in the attentional-cuing task and were unaffected by the number of notes that had to be imagined. However, timing judgments suffered greatly. In fact, the range of fixed temporal deviations was insufficient to find a Δt that resulted in reliable *incongruous* ratings in the imagine condition. This result, as well as an observation that individual listeners differed considerably in their acuity, prompted us to modify the procedure and estimate individual thresholds.

EXPERIMENT 2

In this experiment, we used a standard psychophysical two-alternative forced choice (2AFC) staircase procedure

for estimating image acuity. This procedure allowed us to estimate each listener's point of subjective equivalence in terms of the offset of the image from perfectly in tune or on time, as well as the sharpness of the image in terms of the difference between the two threshold estimates for each dimension in each condition. Thresholds were estimated simultaneously for both directions of deviation along each dimension—that is, flat and sharp for pitch and early and late for time in both the listen and the imagine conditions.

Method

Subjects. Fourteen Dartmouth College undergraduate students (8 of them female) from the introductory psychology course served as listeners in the experiment in return for partial course credit. Their age was 20.1 ± 1.2 years (mean ± standard deviation). All were right-handed. One reported possessing absolute pitch. Eleven listeners had at least 1 year of formal musical training for voice or instrument, and among those, the amount of training was 7.2 ± 5.1 years. All the listeners provided informed consent in accordance with the guidelines of the Committee for the Protection of Human Subjects at Dartmouth. None of the listeners had participated in the previous experiment.

Stimuli. The stimuli were synthesized with the same parameters as those in Experiment 1. However, only the ascending D-major scale (starting at D_4, 293.67 Hz) was used, and in the imagine conditions, the listeners heard five notes and imagined three. As in Experiment 1, a set of sound files was created, with each sound file corresponding to a level of deviation in either pitch or time. Between −300 and −50 msec, temporal deviation steps were 10 msec. Between −50 and 0 msec, the step size was 5 msec. The same step sizes were used for deviations greater than 0. Pitch deviations ran from −50 to +50 cents in 2.5-cent intervals.

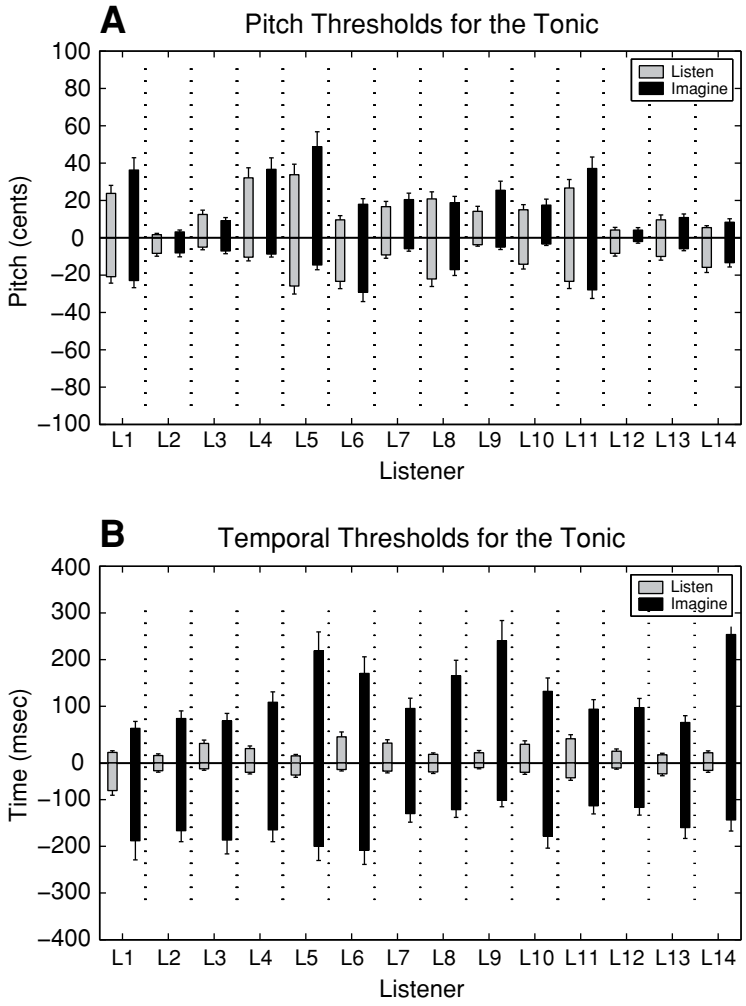

Figure 2. Tuning and timing thresholds for individual listeners judging (A) the intonation of the tonic, which is the final note of an ascending major scale and the most representative note of the key, or (B) the timing of the tonic. Gray bars are data from the listen condition, in which the listeners heard the seven notes of the scale preceding the final note. Black bars are data from the imagine condition, in which listeners heard five notes of the scale and imagined the remaining three. Error bars represent 1 *SEM*, based on the six turnaround points used to calculate each threshold.

Table 2
Image Widths in Experiments 2, 3, and 4

	Pitch		Time	
Experiment	Listen	Imagine	Listen	Imagine
Experiment 2: Eighth note (tonic)	30 ± 15	33 ± 20	51 ± 18	286 ± 71
Experiment 3: Seventh note (leading tone)	30 ± 14	47 ± 24	70 ± 12	264 ± 74
Experiment 4: Tonic		78 ± 40		
Experiment 4: Leading tone		107 ± 32		

Note—All values are given as mean ± standard deviation. Pitch values are given in cents. Time values are given in milliseconds.

Procedure. Thresholds were determined using a 2AFC task. Each trial had the following structure. The listeners saw the cues specifying the type of task (pitch or time judgment) and the mode of imagery (listen or imagine) and heard the first of two sequences. One sequence was the standard sequence in which the eighth note of the ascending scale was in tune and on time. The other was a probe sequence in which the eighth note was either mistuned or mistimed, depending on the task modality of the current trial. As in Experiment 1, the listeners heard all eight notes of the ascending scale in the listen conditions, whereas they heard five and imagined three in the imagine conditions, with a probe note occurring at or around the time at which the eighth note should have occurred (4,200 msec). Immediately upon termination of the first sequence, 1 of 10 different tonal masks was played. The tonal masks were random sequences of four tones within one semitone of the standard probe pitch (587 Hz). Each tone had the same timbre as the notes in the task sequences, was 250 msec in duration, and was played once. The second sequence began 1 sec following the termination of the mask. After hearing the second sequence, the listener had to specify, by pressing one of two buttons, which of the sequences contained the mistuned/mistimed probe note. The order of the standard (in-tune or on-time) sequence and the probe sequence was randomized.

The degree of deviation of the probe note was adjusted using a slightly modified two-down/one-up staircase procedure according to the following rules. The initial deviation was set to the maximum of the range—for example, ±300 msec or ±50 cents. In order to bring the listeners to a range around their thresholds more quickly, the amount of deviation was reduced on each trial by one step until two errors were made. At this turnaround point, the amount of deviation was increased by one step on each trial until a correct response was made. Two successive correct responses resulted in a one-step reduction of the deviation. The procedure continued until 10 turnarounds had been accomplished in each condition.

Conditions were randomly interleaved across trials so that thresholds could be estimated for both listen and imagine in both the time and the pitch tasks in parallel. As in Experiment 1, the listeners were instructed on the tasks and received short blocks of practice trials until they felt that they understood the tasks. The listeners were instructed to refrain from making any movements or vocalizations that would help them keep time during the imagery trials. The listeners were given the opportunity to take a break every 55 trials before continuing. The total number of trials varied across listeners (264–354) but was ~300 on average. The procedure lasted approximately 1.5 h. The listeners then completed a questionnaire, as in Experiment 1.

Data analysis. Acuity thresholds were defined for each subject as the average of the deviation values at turnaround points 5–10. Two properties of the mental images were of particular interest. The first was the width of the mental image. The width was defined as the difference between the positive and the negative thresholds. The second was the offset of the center of the mental image tuning curve, defined as the average of the positive and negative thresholds. The question of whether image acuity, as defined by the width, remained constant between the listen and the imagine conditions was addressed using a paired t test. Similarly, the question of whether the center of the mental image differed between the two conditions was assessed with a paired t test comparing the listen and the imagine conditions. Finally, one-sample t tests were performed separately for the listen and the imagine conditions, testing whether the average center of the image was equal to a deviation of zero or whether the listeners as a population tended to systematically shift their mental images in one direction or another. All t tests were two-tailed.

Results

The width of the pitch component of the images did not differ significantly between the listen and the imagine conditions [$t(1,13) = 1.2$, n.s.; see Figure 2A and Table 2]. The average widths of the pitch images were 30 cents in the listen condition and 33 cents in the imagine condition. Given a fundamental frequency of 587 Hz for the target, these widths correspond to 10.3 Hz in the listen condition and 11.2 Hz in the imagine condition. Thus, the amount of mistuning at threshold in both conditions was about 1% of an in-tune target. The average offset of the image in the imagine condition was 4.3 cents more positive (sharp) than zero [$t(1,13) = 2.5$, $p < .03$; see Table 3]. The offset in the imagine condition was significantly more positive (by 3.4 cents) than was the center of the image in the listen condition [$t(1,13) = 3.6$, $p < .004$]. Figure 2A illustrates that the individual listeners varied in their tendency to imagine sharp or flat. For example, L2's images were very narrow in both the listen and the imagine conditions but

Table 3
Image Offsets in Experiments 2, 3, and 4

	Pitch		Time	
Experiment	Listen	Imagine	Listen	Imagine
Experiment 2: Eighth note	0.9 ± 4.4	4.3 ± 6.2	4 ± 10	9 ± 40
Experiment 3: Seventh note	2.6 ± 9.0	4.4 ± 6.8	9 ± 6	20 ± 34
Experiment 4: Tonic		−0.9 ± 15.0		
Experiment 4: Leading tone		4.0 ± 13.0		

Note—All values are given as mean ± standard deviation. Pitch values are given in cents. Time values are given in milliseconds.

consistently flat, whereas L4's images were much broader and consistently sharp. Remarkably, the images of one listener (L12) were more precise in the imagine condition.

Unlike pitch images, temporal images were much wider in all the listeners when the notes leading up to the probe note had to be imagined than when they were heard (see Figure 2B and Table 2). The average image width increased by 235 ± 22 msec (mean ± standard error) from 51 to 286 msec. Despite the increase in image width, the offset of the image in the imagine condition was neither earlier nor later than that in the listen condition [$t(1,13) = 0.4$, n.s.]. When compared with the canonical onset of the probe note, neither of the measured offsets differed significantly [listen, $t(1,13) = 1.7$, n.s.; imagine, $t(1,13) = 0.9$, n.s.; Table 3]. Thus, although the listeners were much less accurate in judging when the probe note was to occur when asked to imagine the preceding notes, as a group, they did not systematically move their image or expectation earlier or later in time. Nevertheless, inspection of the individual listener data in Figure 2B suggests that the temporal images of some listeners were biased toward earlier times (e.g., L1 and L3), whereas those of others were delayed (e.g., L9 and L14). In addition, there was intersubject variability in the width of the temporal component of the images, with some listeners (e.g., L7 and L12) showing higher acuity than did others (e.g., L5 and L6).

Discussion

The results of Experiment 2 confirmed the observation in Experiment 1 that the temporal component of the images suffered greatly when the listeners were asked to imagine several notes preceding the target event, whereas the images along the pitch dimension remained tightly focused. The temporal thresholds in the listen condition are in direct agreement with thresholds for detecting deviations from isochrony at similar interstimulus intervals (Ehrle & Samson, 2005).

EXPERIMENT 3

In order to facilitate image formation in the present set of experiments, we chose to use ascending scales, because they are relatively simple and familiar, yet they force listeners to move their mental images in pitch space. When ascending scales are used, a convenient location around which to place probes is the tonic, which serves as both the starting note and the final note (an octave above the starting note). The tonic has the property that it is the most representative member of major and minor keys (Krumhansl, 1990) and, when primed with a scale, tends to be categorized most quickly and accurately as belonging to the key (Janata & Reisberg, 1988). Thus, in Experiments 1 and 2, the listeners may have been evaluating the tuning of the probe note relative to a mental image of the tonic that was primed by the heard notes in the scale, rather than arriving at the mental image of the tonic by imagining all of the notes leading up to it. In Experiment 3, we sought to eliminate the potential use of this alternative strategy by terminating the sequence with a probe note on the seventh degree (the leading tone) of the scale. The leading tone is regarded as one of the least characteristic members of a major key, and its representation is, therefore, not strongly primed by the first five notes of the ascending scale. We expected that if the listeners were not forming accurate mental images of each note remaining in the scale but, instead, relying on a primed tonal context, the tuning curve associated with probes around the seventh degree of the scale would be broader and shifted upward in pitch, in comparison with images formed for the eighth and more stable degree of the scale. An upward shift toward the tonic might arise via biasing of the leading tone's representation by the strongly primed tonic to which the leading tone tends to resolve. In other words, if across many trials numerous conflicts arise between the pitch that is to be imagined and its upper neighbor that has a stronger representation based on the context, the outcome might be an image that is both broader and shifted upward.

Method

Subjects. Fourteen members of the Dartmouth community (7 of them female) served as listeners in the experiment and received $10/h for their participation. Their age was 23.1 ± 6.5 years (mean ± standard deviation). All were right-handed. Ten listeners had at least 1 year of formal musical training for voice or instrument, and among those, the amount of training was 8.1 ± 4.7 years. All the listeners provided informed consent in accordance with the guidelines of the Committee for the Protection of Human Subjects at Dartmouth. None of the listeners had participated in either of the previous experiments.

Stimuli. The stimuli were identical to those used in Experiment 2, with the exception that the note sequences ended on the note $C\sharp_5$, the seventh degree of the D-major scale. Thus, probe notes varied in pitch around 554.37 Hz.

Procedure. The procedures used for obtaining pitch and temporal acuity thresholds were the same as those in Experiment 2, and the listeners completed a questionnaire as in Experiment 1.

Data analysis. The data were analyzed as in Experiment 2. In addition, a mixed model analysis was implemented in SAS in order to assess any differences between Experiments 2 and 3 in terms of image widths and offsets. Narrower pitch images in Experiment 2 might indicate that the listeners benefited from a coincidence of the probe pitch with a primed representation of the tonic. Further evidence that the listeners relied on a primed representation of the tonic, instead of generating the correct sequence of mental pitch images, when making probe note judgments would be a dramatic positive offset (on the order of one semitone) of pitch images in Experiment 3. The parameters of the mixed model analysis were the same as those in Experiment 2. One listener's (L13) thresholds were considerably worse than those of all the other listeners in all the conditions (see Figure 3), so this person's data were excluded from the statistical analyses.

Results

In contrast to Experiment 2, pitch images were significantly broader in the imagine condition than in the listen condition [16.5 ± 4.9 cents; $t(12) = 3.35$, $p = .0058$; see Figure 3A and Table 2]. The offset of the pitch image did not differ significantly across the two conditions [1.8 ± 1.9 cents; $t(12) = 0.92$, n.s.], although the image was significantly sharper than zero in both the listen [2.6 ± 0.8

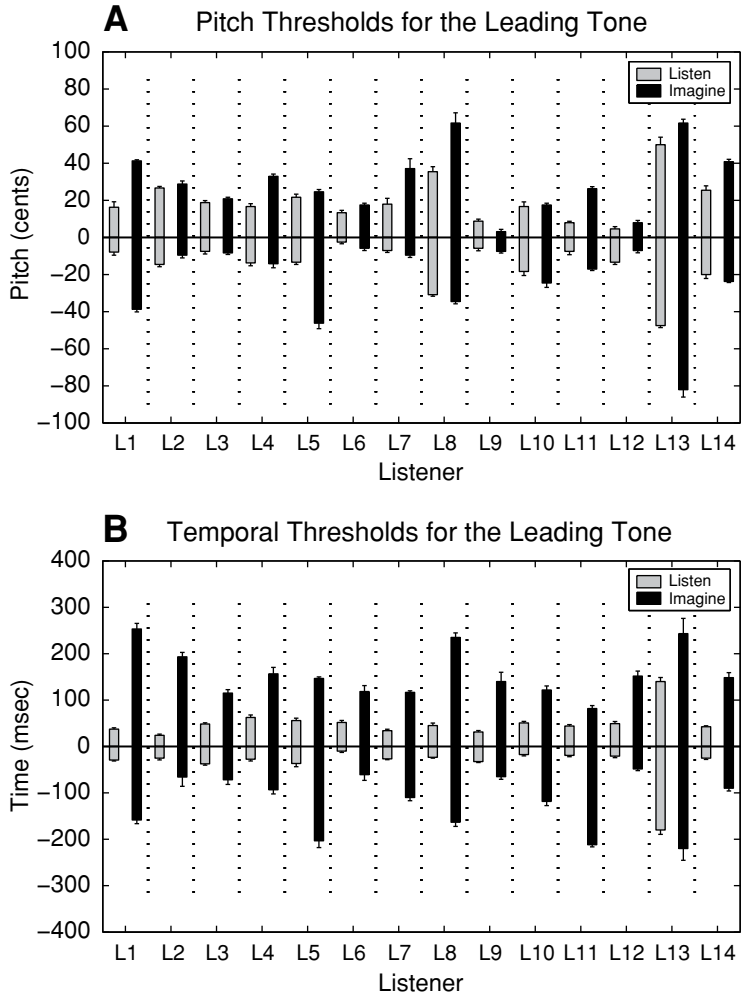

Figure 3. Tuning and timing thresholds for individual listeners judging (A) the intonation of the leading tone, the penultimate note of an ascending major scale, or (B) the timing of the leading tone. In the listen condition (gray bars), six notes were presented prior to the target note, whereas in the imagine condition (black bars), four notes were presented and three were imagined. Error bars represent 1 *SEM*, based on the six turnaround points used to calculate each threshold.

cents; $t(1,12) = 3.18$, $p = .0095$] and the imagine [4.4 $\pm$ 2.0 cents; $t(1,12) = 2.2$, $p = .0480$] conditions. As in Experiment 2, there was considerable variability in image acuity across listeners.

As in Experiment 2, temporal images were significantly broader (mean difference = 195 $\pm$ 21 msec) in the imagine than in the listen condition for all the listeners (see Figure 3B and Table 3). Although the center of the temporal images did not differ significantly between conditions [11.46 msec; $t(1,12) = 1.01$, n.s.], the expectation for a slight, 9.5-msec delay in the listen condition was significant [$t(1,12) = 5.02$, $p = .0003$], and there was a tendency, albeit more variable, toward the same in the imagine conditions [19.9 msec; $t(1,12) = 2.05$, $p = .0625$; see Table 3].

The combined analysis of pitch image width estimates from Experiments 2 and 3 identified a significant main effect of imagery condition [$F(1,25) = 13.19$, $p = .0013$], in which images were broader, on average, in the imagine condition by 9.6 cents. This increased width was attributable to the imagine condition in Experiment 3, as suggested by a significant interaction between imagery condition and terminal note [$F(1,25) = 6.87$, $p = .0147$] and the pairwise comparison between imagery conditions in Experiment 3 noted above. Neither a pairwise comparison of image widths between Experiments 2 and 3 for the imagine condition nor one for the listen condition showed a significant difference between the two experiments. Thus, the broadening of the image width when the leading tone is imagined is rather modest. Pitch images in the imagine condition were more sharp by 2.5 cents [$F(1,25) = 6.25$, $p = .0193$], but there was neither a main effect of target note nor an interaction of target note and trial type.

As was expected, the combined analysis of offsets in the temporal component showed no significant effects, and the analysis of temporal image width showed only a highly significant main effect of trial type [$F(1,25) = 208.21, p < .0001$], but neither a main effect of target note nor an interaction of target note and trial type.

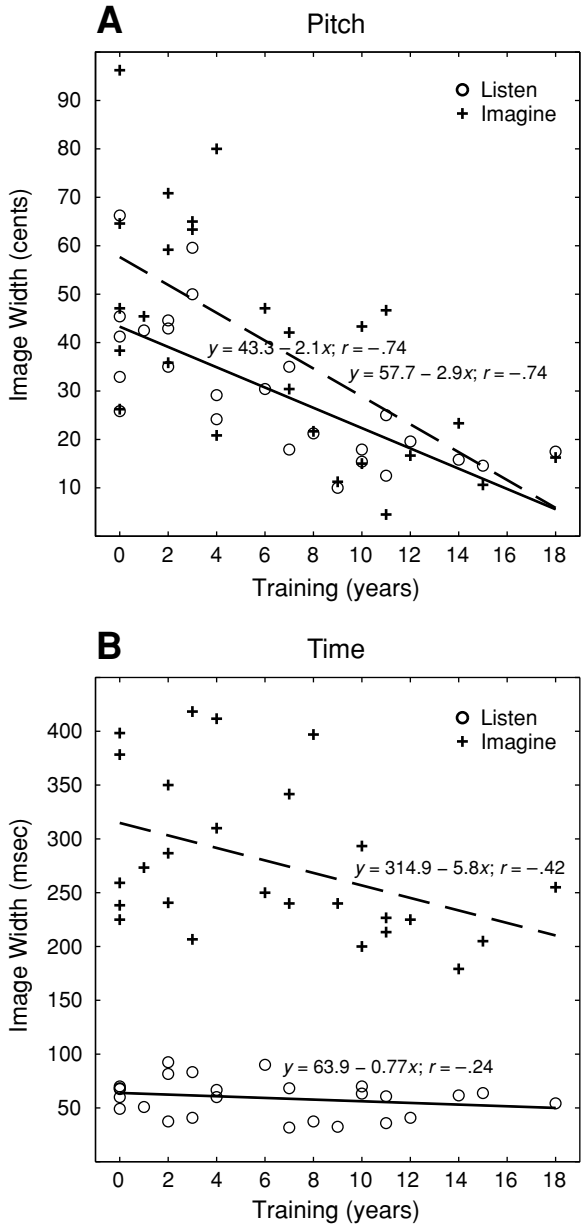

Figure 4. Pitch and temporal image acuity improve with music training. For the 22 listeners in Experiments 2 and 3 for whom the number of years of musical training was known, the duration of musical training was positively correlated with smaller thresholds in both the listen and the imagine conditions for pitch (panel A). Temporal images showed an improvement with musical training only in the imagine condition (panel B). The solid line is the slope of the regression line from the listen conditions, and the dashed line is the slope for the imagine conditions.

Finally, we calculated the correlations between the duration of musical training and each of our dependent measures of auditory image acuity. Musical training significantly influenced the acuity of auditory images, primarily along the pitch dimension (Figure 4A). Pooled across the 22 listeners in Experiments 2 and 3 for whom musical training data were available, the width of the image was negatively correlated with years of musical training in both the listen conditions [$r(20) = -.74, p < .0001$] and the imagine conditions [$r(20) = -.64, p = .0004$]. Although the width of temporal images was uniformly good when the listeners heard all of the notes preceding the target, more musically trained listeners again showed significantly tighter thresholds when asked to imagine three notes [$r(20) = -.42, p = .0320$; see Figure 4B].

Discussion

Experiment 3 replicated the results of Experiments 1 and 2, showing that the temporal component of auditory images is considerably less precise when multiple notes in a scale must be imagined than when the timing of the next event in an isochronous note sequence is to be evaluated. The small but significant broadening of images in the pitch dimension in the imagery task suggested that the listeners had a more difficult time forming an accurate image of the leading tone than of the tonic. Because their images of the leading tone were not biased upward in pitch toward the tonic, it is unlikely that the broader thresholds arose from a blending or confusion of an image for the leading tone with a primed image for the tonic. It is possible that the process of imagining a sequence of pitches is inherently associated with increased variance that results in a broader mental image for the final pitch in the sequence than with the expectancy associated with the next heard pitch. Such an interpretation would suggest that the slightly narrower width of the image for the tonic reflects a facilitation of image acuity by the primed tonal context. Such an account needs further support by showing that image acuity broadens similarly for other notes surrounding the tonic, such as a continuation of the scale to the supertonic. However, previous work shows a context-dependent facilitation of tuning judgments when the target pitch is at the end of a familiar melody and, to a lesser extent, at the end of a repeating or alternating tone sequence, in comparison with judgments made on isolated pairs of tones (Warrier & Zatorre, 2002).

EXPERIMENT 4

We performed a final experiment to test whether the comparable acuity of pitch images in the imagery conditions of Experiments 2 and 3 might be attributable solely to priming of the key to which the probes belonged. Given that the leading tone is judged to be much less strongly associated with a primed key than is the tonic (Janata & Reisberg, 1988; Krumhansl, 1990), it is unlikely that the comparable image acuity for tonic and leading tone probes in the imagery conditions in Experiments 2 and 3 is attributable to priming effects alone. However, to assess

more thoroughly the relative influence of tonal priming versus imagery instructions on the listeners' intonation judgments, we performed a within-subjects comparison of tonic and leading tone probe judgments, using slightly modified versions of the stimulus materials and task.

Listeners performed a 2AFC task in which they heard the initial four notes of an ascending major scale and discriminated among the same set of pitch probes, as in Experiments 2 and 3. However, both the stimuli and the task differed in one regard. First, the probe always occurred earlier, with the intent of rendering a strategy of isochronous image formation of the remaining notes useless. Second, because we wanted to discourage any explicit strategy of imagery, we gave the listeners a *context membership task*, in which they had to determine which of the two probe items was a note from the key (context) that was suggested/primed by the first four notes of the scale. Since the in-tune probe is technically a member of the primed context, whereas a mistuned probe is not, the task serves to estimate the degree to which the intonation judgments in Experiments 1–3 were based on priming of the key alone.

We were interested in examining two questions. First, would the image widths of the leading tone and the tonic differ from one another? The hypothesis that each probe is primed equally by the context would have to be rejected if the two differed. Second, in order to assess any relative benefit of an explicit instruction to imagine the remaining notes in the scale over no imagery instruction whatsoever, we compared the image widths obtained in Experiments 2–4. If our results in Experiments 2 and 3 were driven primarily by priming of the key, image widths should not change in this experiment. However, if the use of an imagery strategy is explicitly supported by both task instructions and the stimulus structure (as in Experiments 2 and 3), we would expect image widths to broaden when imagery instructions are absent and the assumptions about when the probes should occur in an isochronously presented scale are disrupted.

Method

Subjects. Twenty-two members of the University of California, Davis community (13 of them female) served as listeners in the experiment and received either course credit or $10/h for their participation. One of the listeners did not complete the experiment, on account of fatigue, and another was excluded on account of hearing problems. The remaining listeners ranged in age from 18 to 26 years [20.6 ± 2.0 years (mean ± standard deviation)]. Nineteen were right-handed. Fifteen listeners reported having undergone at least 1 year of formal musical training for voice or instrument (range, 2–15 years), and among them, the amount of training was 6.3 ± 4.4 years. All the listeners provided informed consent in accordance with the guidelines of the Institutional Review Board at the University of California, Davis. None of the listeners had participated in any of the previous experiments.

Stimuli. The stimuli used were identical to those used in the pitch conditions in Experiments 2 and 3, with the exception that the probes were heard 1,600 msec following the onset of the final (fourth) note in the prime. With respect to an isochronous continuation of the scale, the probe occurred both considerably earlier than would be expected and 200 msec earlier than the sixth note of the scale would have been heard. Both of these features of probe timing were intended to disrupt the imagery strategy that was encouraged in Experiments 1–3.

Procedure. The listeners completed two warm-up exercises before we obtained thresholds. First, in order to familiarize them with a 2AFC task structure, we had the listeners perform eight trials in which each interval had two pitches about which they made *same*/*different* judgments in order to indicate which interval contained two different pitches. The frequency difference varied between 35 and 50 cents, in order to make it readily detectable. Once the listeners were able to make the discriminations reliably, they performed eight practice trials of the actual task. The listeners were instructed that each trial was split into two parts and that, in each part, they would hear the first four notes of a major scale and then a probe that either did or did not belong to the same key/context that the first four notes belonged to. Their task was to indicate which part of the trial had the probe note that did not fit into the context. They pressed a button with their left or right index finger to indicate whether they thought that the out-of-context probe was in the first or the second part, respectively. Once the listener felt comfortable with the task structure, we proceeded to obtain the upper and lower thresholds for both the leading tone and the tonic simultaneously. Following the threshold estimation procedure, the listeners completed a brief questionnaire to probe their strategies and awareness of different aspects of the task.

In earlier tests of the procedure, we noticed that some trajectories of pilot subjects did not converge toward the probe notes but traveled, instead, to the boundary of one semitone that we had set for our stimuli. In such cases, the threshold estimation procedure for that particular threshold was terminated if the subject was stuck at the boundary for three trials.

Data analysis. The data were analyzed using mixed model analyses of image width and offset, as in the previous experiments.

Results

In accordance with our predictions, Figure 5 and Table 2 illustrate that the change in stimuli and task resulted in a significant decrement in image acuity for the leading tone, relative to the tonic [30 ± 6.8 cents; $t(18) = 4.33, p = .0004$]. A comparison of tonic and leading tone image widths in Experiment 4 with their counterparts in Experiments 2 and 3, respectively, indicated that image widths were significantly larger in Experiment 4 for both probes [tonic, 44 ± 11 cents; $t(44) = 4.03$, $p = .0002$; leading tone, 60 ± 11 cents; $t(57) = 5.49, p < .0001$]. In 14 of 20 listeners in Experiment 4, image widths were considerably larger for the leading tone than for the tonic. The effect of type of task on the relative image widths for the leading tone and tonic was captured in a highly significant experiment × probe type interaction [$F(4,33.9) = 78.06, p < .0001$]. An analysis of the image offsets indicated no significant differences from zero or between probe types (Table 3).

Most listeners (17/20) were aware that there were two different probe tones. When asked which probe tone (lower or higher) fit better with the preceding context, 4 of the 17 indicated that the leading tone fit better, and 13 indicated that the tonic did. This difference in ratings fits with standard results in the tonal hierarchy literature (e.g., Krumhansl, 1990). Ratings of differential difficulty of performing the task for each of the probe notes on a 7-point scale (1 = *more difficult for lower probe*, 4 = *no difference*, 7 = *more difficult for the higher probe*), showed no systematic difference between the two probes [3.9 ± 0.6; $t(16) = -.48$, n.s.], although the distribution

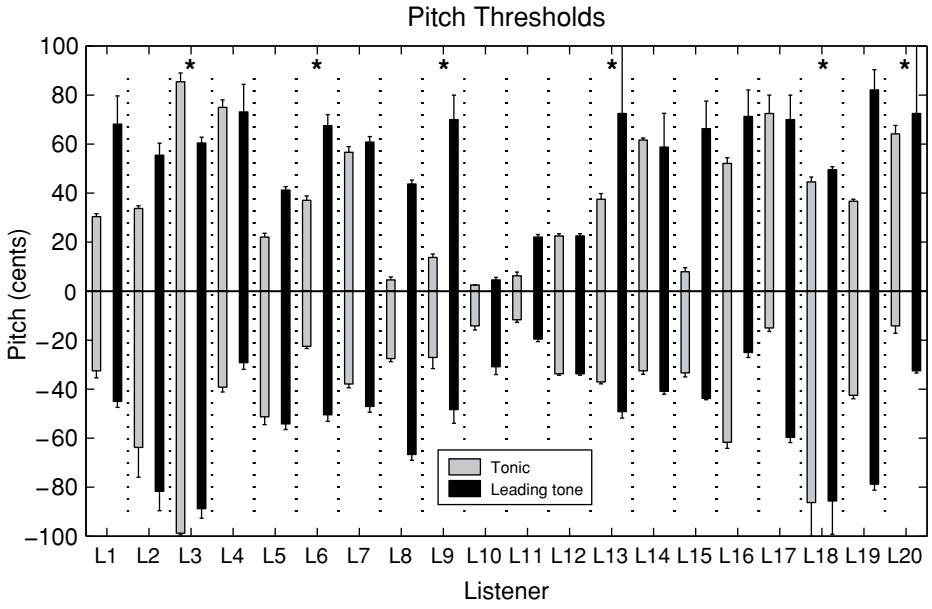

Figure 5. Tuning thresholds for individual listeners who made key membership judgments separately for the leading tone and the tonic, relative to their mistuned counterparts. In all cases, the listeners heard the initial four notes of an ascending major scale in order to prime the key. No imagery instructions were given. Gray bars are data from the condition in which the listeners made judgments about the tonic (eighth) scale degree, and black bars are data from judgments about the leading tone (seventh) scale degree. Asterisks indicate listeners who reported using imagery sometimes to never. All of the other listeners reported forming images at least sometimes.

was bimodal, with 9/17 listeners giving a rating of 1 or 2 and 6/17 listeners giving a rating of 6 or 7. When asked to rate how confusing the task was (1 = *not at all confusing*, 4 = *somewhat confusing*, 7 = *very confusing*), the listeners gave it a rating of 2.6 ± 0.4. Additional evidence that the tonic served as a stronger perceptual magnet, even when mistuned, comes from the observation that the upper thresholds for the leading tone migrated to the upper boundary of mistuning in 50% of the subjects, reaching one semitone (the tonic) in all of the cases (8/10) in which the maximal mistuning was one semitone. Thus, in the trials in which the level of upward mistuning of the leading tone had almost reached the tonic, the listeners directly compared the relative stability of the leading tone and tonic with respect to the priming context and judged the tonic to be more stable. In contrast, only 2 listeners migrated to the upper boundary when evaluating the tonic. With regard to flat mistunings, 3 listeners migrated to the lower boundary for both the leading tone and the tonic.

We also queried the listeners about their use of imagery in performing the task, asking them, "As the trials went by, did you find yourself imagining the probe note in your mind before you actually heard it? (1 = *yes, always*, 4 = *sometimes*, 7 = *no, never*)." The mean response was 2.6 ± 0.4, indicating that many listeners found themselves trying to form images of the probe notes. When asked to rate how vivid their mental images were (1 = *very vivid*, 7 = *had no image*), the mean response was 3.5 ± 0.4. When asked whether they had imagined all of the missing notes of the scale or just the single probe note, 41.2% answered that they had imagined all of the missing notes, whereas 58.5% responded that they had imagined only the probe note.

As in Experiments 2 and 3, the amount of experience with playing a musical instrument had a significant effect on image widths (Figure 6), although there was a significant interaction of probe and amount of experience [$F(1,18) = 4.82$, $p < .05$]. The interaction was due to a significant negative correlation between image width and amount of musical experience for the tonic ($R^2 = .20$, $p < .05$), but not for the leading tone ($R^2 = .01$, n.s.). A mixed model evaluating image widths across Experiments 2–4 and using experience as a covariate showed an overall significant main effect of experience [$F(1,44) = 11.69$, $p = .0014$] and an experience × probe interaction [$F(1,21) = 5.52$, $p = .0285$] but no experience × experiment interaction [$F(2,47) = 1.62$, n.s.]. The fact that the latter interaction was not significant is important insofar as it provides some reassurance that the differences in image widths between Experiments 2 and 3 and Experiment 4 are not attributable to differences between Dartmouth and University of California, Davis students. Accounting for the variance associated with experience did not change the significance of the outcome of any pairwise comparison between tonic and leading tone probes.

Discussion

A comparison of the results from Experiment 4 with those from Experiments 2 and 3 served to assess the relative influence of contextual priming versus imagery instructions on the width of tuning of mental pitch images

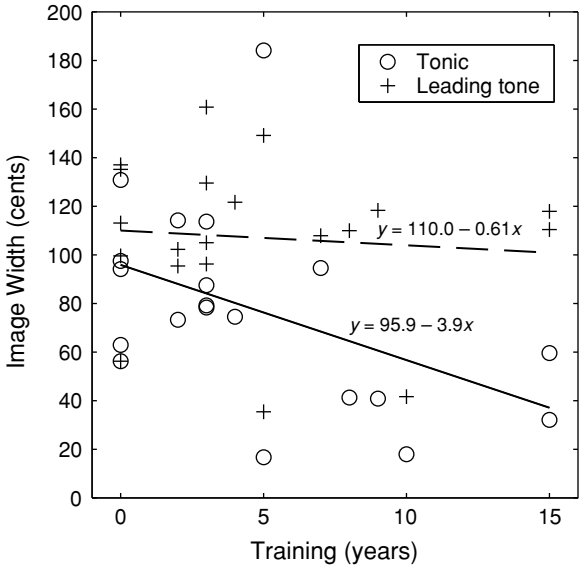

Figure 6. Correlation of musical training and tuning curves for judgments about the tonic (solid regression line) and leading tone (dashed regression line). Increased musical training results in significantly better performance when the tonic is judged.

in the context of tuning judgments about probe tones. The large increase in image widths for both the tonic and the leading tone in Experiment 4 indicated that priming of a tonal (key) context could not explain the image acuity in the earlier experiments. Because we manipulated aspects of both the task and the stimulus, it is interesting to consider the possible causes of the pattern of image acuity decrements we observed.

One possible reason is the absence of explicit imagery instructions. Interestingly, most listeners reported frequently attempting some imagery strategy, although the data in Figure 5 suggest that relatively few listeners—for example, L8, L10–12, and L15—benefited from it. What the dissociation between listener reports and the measured thresholds nicely illustrates is that the quality of imagery can be highly variable. In other words, not all attempts at mental imagery are equally effective. Even if an imagery strategy is adopted, thus effectively turning the task into a type of imagery task regardless of the experimenter instructions, certain conditions must be met in order to achieve good image quality.

One stimulus factor that may have hampered the listeners in their attempt to imagine the probe was the temporal incongruity of the probe. The probe arrived earlier than either the leading tone or the tonic would have in an isochronous continuation of the scale. Even though Experiments 1–3 showed that temporal acuity was considerably more impaired than pitch acuity under imagery conditions, suggesting that temporal precision in an image may not be a prerequisite for forming an accurate pitch image, a gross temporal violation such as that in Experiment 4 might, nonetheless, interact with the ability to form images of pitches that normally would have occurred at specific later times. Such structural interactions of tonal context and time were documented by Jones and Boltz (1989) in an experiment in which listeners extrapolated the ending time of melodies missing the final 1–5 notes. Extrapolated ending times could be induced to deviate from the expected time (1,200 msec) as a function of the preceding temporal and harmonic accent structure, indicating that the perceived accent structure could shape expectancies reaching several events into the future. In our situation, the probe violated both the isochronous context and the pitch that would normally have been expected around that time if the ascending scale had been continued normally.

Image acuity may have been impaired also by the unpredictability of which scale degree (leading tone or tonic) would be the focus of any given trial. If the listeners formed an image of a probe, it may not have been an image of the correct probe for that trial. A mismatch between imagined probe and actual probe on the first part of the 2AFC trial might be expected to introduce uncertainty into the judgment of whether or not the probe that actually sounded was part of the primed key, which, in turn, could manifest itself as a broader image.

A different result from Experiment 4 is consistent with results that would be predicted by one of two priming accounts. Our observation that mental images associated with the leading tone were significantly less focused than those associated with the tonic is in accordance with probe tone studies of tonal hierarchies that have shown the leading tone to be judged as less strongly associated with a key priming context than is the tonic (e.g., Janata & Reisberg, 1988; Krumhansl, 1990). Specifically, we find that when conditions for accurate image formation are not optimal and contextual priming effects can presumably dominate, judgments about the key membership of mistuned scale degrees are more precise for the tonic, the more stable note in the hierarchy. One must point out, however, that the results could also be explained by repetition priming, arising from the fact that the pitch class of the tonic was heard in the priming context, whereas the pitch class of the leading tone was not heard (cf. Oram & Cuddy, 1995). Because we were interested in keeping the stimuli in Experiment 4 as similar as possible to those in Experiments 2 and 3, we did not compare the acuity of two pitch classes that were not part of the priming sequence but that differed in their positions in the tonal hierarchy.

GENERAL DISCUSSION

We investigated the acuity of mental auditory images, using psychophysical threshold estimation procedures in a series of four experiments. Experiments 1–3 directly compared the acuity of *expectancies* for immediately following auditory events with the acuity of the last *image* in a sequence of auditory images retrieved from memory following an acoustic cue. In these experiments, we found that the acuity of auditory images depended on the dimension of them that the listeners were asked to judge. Whereas the listeners formed images for pitch that were as accurate when they imagined two notes preceding a target

as when they heard all of the notes preceding a target, their images in time were impaired when imagining of multiple notes preceding the target was required. By having the listeners perform both an attentional-cuing task (the listen conditions in Experiments 1–3) and an imagery task (the imagine conditions in Experiments 1–3), we were able to provide evidence that the mental representation of pitch at the point of comparison with an incoming acoustic stimulus is practically equivalent in the two tasks. Absent direct neurophysiological data, we cannot prove that the two mental representations are the same in terms of their activation patterns in the neural substrates that are required for the discrimination judgment. However, we have reduced the need to posit that pitch expectations and pitch images differ. Indeed, the idea of *images as anticipations* put forward by Neisser (1976) provides a unifying framework for our data and for thinking about the relationship of attention and imagery tasks. Experiment 4 served to illustrate that the listeners in Experiments 1–3 were not basing their tuning decisions purely on the relationship of the probe note to the sense of key that was established by the priming sequence. Rather, the results indicate that some combination of explicit knowledge about what the probe would be, its approximate timing, and instructions to imagine all of the notes leading up to the probe were necessary for increased image fidelity.

Although our tasks and pitch acuity data fit well with Neisser's (1976) *mental image as anticipation* framework, the divergence of pitch and temporal acuity across the attentional-cuing and imagery contexts presents more of a puzzle. In other words, why do we accurately anticipate the pitch but not necessarily the exact moment at which it will occur? Neuropsychological data suggest that these dimensions are not always tightly bound. Amusics—stroke patients or neurologically intact individuals with pronounced impairments on tasks requiring fine-grained pitch judgments—show better performance, sometimes even on par with control listeners, on temporal judgments, relative to pitch judgments, thereby indicating that at least in such individuals, the neural substrates supporting temporal and pitch judgments differ (Ayotte, Peretz, & Hyde, 2002; Hyde & Peretz, 2004; Peretz et al., 2002; Peretz & Kolinsky, 1993). It is, therefore, not surprising that mental images formed on each dimension can show different tuning characteristics, even if listeners are instructed to imagine objects in which these dimensions are integrated—for example, the notes of a melody.

Our data lead us to conclude that images in time are more susceptible to distortion in the absence of external stimuli than are pitch images. The broader underlying issue is the manner in which timing mechanisms in the brain interact with control processes, such as attention or working memory, that are implicated in voluntary image formation, as well as the behavioral measures used to examine the timing process. Timing processes in the brain have been examined in a variety of different contexts (Block & Zakay, 1997; Ehrle & Samson, 2005; Ivry, 1996; Jones, 1993; McAuley & Jones, 2003; Schöner, 2002; Wearden, 2003; Wright & Fitzgerald, 2004), but a detailed discussion of how each of those paradigms might map onto the tasks we used is beyond the scope of this discussion. We will focus, therefore, on three possibilities for the divergence of temporal acuity that we observed across our listen and imagine conditions.

In the first, the ability of isochronous tone sequences to automatically induce expectations for events to occur at the established interonset interval is examined. Pitch-matching judgments in a working memory task are more accurate if the target pitch occurs at an interonset interval that is established by a sequence of task-irrelevant distractor tones presented with an isochronous rhythm (Jones, Moynihan, MacKenzie, & Puente, 2002). Even if the last "pacing" note is omitted, deviations as small as 15 msec from the established period of 600 msec result in impaired pitch-matching performance, indicating that the temporal expectation persists for at least one extra period. This result supports a model in which isochronous stimulus sequences serve to orient attention in time by inducing an oscillatory process that entrains to the regularity of the note sequence preceding the target (Barnes & Jones, 2000; Large & Jones, 1999; McAuley & Jones, 2003). Our data in the listen condition are entirely consistent with this account and with the temporal precision observed by Jones and colleagues.

In order to juxtapose our imagery results against an attentional oscillator context, it is important to point to a distinction between bottom-up and top-down temporal expectations suggested by Barnes and Jones (2000). The crux of the distinction is that entrained attentional processes that give rise to temporal expectations depend on external input. Thus, these temporal expectations are the result of a passive, stimulus-driven, bottom-up process, rather than a top-down process guided by voluntary orienting of attention. Indeed, in their tasks, there was no requirement for subjects to generate either covert or overt events during the temporal expectation-inducing isochronous sequence, so there is presumably relatively little top-down influence on the expectations. Our imagery task, on the other hand, can be thought of as a top-down process, because it involves the explicit formation of images for each event that remains in the sequence. Further experiments will be needed to determine how internally driven temporal expectations fit into a dynamic attending framework.

A second explanation for the deterioration of the acuity of temporal images in our experiments takes into consideration the sensorimotor constraints we placed on the listeners. We explicitly instructed them to refrain from making any movements, including vocalizations, that would help them keep time. Although we did not videotape the listeners or obtain myographic recordings to objectively verify their compliance with our instructions, there is no evidence that any of them benefited from covert timekeeping movements they may have made. It would be of interest to know to what extent the sensorimotor feedback that would be provided by allowing listeners to tap their fingers or feet would assist in maintaining temporal acuity. Studies of synchronization and continuation tapping provide a partial answer to this question. During the

production of 580-msec intervals, the standard deviation of interresponse intervals is between 15 and 20 msec in both synchronization and continuation conditions, and the mean interresponse interval in the continuation condition is a modest 10 msec shorter than the target interval (Semjen, Schulze, & Vorberg, 2000). The relative similarity in timing variability between externally and internally guided tapping behavior, when contrasted with the large differences in time estimation observed in our experiments, suggests that the sensorimotor component of the tapping tasks is essential to maintaining the accuracy of the internal timekeeper across multiple mentally generated events when an external pacing signal is absent. This account is also consistent with the notion that attentional oscillators may require some form of external input to maintain a precisely timed attentional focus.

Finally, the relative difficulty in generating mental images in time with the restrictions we placed on our listeners might have also biased them toward using a different imagery strategy. Specifically, given that the in-tune probe notes in Experiments 1–3 were unique and, therefore, fully predictable, the listeners may have formed a memory of the prime and probe that they could use to accurately place their image in pitch without regard for the exact timing. In other words, since the listeners knew which pitch would be probed, there was less incentive to arrive at the proper pitch location incrementally by imagining each successive note than might have been the case if the pitch location to be probed had not been known ahead of time.

Similarly, when performing the time judgments in the imagery trials, the listeners may have formed a memory trace of the interval separating the last context note and the probe, instead of imagining each of the notes leading up to the probe at the right time. Thus, the increased variance of temporal estimates we observed may have been a product of the increased temporal separation from the last reference interval, as would be expected given the extensive interval timing literature. Because our experiments were not structured to verify that the listeners were imagining every note leading up to the probe note, we cannot rule out the possibility that this alternate strategy was used.

Finally, we found significant correlations between the amount of musical training and image thresholds on both the intonation and the context membership tasks. More musical training was accompanied by more narrowly tuned images, primarily along the pitch dimension. Although we did not find an effect of musical training on the ability to discriminate a deviation of the probe note from isochrony (cf. Ehrle & Samson, 2005), we did find that increased musical training was correlated with a better ability to judge whether a probe note occurred on time in the imagery conditions. With regard to pitch images, we observed a more interesting dissociation across Experiments 2–4, in that the benefit of musical training in the context membership task was apparent for the tonic, but not for the leading tone. It might be expected that musicians would have benefited from their training when making judgments about the leading tone also. One possible reason for the discrepancy between the tonic and the leading tone image widths as a function of musical training is the observation that ratings of how well a tonic probe fits with a preceding tonal context are higher among musicians than among nonmusicians, whereas ratings for diatonic tones outside the major triad are comparable for the two groups, if not lower for musicians (Halpern, Kwak, Bartlett, & Dowling, 1996). Thus, if the listeners in our study were basing their membership judgments on a sense of how well the tones fit, musicians would be expected to benefit when judging trials with the tonic.

What remains to be determined is whether musically trained listeners showed the benefits they did because a task in which they are asked to listen carefully and make subtle acoustic discriminations comes more naturally to them and they find it easier to focus their auditory attention as directed, or whether the sharper tuning curves reflect a finer tuning of the neural circuitry underlying auditory attention and image formation. Recent recordings of brain electrical activity in musically trained and untrained listeners performing a pitch discrimination judgment at threshold showed that musical training was unrelated to the strength of a brain potential index of automatic processing of acoustic deviance, the mismatch negativity, whereas musical training did increase the magnitude of event-related potentials associated with attentional processing when finer discriminations were made (Tervaniemi, Just, Koelsch, Widmann, & Schroger, 2005).

Overall, our results support a view that auditory images are multifaceted and that their acuity along any given dimension depends, in part, on the context in which they are formed, the manner in which they are probed, and the musical expertise of the listener. The data suggest that images on the dimension of pitch are isomorphic when generated by attention-cuing and imagery tasks. Electrophysiological studies showing similar responses of the auditory cortex to expected/imagined auditory events and their heard counterparts provide more direct evidence for this claim (Hughes et al., 2001; Janata, 2001) and can, perhaps, shed light on how the temporal acuity of auditory images is shaped.

REFERENCES

Ayotte, J., Peretz, I., & Hyde, K. (2002). Congenital amusia: A group study of adults afflicted with a music-specific disorder. *Brain*, **125**, 238-251.

Barnes, R., & Jones, M. R. (2000). Expectancy, attention, and time. *Cognitive Psychology*, **41**, 254-311.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective duration judgments: A meta-analytic review. *Psychonomic Bulletin & Review*, **4**, 184-197.

Crowder, R. G. (1989). Imagery for musical timbre. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 472-478.

Deutsch, D. (1970). Tones and numbers: Specificity of interference in immediate memory. *Science*, **168**, 1604-1605.

Dowling, W. J., Lung, K. M.-T., & Herrbold, S. (1987). Aiming attention in pitch and time in the perception of interleaved melodies. *Perception & Psychophysics*, **41**, 642-656.

Ehrle, N., & Samson, S. (2005). Auditory discrimination of anisochrony: Influence of the tempo and musical backgrounds of listeners. *Brain & Cognition*, **58**, 133-147.

Farah, M. J. (2000). The neural bases of mental imagery. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 965-974). Cambridge, MA: MIT Press.

Farah, M. J., & Smith, A. F. (1983). Perceptual interference and facilitation with auditory imagery. *Perception & Psychophysics*, **33**, 475-478.

Halpern, A. R. (1988). Mental scanning in auditory imagery for songs. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 434-443.

Halpern, A. R., Kwak, S., Bartlett, J. C., & Dowling, W. J. (1996). Effects of aging and musical experience on the representation of tonal hierarchies. *Psychology & Aging*, **11**, 235-246.

Halpern, A. R., & Zatorre, R. J. (1999). When that tune runs through your head: A PET investigation of auditory imagery for familiar melodies. *Cerebral Cortex*, **9**, 697-704.

Hubbard, T. L., & Stoeckig, K. (1988). Musical imagery: Generation of tones and chords. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 656-667.

Hughes, H. C., Darcey, T. M., Barkan, H. I., Williamson, P. D., Roberts, D. W., & Aslin, C. H. (2001). Responses of human auditory association cortex to the omission of an expected acoustic event. *NeuroImage*, **13**, 1073-1089.

Hyde, K. L., & Peretz, I. (2004). Brains that are out of tune but in time. *Psychological Science*, **15**, 356-360.

Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology*, **6**, 851-857.

Janata, P. (2001). Brain electrical activity evoked by mental formation of auditory expectations and images. *Brain Topography*, **13**, 169-193.

Janata, P., & Reisberg, D. (1988). Response-time measures as a means of exploring tonal hierarchies. *Music Perception*, **6**, 161-172.

Jones, M. R. (1981). Music as a stimulus for psychological motion: I. Some determinants of expectancies. *Psychomusicology*, **1**, 34-51.

Jones, M. R. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 69-112). Oxford: Oxford University Press.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review*, **96**, 459-491.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, **13**, 313-319.

Kalakoski, V. (2001). Musical imagery and working memory. In R. I. Godøy & H. Jørgensen (Eds.), *Musical imagery* (pp. 43-55). Lisse, The Netherlands: Swets & Zeitlinger.

Keller, T. A., Cowan, N., & Saults, J. S. (1995). Can auditory memory for tone pitch be rehearsed? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 635-645.

Kosslyn, S. M., & Thompson, W. L. (2000). Shared mechanisms in visual imagery and visual perception: Insights from cognitive neuroscience. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 975-985). Cambridge, MA: MIT Press.

Kraemer, D. J. M., Macrae, C. N., Green, A. E., & Kelley, W. M. (2005). Musical imagery: Sound of silence activates auditory cortex. *Nature*, **434**, 158.

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

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review*, **106**, 119-159.

Littell, R. C., Milliken, G. A., Stroup, W. W., & Wolfinger, R. D. (1996). *SAS system for mixed models*. Cary, NC: SAS Institute.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception & Performance*, **29**, 1102-1125.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman.

Oram, N., & Cuddy, L. L. (1995). Responsiveness of Western adults to pitch-distributional information in melodic sequences. *Psychological Research*, **57**, 103-118.

Pechmann, T., & Mohr, G. (1992). Interference in memory for tonal pitch: Implications for a working-memory model. *Memory & Cognition*, **20**, 314-320.

Peretz, I., Ayotte, J., Zatorre, R. J., Mehler, J., Ahad, P., Penhune, V. B., & Jutras, B. (2002). Congenital amusia: A disorder of fine-grained pitch discrimination. *Neuron*, **33**, 185-191.

Peretz, I., & Kolinsky, R. (1993). Boundaries of separability between melody and rhythm in music discrimination: A neuropsychological perspective. *Quarterly Journal of Experimental Psychology*, **46A**, 301-325.

Schöner, G. (2002). Timing, clocks, and dynamical systems. *Brain & Cognition*, **48**, 31-51.

Semjen, A., Schulze, H. H., & Vorberg, D. (2000). Timing precision in continuation and synchronization tapping. *Psychological Research*, **63**, 137-147.

Tervaniemi, M., Just, V., Koelsch, S., Widmann, A., & Schroger, E. (2005). Pitch discrimination accuracy in musicians vs nonmusicians: An event-related potential and behavioral study. *Experimental Brain Research*, **161**, 1-10.

Warrier, C. M., & Zatorre, R. J. (2002). Influence of tonal context and timbral variation on perception of pitch. *Perception & Psychophysics*, **64**, 198-207.

Wearden, J. H. (2003). Applying the scalar timing model to human time psychology: Progress and challenges. In H. Helfrich (Ed.), *Time and mind II: Information processing perspectives* (pp. 21-39). Göttingen: Hofgrefe & Huber.

Wright, B. A., & Fitzgerald, M. B. (2004). The time course of attention in a simple auditory detection task. *Perception & Psychophysics*, **66**, 508-516.

Zatorre, R. J., Halpern, A. R., Perry, D. W., Meyer, E., & Evans, A. C. (1996). Hearing in the mind's ear: A PET investigation of musical imagery and perception. *Journal of Cognitive Neuroscience*, **8**, 29-46.

(Manuscript received November 12, 2004;
revision accepted for publication September 7, 2005.)