Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 4, 656–667

Musical Imagery: Generation of Tones and Chords

Timothy L. Hubbard and Keiko Stoeckig
Dartmouth College

Four experiments are reported that examine subjects' ability to form and use images of tones and chords. In Experiments 1 and 3, subjects heard a cue tone or chord and formed an image of a tone or chord one whole step in pitch above the cue. This image was then compared to a probe tone or chord that was either the same as the image in pitch, different from the image in pitch and harmonically closely related, or different and harmonically distantly related. Image formation times and response times were measured. In Experiment 3 a random-tone mask was used to control for possible contributions of the cue in echoic memory. In both experiments tone images were formed faster than chord images, a result consistent with the idea of structural complexity as a determinant of image formation time. Response times and accuracy rates were found to parallel results found in music perception studies, results consistent with the idea of shared mechanisms in the processing of musical images and percepts. Experiments 2 and 4 were control experiments examining the possible influence of demand characteristics and subjects' knowledge. In Experiment 2 subjects predicted the results of "an average person" using imagery, and in Experiment 4 subjects completed the task without receiving imagery instructions. The findings ruled out the possibility that demand characteristics and subjects' knowledge were solely responsible for the results of Experiments 1 and 3 and support the argument for the role of imagery.

In the past decade there has been a virtual explosion of research in mental imagery (Block, 1981; Kosslyn, 1980; Shepard & Cooper, 1982), but researchers have focused almost exclusively on visual forms of imagery. Few studies have examined the functional or structural nature of auditory or other types of nonvisual imagery. Studies that have included measures of auditory imagery have usually addressed the functional aspects of auditory images, but the results have been mixed, sometimes showing facilitative (Farah & Smith, 1983) or inhibitive (Segal & Fusella, 1970) effects of auditory imagery on auditory perception.

One of the earlier studies involving the functional aspects of auditory imagery was conducted by Segal and Fusella (1970). Subjects were instructed to generate either visual or auditory images while simultaneously attempting to perceive either a visual or auditory stimulus. Cross-modal imagery did not affect perception, but same-modal imaging decreased perceptual sensitivity—more specifically, generation of an auditory image decreased ability to perceive an auditory stimulus. Conversely, Farah and Smith (1983) found that imaging a tone of a specific frequency increased ability to detect a subsequently presented tone of the same frequency by lowering the threshold (increasing *d'*) for detection. Although the focus of these experiments was an examination of the functional aspects of auditory imagery in the facilitation or inhibition of perception, the results also suggest that frequency may be represented as a structural property in auditory images.

It is difficult, however, to gain broad insight into either the functional or structural nature of auditory imagery from these studies, given the limited nature of the stimuli that were used. Because ordinary auditory stimuli are usually complex and embedded in a meaningful context, it is possible that these results are not an accurate reflection of the total effects of auditory imagery. Specifically, the relationship between perceived and imaged frequencies may bear upon the processing or judging of auditory stimuli. For example, Farah and Smith (1983) found that accuracy rates were better for detection of same-frequency targets relative to different-frequency targets. The results from melody recognition studies in music perception, however, suggest that accuracy in identifying a target depends on the harmonic relatedness of the target to the imaged tone, rather than on the absolute difference in frequency per se. Accuracy for tone targets different from, and yet harmonically distantly related to, a comparison tone in a recognition task is better than for different targets that are harmonically closely related, and in some cases may even be better than accuracy for same targets (e.g., Cuddy, Cohen, & Miller, 1979; Dewar, Cuddy, & Mewhort, 1977; Dowling, 1978); this pattern of accuracy rates also occurs when sequences of chords are used (e.g., Bharucha, 1984; Bharucha & Krumhansl, 1983).

If an image is capable of serving as a comparison stimulus in a music recognition task, then similarity judgments should be affected by the harmonic relation between an imaged stimulus and a comparison probe. It has been suggested that visual percepts and images share a common representation at some levels of processing (Finke, 1985; Shepard & Podgorny, 1978); perhaps musical percepts and images also share a common level of representation. If there are such similarities in the processing of musical images and percepts, then the

---

The authors thank Jamshed Bharucha for use of the Music Perception Laboratory at Dartmouth College and Jack Baird, Jamshed Bharucha, Robert Crowder, Margaret J. Intons-Peterson, George Wolford, and two anonymous reviewers for helpful comments on earlier versions of this manuscript.

Correspondence concerning this article should be addressed either to Timothy L. Hubbard, who is now at Whitely Psychology Laboratory, Franklin & Marshall College, P. O. Box 3003, Lancaster, Pennsylvania 17604, or Keiko Stoeckig, Department of Psychology, Glassboro State College, Glassboro, New Jersey 08028.

responses to an imaged stimulus should mirror those to a perceived stimulus. Weber and Brown (1986) suggested that common representations underlie both musical production and imaged musical production, so it is reasonable to predict that processing of imaged musical stimuli should result in patterns of responses similar to those of perceived musical stimuli.

Auditory imagery consists not only of musical imagery but also includes imagery of naturally occurring sounds in the environment. Intons-Peterson (1980) presented subjects with lists of phrases that named common environmental sounds. One group of subjects rated each name for typical loudness and another group formed images of each sound. No relationship between the loudness rating and the image formation time was found. However, the "environmental" sounds used by Intons-Peterson differed along many dimensions, not just loudness. It is quite possible that the uncontrolled values on other dimensions obscured any relationship of loudness and image formation time. Some other dimension, such as pitch or complexity, may be related to image formation time. However, it is often difficult to manipulate dimensions independently in naturally occurring sounds without distorting those sounds. The four dimensions by which musical events can be specified—loudness, duration, pitch, and timbre (Dowling & Harwood, 1986)—can generally be varied without losing the meaningfulness of the music, which make musical events (e.g., tones, chords, and melodies) ideal for the study of auditory imagery.

Although no studies have explicitly examined generation times for musical images, a few studies have examined the generation times of visual images. Kosslyn, Reiser, Farah, and Fliegel (1983, Experiment 2) presented subjects with line drawings of 12 different animals. Three versions of each animal were used. In one version, all of an animal was drawn on a single page. In the second version, the drawing was split up so that the first page included the body and the second page showed the appendages. The third version divided the body parts over five pages. Kosslyn et al. found a linear relationship between the number of pages presented and the time required to form the image. Weber and Harnish (1974) found that images of longer words required more time to form than images of shorter words. Both of these studies suggest that it takes less time to image simpler stimuli than it does to image more complex stimuli. On this basis we predict that physically more complex musical stimuli, such as chords (which consist of a group of tones sounded simultaneously), will require more time to image than physically less complex stimuli, such as single tones. To borrow from Kosslyn et al., the number of pages in which the animal stimuli were presented might be regarded as analogous to the number of tones that comprise a musical chord. Just as subjects constructed meaningful images of animals by "gluing" the disparate pages together, so might they "glue" individual tones together to produce a meaningful chord. Generation time of images of chords would then be an additive function of the number of constituent tones.

In the following set of experiments, we asked subjects to form images of musically meaningful events—complex tones and chords—and compare their images with external auditory stimuli that were meaningfully (i.e., harmonically) related to the images. To the extent that images function like actually perceived musical stimuli, the pattern of accuracy rates should parallel that found in previous music perception studies (such as Bharucha, 1984; Bharucha & Krumhansl, 1983; Cuddy et al., 1979; Dewar et al., 1977; Dowling, 1978). Reaction times have generally not been measured in melody recognition studies in music perception, but there is some evidence to suggest that increased accuracy should coincide with faster correct decisions (Bharucha & Stoeckig, 1986, 1987b).

Experiment 1

In this experiment, we collected measures of image formation time, judgment time, and judgment accuracy. Subjects heard either a cue tone or a chord on each trial and then formed an image of a tone or chord one whole step higher in pitch. For example, if the cue was the tone B$^\flat$, then subjects imaged the tone C; if the cue was the chord B$^\flat$ major, consisting of the tones B$^\flat$, D, and F, then the subjects imaged the chord C major, consisting of the tones C, E, and G. If subjects generate chords by imaging a piecemeal sequence of the constituent tones (as suggested by the results of Kosslyn et al., 1983), then generation times for chords should be greater than those for single tones. If subjects generate chords by imaging the constituent tones concurrently, then generation times for chords should not differ from those for single tones.

After the subjects formed their images, they were presented with a probe tone or chord which they judged as being the same as or different from their imaged tones or chords. Half of the probes matched the imaged pitches, half did not. Of the probe trials not matching the image, half were of tones or chords that were harmonically closely related (henceforth referred to as *related*) and half were harmonically distantly related (henceforth referred to as *unrelated*) to the image. Because prior context can influence how tones and chords are perceived (e.g., Bharucha, 1984; Bharucha & Stoeckig, 1986, 1987b; Cuddy et al., 1979; Dewar et al., 1977; Dowling, 1978) and the prior context consisted of both the cue and the image, it was necessary to ensure that differences between the different-related and different-unrelated conditions were due to the harmonic properties of the image and not to those of the probe. Therefore, we chose different-related probes and different-unrelated probes that were equidistant (on the circle of fifths) from the cue.[1] If, as suggested by Neisser (1976),

---

[1] It is possible that perceived relatedness may depend not only on the distance along the circle of fifths, but also on the direction traversed (i.e., a given distance along the circle in one direction may not yield the same degree of perceptual relatedness as an equal distance in the opposite direction). However, there is evidence to suggest that, aside from those points immediately adjacent to a given starting point, traveling equal distances along either side of the circle does yield approximately equal degrees of perceived relatedness. In a harmonic priming paradigm, Bharucha and Stoeckig (1987a) had subjects make speeded judgments about chord pairs that were closely related (I paired with V or IV, both immediately adjacent to I along the circle of fifths; e.g., C with G or F), neutrally related (I paired with chord based on the submediant or the minor third above, three steps on either side of the circle; e.g., C with A or E$^\flat$), or distantly

images function through the existing perceptual schemata—in this case, the existing schemata of harmony (e.g., see Bharucha, 1987; Bharucha & Stoeckig, 1986; Deutsch & Feroe, 1981; Krumhansl, 1979; Shepard, 1982)—then judgments to same targets should be the fastest and most accurate, followed by judgments to different-unrelated targets, with judgments to different-related targets the slowest and least accurate.

Method

Subjects. The subjects in all experiments were introductory psychology students at Dartmouth College who participated as subjects for partial course credit. No subject participated in more than one experiment, and all subjects were naive to the hypotheses until after all data had been collected. Twenty-eight undergraduates participated in the current experiment. Data from 7 subjects were discarded due to subjects' failure to follow instructions, resulting in a total of 21 subjects. Musical training, as measured by number of years of formal instrument or voice lessons, ranged from no training to 21 years of training, with a mean of 6.4 years and a median of 6.0 years. All subjects reported normal hearing; no subjects reported having absolute pitch.

Apparatus. Stimuli were synthesized by an Apple Macintosh microcomputer and presented to subjects through a Sansui A-707 amplifier and Sennheisser HD-410 headphones.

Stimuli. Cues and probes consisted of the 12 pitches of the chromatic scale and the 12 major chords based on those pitches. Frequencies were based on an equal tempered tuning fixed at an A of 440 Hz. Tones were synthesized by imposing an envelope over a five octave range (starting at 65.41 Hz) and tapering off to the loudness threshold at each end (see Shepard, 1964). The three component tones of each chord were sampled from each of these five octaves and added in proportion to their envelope amplitudes (see Krumhansl, Bharucha, & Kessler, 1982). Tones and chords constructed in this way are as close as possible to true tone or chord chroma, without the influence of pitch height. The waveform of each of the 15 component tones contained the first four harmonics with equal amplitudes, creating an organ-like timbre.

Each of the 12 tones and 12 chords occurred equally often as a cue in cue–probe pairs. Cue–probe pairs consisted only of either tone–tone or chord–chord pairs. For both tone and chord trials, cues were paired with probes such that the probe and the tone or chord one whole step (two semitones) above the cue were the same in pitch, different in pitch and harmonically related, or different in pitch and harmonically unrelated. The cue was harmonically equidistant to the different-related and different-unrelated probes. Same trials were presented twice so that the probability of a same or different response on any given trial was .5. Same-1 and same-2 trials were identical and were randomly dispersed among the other trial types and each other. This resulted in four types of trials: same-1, same-2, different-related, and different-unrelated. There were a total of 96 trials, which each subject received in a different random order.

The cue–probe pairs were specified by the distance between the cue and probe in the tone trials, or between the roots of the chords in the chord trials. In both tone and chord trials, the same probe was two semitones above the cue, the different-related probe was four semitones above, and the different-unrelated probe was eight semitones above. For example, if the cue was B♭, then the image was C, the different-related probe was D, and the different-unrelated probe was F♮ (see Table 1).

Procedure. Upon arrival subjects received a brief definition and example of a tone, a chord, and a musical interval of one whole step (2 semitones). A whole step was defined as "the distance between Do and Re in the Do-Re-Mi scale." Imaging tones and chords was also described as " 'hearing' the tone or chord 'in your head' in the same way that you would if you were to remember the sound of your mother's voice, a siren, or the sound of a car door slamming." Subjects were then given an opportunity to practice forming images of tones and chords in a training session. The training session consisted of both tone trials and chord trials, identical in structure except that in tone trials the cue and image were tones, whereas in chord trials the cue and image were chords. Subjects initiated each trial by pressing a key on the keyboard. Their keypress resulted in a 2-s presentation of the cue. After the cue had ended the subject was instructed to form an auditory image that was one whole step (two semitones) above the cue for tone trials. In chord trials, subjects were to image a chord in which the constituent tones of the chord were raised one whole step above the equivalent tones in the cue. For example, in a tone trial, if the cue was B♭, then the image was C; in a chord trial, if the cue was B♭ major, then the image was C major (see Table 1). When the image was formed, the subject pressed the space bar and a comparison tone or chord played for 2 s. This comparison was one whole step above the cue and allowed subjects to assess the accuracy of their images. Tone trials and chord trials were intermixed in a random fashion. There were 24 training trials, with additional training given if requested.

After the training session, subjects went immediately to the experimental session. Again there were two types of trials, tone trials and chord trials, both having the same structure. Subjects initiated each trial by pressing the "next trial" key on the keyboard. A 2-s cue was played. When the cue stopped playing, the subject was instructed to form an image one whole step (2 semitones) above the cue and to press the space bar when the correct image was formed. This was immediately followed by a 2-s probe, which either matched the subject's image in pitch(es) or was of different pitch(es). The subject's task was to compare the pitch(es) of the probe with those of the image, to determine whether they were the same or different, and to respond by pressing the appropriately marked key. Subjects were instructed to respond as quickly as possible while being as accurate

Table 1
Examples of Cues, Images, and Targets

Cue	Image	Target Related	Unrelated
	Experiments 1 and 3		
Tone: B♭	C	D	F♮
Chord: B♭ major	C major	D major	F♮ major
	Experiment 4		
Tone: B♭	none	D	F♮
Chord: B♭ major	none	D major	F♮ major

---

related (I paired with chord based on the tritone, maximally distant, i.e., opposite side of the circle; e.g., C with F♮). As expected, closely related chords were primed more than neutrally related chords, and neutrally related chords were primed more than distantly related chords. However, although there was a significant difference in priming between the two closely related chord pairs, no such difference was found between the two neutrally related chord pairs. This suggests that perceived relatedness may be equivalent for points equidistant from a given chord, given that the distance is greater than two steps away.

as possible. Subjects were told after each response whether it was correct or incorrect (i.e., whether, in fact, their image should have matched the probe). Summary feedback of mean response time and accuracy was given after every 24 trials.

Subjects were given 16 practice trials at the beginning of the session, with more practice if requested. Practice trials were randomly drawn from the experimental trial set. After the experimental session, each subject was debriefed and asked to complete a musical background questionnaire. The entire session took about 45 min.

Results

Time between the offset of the cue and the subject's space bar press was used as an approximation of the time to generate the auditory image (image formation time, or IFT). Data from 5 subjects with fewer than 90% usable IFTs were omitted from all subsequent analyses due to failure to follow instructions. Data from 2 additional subjects whose IFTs were greater than 2.5 standard deviations from the mean were omitted from all subsequent analyses to ensure that all subjects were following the same instructions. Image formation times from the remaining 21 subjects less than or equal to zero were discarded (2.93% of the data). As predicted, mean IFTs for chords (517 ms) were longer than IFTs for tones (452 ms), $F(1, 20) = 7.23, p = .014$.

Preliminary analysis revealed no difference in reaction times (RT), $F(1, 20) < 1, p > .34$, or accuracy rates, $F(1, 20) < 1, p > .90$, between same-1 and same-2 conditions; therefore, the two same conditions were combined into one overall same condition for all subsequent analyses. Response times greater than 2,000 ms or less than zero were discarded (0.60% of the data). Mean RTs for correct responses were analyzed using a multivariate analysis of variance (MANOVA) for repeated measures (see O'Brien & Kaiser, 1985), with two within-subjects factors: cue (tone vs. chord) and target (same, different-related, and different-unrelated).

As shown in Figure 1, there was a significant main effect of type of target, $F(2, 19) = 6.18, p = .009$, for the response times. A univariate orthogonal contrast showed that same responses were significantly faster than different responses, $F(1, 20) = 12.42, p = .002$. The cue main effect and the Cue × Target interaction were nonsignificant, both $F\text{s} < 1$, both $p\text{s} > .50$.

The accuracy rates for all conditions are also shown in Figure 1. Accuracy was greater for tones (94%) than for chords (90%), $F(1, 20) = 5.94, p = .024$. There was a main effect of target type, $F(2, 19) = 15.65, p < .001$, in which subjects were more accurate for same targets than for different targets, $F(1, 20) = 26.25, p < .001$, and accuracy was greater for different-unrelated targets than for different-related targets, $F(1, 20) = 32.37, p < .001$. There was also a significant Cue × Target interaction effect, $F(2, 19) = 4.11, p = .033$. Separate analyses for tones and chords showed a target effect for tones, $F(2, 19) = 6.16, p = .009$, in which accuracy for same targets was marginally better than that for different targets, $F(1, 20) = 2.78, p = .111$, and accuracy for different-unrelated targets was greater than that for different-related targets, $F(1, 20) = 10.17, p = .005$. There was also a target effect for chords, $F(2, 19) = 14.20, p < .001$. Accuracy was greater for same targets than for different targets, $F(1, 20) = 24.43, p < .001$, and was also greater for different-unrelated relative to different-related targets, $F(1, 20) = 28.91, p < .001$.

To determine the effects of musical training on IFTs, RTs, and accuracy rates, we reanalyzed the data with years of musical training as an additional factor. We recoded this factor into a dichotomous variable, using a median split to

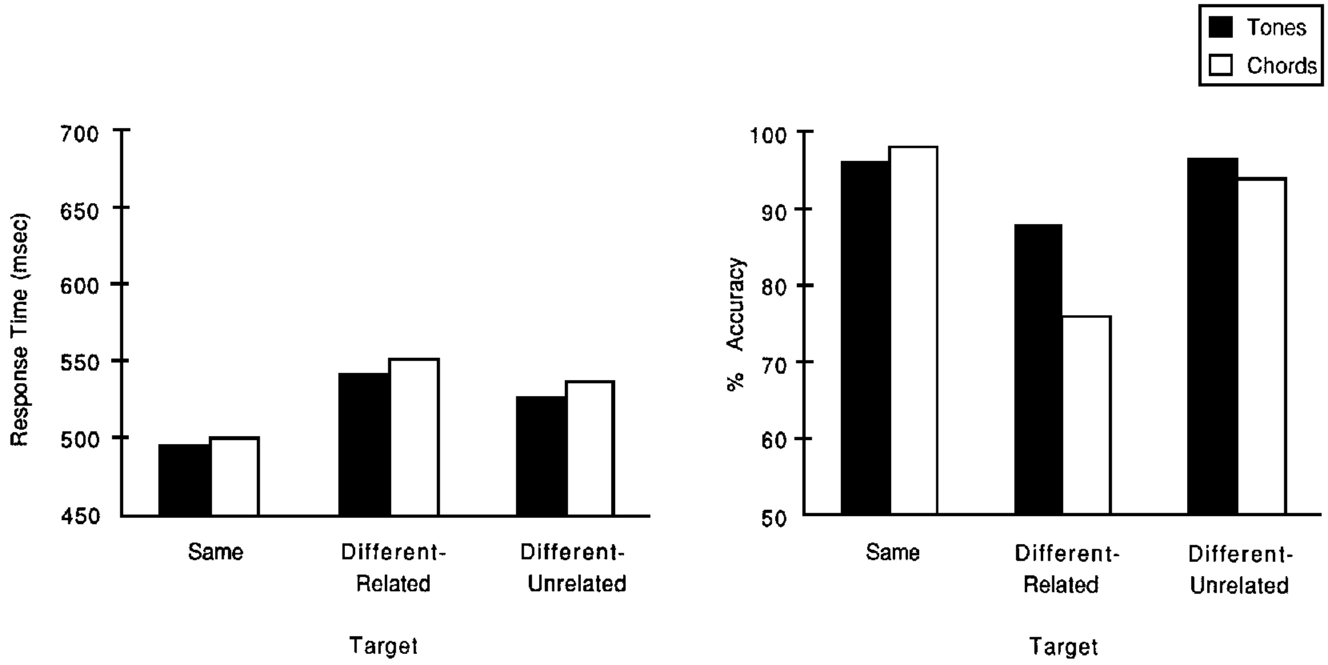

Figure 1. Response times and accuracy rates for same/different judgments in Experiment 1.

classify subjects as either highly trained or untrained. Although there were trends for highly trained subjects to have shorter IFTs, quicker RTs, and greater accuracy rates than untrained subjects, we found no significant differences, all *F*s < 1.25, all *p*s $> .25$.

Discussion

As predicted, image formation time for chords was longer than image formation time for tones. Although this difference may be attributed to the more complex physical structure of the chords, the magnitude of the difference is not as large as would be expected if subjects formed their chord images by serially imaging each constituent tone. Chord images may have been formed by imaging each of the constituent tones in parallel, with perhaps an additional amount of time needed to coalesce the tones into a single, well-formed chord. Alternatively, the chord images could have resulted from the activation of previously formed, stored complex representations of chords. It may take longer to activate chord representations than tone representations because of the potentially more complex structure of chord representations. The current data cannot distinguish between these two alternatives.

The pattern of accuracy rates and response times shows the predicted influence of harmonic relatedness in the decision task. Accuracy rates for same targets were higher than for different targets, and accuracy rates for different-unrelated targets were higher than for different-related targets. Although this pattern occurred for both tones and chords, it was particularly pronounced for chords, perhaps reflecting the greater role of chords in instantiating a key in Western music. Response time was also found to be faster for correct same responses than for correct different responses. The differential target effects found in decision times and accuracy rates for both tones and chords are in accord with the pattern of accuracies found in music recognition studies and the predictions of response times based on those accuracy rates. This pattern of results suggests that subjects used their images in making their decisions and that these images formed a musical context similar to that formed by musical percepts.

It is possible, however, that these results were not due to subjects' use of imagery. Although all subjects reported using imagery, there may be alternative explanations. Imagery experiments are particularly vulnerable to demand characteristics; that is, subjects may deduce the aims of the experiment and adjust their responses accordingly (see Mitchell & Richman, 1980). Similarly, Pylyshyn (1981, 1984) has suggested that results previously attributed to properties of imagery are actually due to subjects' beliefs, desires, and/or general knowledge. Pylyshyn refers to such beliefs and knowledge as *tacit knowledge*.[2] According to this view, the findings of Experiment 1 would result from subjects possessing conscious knowledge of harmonic structure, identifying the harmonic relationship between cue and probe on every trial, predicting the particular pattern of results that should result from such harmonic relationships, and consciously modifying their responses on every trial to produce a pattern consistent with this pattern.

The alternative explanations of demand characteristics and subjects' use of tacit knowledge are consistent with the results of Experiment 1, but the claim that subjects utilized imagery may be supported in several ways. First, a difference is obtained in the time required to form images. If subjects are not constructing images, then there is no a priori reason for the longer IFT for chords. It is not clear why propositional or other accounts would predict the longer IFTs. Second, the pattern of results for the accuracy rates parallels patterns that have been found in music recognition tasks. Because imagery is often proposed to be equivalent to perception at some levels of processing (Farah, 1985; Finke, 1985; Shepard & Podgorny, 1978), similar response patterns for both imagery and perceptual stimuli would be expected. Third, it seems highly unlikely that subjects could identify the harmonic relationship, determine what their accuracy and response latency should be, and adjust their response accordingly, all within an average span of under a second.

Nevertheless, in order to rule out the alternative explanations of demand characteristics and the use of tacit knowledge, we decided to determine if a new group of subjects would be able to predict the performance of subjects using imagery. If they could make accurate predictions, this would not necessarily invalidate an imagery explanation (cf. Denis & Carfantan, 1985), but if they could not make accurate predictions, it would be substantial evidence that IFTs, RTs, and accuracy rates obtained in Experiment 1 were due to structural properties of the image and not to demand characteristics or subjects' use of tacit knowledge.

Experiment 2

In the second experiment we asked naive subjects to predict, without using imagery, how long it would take an average person to generate images of tones and chords. They then guessed the order, from fastest to slowest, in which the average

---

[2] In the present case, knowledge of harmonic structure may be considered tacit, but not in the sense in which Pylyshyn refers. Pylyshyn uses the term *tacit knowledge* to refer to the beliefs, goals, and desires of a subject that may influence apparent properties of imagery. If changing a subject's beliefs, goals, and desires changes a subject's performance on a task, then that task is said to be *cognitively penetrable*. Pylyshyn claims that tasks which are cognitively penetrable must be explained by reference to symbolic nonimaginal processes rather than properties or mechanisms of imagery per se. A nonmusician's knowledge of musical structure, much like knowledge of grammatical structure, may be considered tacit only in that, because it is learned from exposure to a culture's music and not through explicit instruction, it is verbally inaccessible. Relatively long-term exposure to music may be necessary for developing such a knowledge of harmonic structure, but once these representational structures are developed they may constrain the properties exhibited by imagery (especially if imagery utilizes perceptual schemata—see Neisser, 1976). A change in a subject's immediate beliefs, goals, and desires would not be expected to change the representational structures that have been built up from long exposure; hence, the representation of harmonic structure would be relatively impenetrable by Pylyshyn's standards. Therefore, even though knowledge of harmonic structure is often "tacitly" acquired, this is not the type of tacit knowledge referred to by Pylyshyn.

person could make a same/different judgment about an imaged tone or chord. There is no a priori reason why naive subjects should think image generation time for chords should be longer than that for tones, so we predicted that these control subjects should exhibit no difference in their guessed image formation times for tones and chords. Likewise, we would not expect their decision time rankings to precisely match the actual ordering of outcomes in Experiment 1. If control subjects cannot accurately predict what subjects using imagery actually do, it is difficult to see how alternative explanations like subjects' use of tacit knowledge or demand characteristics can account for the data in Experiment 1. Of course, such an outcome would not definitively prove that imagery subjects did, in fact, use imagery. If other strategies are ruled out, though, it would offer convergent validity supporting the claim that subjects were indeed using imagery.

Method

Subjects. Eighteen undergraduates from the same subject pool used in the previous experiment participated as subjects. Data from 4 subjects were discarded due to subjects' failure to follow instructions, resulting in a total of 14 subjects. Musical training, as measured by number of years of formal instrument or voice lessons, ranged from no training to 21 years of training, with a mean of 7.1 years and a median of 6.5 years. All subjects reported normal hearing; no subjects reported having absolute pitch.

Apparatus. The apparatus was the same as that used in Experiment 1.

Stimuli. Tones and chords were synthesized in the same manner as in Experiment 1.

Procedure. The procedure was the same as in Experiment 1 with the following exceptions. Subjects were informed that the experiment involved two groups of subjects doing two different tasks. According to the instructions, in each trial subjects in both groups would be presented with a cue, which would be either a tone or a chord. One group of subjects would be asked to actually form an image one whole step above each cue. The other group of subjects would be asked to guess how long it would take the average person to form an image one whole step above each cue.

All subjects in this experiment, however, were actually told that they were in the "guessing group." We emphasized that they were not to actually form the images, but only to make a "best guess" as to how long it might take an average person to form the image. After the cue finished playing, subjects waited for the amount of time they thought necessary to form the required image and then pressed the space bar. After completing the guessing trials, subjects read a scenario in which another person imaged either a tone or a chord. They then rank-ordered the predicted speed of the other person's same/different judgments to other tones or chords that were related to the hypothetical image in one of three ways: (a) same pitch(es) as image, (b) different pitch(es) but harmonically related to image, and (c) different pitch(es) but harmonically unrelated to image. The order in which the choices were listed on the answer sheet was counterbalanced across subjects. The entire session took about 30 min.

Results

Data from 2 subjects with fewer than 90% usable IFTs were omitted from analysis due to failure to follow instructions. Data from 2 additional subjects whose IFTs were greater than 2.5 standard deviations from the mean were also omitted from analysis, to comply with the IFT criteria used in Experiment 1. Guessed image formation times from the remaining 14 subjects less than or equal to zero were discarded (0.67% of the data). Mean tone and chord "guessed" IFTs were compared. There was no difference between guessed IFTs for tones (1,843 ms) and chords (1,813 ms), $F(1, 13) = 0.09$, $p = .77$. A comparison using all 18 subjects' data also revealed no difference between guessed tone and chord IFTs (1,995 ms and 1,988 ms, respectively), $F(1, 17) = 0.0$, $p = .96$.

All 18 subjects' rank order predictions were analyzed using separate Friedman two-way analyses of variance (Siegel, 1956) for tones and chords. As can be seen in Table 2, subjects correctly predicted that same responses would be faster than different responses for tones, $\chi^2(2, N = 18) = 8.44$, $p = .01$, but they incorrectly ranked responses to different-unrelated tones as the slowest. Subjects did not differentiate among their rank orders of chord responses, $\chi^2(2, N = 18) = 3.58$, $p = .17$.

As in Experiment 1, we reanalyzed guessed IFTs with degree of musical training as an additional factor, using a median split to classify subjects as either highly trained or untrained. There was a trend for highly trained subjects to have shorter guessed image formation times, although once again, this difference was not significant ($F < 1$, $p > .84$).

Discussion

The alternative explanations of demand characteristics and subjects' use of tacit knowledge for the results obtained in Experiment 1 were not supported, suggesting that subjects in that experiment were, in fact, using imagery. The latencies of the guessed IFTs from Experiment 2 are over three times longer than the IFTs from Experiment 1. The reason for this difference is not clear, but it demonstrates a fundamental difference between *predicting results due to imagery* and *results due to imagery*. The demand characteristics explanation would suggest the results of Experiment 1 were due to nonimagery prediction, but the magnitude of the difference between guessed IFTs and real IFTs offers strong support for the notion that imagery is a process distinct from nonimagery prediction. Similarly, the lack of difference between guessed IFTs for tones and chords in Experiment 2 is contrary to what should result if subjects were either using tacit knowledge or tapping demand characteristics. The evidence thus suggests that subjects who reported using images were indeed using images and not merely propositional or other nonimaginal knowledge to predict the outcomes.

It is possible, however, that the echoic trace of the cue

Table 2
Mean Rank Orderings for Tones and Chords

Hypothetical images	Predicted order of response to hypothetical percept		
	Same	Different-related	Different-unrelated
Tones	1.50	2.19	2.31
Chords	1.59	2.19	2.22

Note. Rank ordering of 1 assigned to fastest response; rank ordering of 3 assigned to slowest response.

contributed to the pattern of results found in Experiment 1. These results seem to suggest that the prior context of the image, and not just the physically presented cue, is responsible for the instantiation of harmonic tonality. Even though the harmonic relatedness between the cue and probes should not have been a factor, we must still question the degree to which the cue is important in this instantiation: Is it possible that the differential effects are arising primarily from the presence of the cue, and only secondarily from the presence of the image? To answer this question, we conducted a further experiment in which a mask consisting of a randomly selected sequence of pitches was played prior to the presentation of the probe in order to erase any echoic traces of the cue.

Experiment 3

In this experiment we attempted to rule out any effect of echoic or short-term memory that may have contributed to subjects' performances. It is possible that subjects relied on a memory trace of the cue rather than constructing an image. To minimize the contributions of an echoic store to the decision task, we replicated Experiment 1 but modified it by presenting a pattern mask of random pitches between the image formation and the presentation of the probe. The interpolation of intervening tones between target and comparison tones has previously been shown to be quite successful in distorting memory for tones (Deutsch, 1982).

Despite the interference with the memory for the cue, it was important that subjects be able to generate an image of the correct pitch. If the random pitch sequence was truly successful at distorting memory for the cue, subjects might not image the correct pitch(es). Therefore, we presented the random pitch mask after the image was formed, rather than immediately after the cue occurred. Although it was possible that the sequence would mask both the cue and the image, we hoped that over the extended period of time the echoic trace of the cue would decay, whereas the image could be reconstructed or would otherwise still be intact (or nearly so) during the same/different decision. The presence of the mask, however, should decrease accuracy rates considerably.

Method

Subjects. Thirty-three undergraduates from the same pool as previous experiments participated as subjects. The data from 12 subjects were discarded due to subjects' failure to follow instructions, leaving a total of 21 subjects. Musical training, as measured by number of years of formal instrument or voice lessons, ranged from 0 to 15 years of training, with a mean of 5.5 and a median of 4.0 years. All subjects reported normal hearing; no subjects reported having absolute pitch.

Apparatus. The apparatus was the same as that used in Experiment 1.

Stimuli. Cue–probe pairs were constructed in the same manner as in Experiment 1. The mask consisted of a sequence of 16 tones, synthesized in the same manner as the tone stimuli in Experiment 1. Each of the 16 tones, lasting 125 ms, was randomly selected in each sequence, and a new sequence was randomly generated for each trial. The entire mask sequence lasted 2 s, with no pause between notes.

Procedure. The procedure was identical to that used in Experiment 1, except that a 2-s mask was interposed between the formation of the image and the presentation of the probe. A 1-s pause preceded and followed the mask. The entire session took about 45 min.

Results

Data from 10 subjects with fewer than 90% usable IFTs were omitted from all subsequent analyses due to failure to follow instructions. Data from 2 additional subjects whose IFTs were greater than 2.5 standard deviations from the mean were also omitted from all subsequent analyses, again to ensure that all subjects were following the same instructions. Image formation times from the remaining 21 subjects that were less than or equal to zero were discarded (1.49% of the data). Mean tone and chord IFTs were compared. Chord IFTs (1,305 ms) were longer than tone IFTs (1,131 ms), $F(1, 20) = 17.48$, $p < .001$.

Preliminary analysis revealed no difference in RTs, $F(1, 20) = 2.0$, $p > .17$, or accuracy rates, $F(1, 20) < 1$, $p > .54$, between same-1 and same-2 conditions; therefore, the two same conditions were combined into one overall same condition for all subsequent analyses. Response times greater than 2,000 ms or less than zero were discarded (6.70% of the data). Mean RTs for correct responses were analyzed using a MANOVA for repeated measures, with two within-subjects factors: cue (tone vs. chord) and target (same, different-related, and different-unrelated). As shown in Figure 2, the effect of cue was significant, with RTs for tones faster than RTs for chords (830 ms and 897 ms, respectively), $F(1, 20) = 37.85$, $p < .001$. Although a target effect was not found, $F < 1$, $p > .53$, the Cue × Target interaction was significant, $F(2, 19) = 9.09$, $p = .002$. Separate analyses for tones and chords showed an effect of target type for tones, $F(2, 19) = 3.53$, $p = .05$. Orthogonal planned contrasts showed only marginally significant results, with responses to same targets slower than those to different targets, $F(1, 20) = 3.32$, $p = .084$, and responses to different-unrelated targets faster than those to different-related targets, $F(1, 20) = 3.42$, $p = .079$. There was also a target effect for chords, $F(2, 19) = 5.76$, $p = .011$, with responses to same targets faster than responses to different targets, $F(1, 20) = 11.57$, $p = .003$.

Accuracy rates for each condition are also shown in Figure 2. Overall accuracy was better for tones than for chords (87% and 78%, respectively), $F(1, 20) = 11.84$, $p = .003$. Accuracy also depended on target type, $F(2, 19) = 13.71$, $p < .001$. Univariate orthogonal contrasts showed that accuracy was better for same targets than for different targets, $F(1, 20) = 15.30$, $p = .001$, replicating the findings of Experiment 1. Accuracy was also better for different-unrelated targets than for different-related targets, $F(1, 20) = 19.87$, $p < .001$, also replicating the results of Experiment 1. There was a significant Cue × Target interaction, $F(2, 19) = 10.21$, $p = .001$. Separate analyses on tones and chords revealed a target effect for tones, $F(2, 19) = 11.42$, $p = .001$, in which accuracy for different targets was marginally better than that for same targets, $F(1, 20) = 2.97$, $p = .100$, and accuracy for different-unrelated tone targets was greater than that for different-related tone targets, $F(1, 20) = 17.84$, $p < .001$. There was also a target effect for chords, $F(2, 19) = 18.13$, $p < .001$. Accuracy for

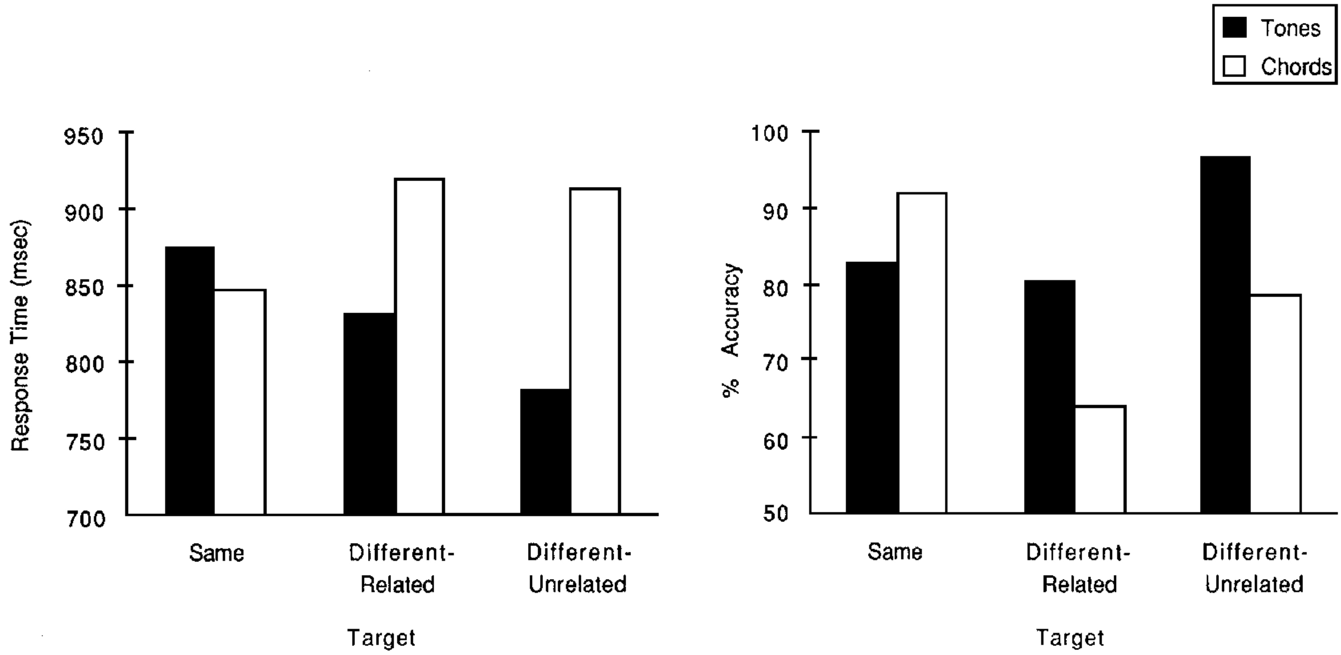

Figure 2. Response times and accuracy rates for same/different judgments in Experiment 3.

same responses was greater than that for different responses, $F(1, 20) = 32.20$, $p < .001$. Accuracy was also better for different-unrelated relative to different-related chord targets, $F(1, 20) = 11.69$, $p = .003$.

As in Experiment 1, we reanalyzed IFTs, RTs, and accuracy rates with degree of musical training as an additional factor, using a median split to classify subjects as either highly trained or untrained. Replicating the results of Experiment 1, we found trends for highly trained subjects to have shorter IFTs, quicker RTs, and greater accuracy rates, but, again, no significant differences were found, all Fs < 1, all ps $> .35$.

Discussion

In general, the pattern of results mirrors that found in Experiment 1. Although the IFTs in Experiment 3 are larger than the IFTs obtained in Experiment 1 by almost a factor of three, the ratio of chord to tone IFTs is constant across both experiments (1.14 in Experiment 1 and 1.15 in Experiment 3). That this should occur in Experiments 1 and 3, where subjects were asked to image, but not in Experiment 2, where subjects were explicitly asked not to image, supports the claim that subjects were indeed imaging when so instructed. In general, both IFTs and RTs are larger than in Experiment 1. This could be due to the use of different subjects in Experiment 3 but more probably reflects the more difficult nature of the current task. A comparison of Figures 1 and 2 shows that the mask appeared to increase the response time and slightly decrease the accuracy rate. Although the echoic trace may still have aided subjects in the construction of their images, the random tone mask prohibited its having any similar facilitation on image maintenance.

The results of the previous experiments have shown that the generation of musical images can influence the subsequent perception of tones or chords. Moreover, these differences appear to be due to properties of the image and not to demand characteristics or subjects' tacit knowledge, because subjects claiming not to image perform in a qualitatively and quantitatively different way than subjects who claim to use imagery. However, it is possible that these patterns of results may not be due to imagery per se but to properties of the mental representation of pitch apart from imagery. It is very difficult to determine whether a person has formed an image or deployed another form of mental representation, especially because it is difficult to conceive of the activation of a pitch representation that does not correspond, in some manner, to an image of that pitch. In spite of these difficulties, we attempted to address this question by seeing if we could obtain the same pattern of results with subjects who performed the same task but who were not explicitly instructed to form images. If the image is eliminated and a different pattern of responses occurs, then the previously observed patterns must clearly be dependent upon properties of tone and chord images.

Experiment 4

We conducted a fourth experiment to determine whether the results obtained in the previous experiments could be attributable to the activation of some form of pitch/chord representation other than imagery and whether subjects could make accurate decisions about pitches without using imagery. Subjects were presented with cues and subsequently asked to distinguish between targets that were either one whole step above the cue or that differed by more than one whole step from the cue. Judging targets to be one whole step above the cue corresponded to making a "same" response in Experiments 1 and 3, whereas judging targets to be other than one whole step above the cue corresponded to making a "different" response. Subjects were not explicitly instructed to use any one strategy in performing this task, and any reference to imagery was explicitly avoided in the instructions.

If the pattern of results from subjects who are not specifically instructed to image, and who later claim not to have done so, does not match the pattern of results found in Experiments 1 and 3, then the previous pattern of findings probably resulted from properties inherent in those subjects' images. If the observed patterns match those obtained earlier, and subjects later claim not to have used imagery in their decision making, then it is possible that the previous results could have been due to properties of pitch representation other than those associated with imagery. However, if the observed patterns match those obtained previously, but subjects claim to have adopted an imagery strategy, even when not explicitly asked to do so, then an explanation for our previous results based on subjects' use of images cannot be ruled out. Indeed, such an outcome would suggest that the distinction between a musical image and a musical representation may not be a valid or useful one.

Method

Subjects. Twenty undergraduates from the same pool used in previous experiments participated as subjects. Musical training, as measured by number of years of formal instrument or voice lessons, ranged from no training to 10 years of training, with a mean of 4.1 years and a median of 3.0 years. All subjects reported normal hearing; no subjects reported having absolute pitch.

Apparatus. The apparatus was the same as that used in Experiment 3.

Stimuli. Cue–probe pairs and masks were constructed in the same manner as in Experiment 3.

Procedure. The procedure was the same as that used in Experiment 3 except that subjects were not instructed to form an image. Instead, the subject's task was to decide as quickly and as accurately as possible whether the probe was one whole step above the cue or greater than one whole step above the cue, and to respond by pressing the appropriately marked key.

Results

Response times greater than 2,000 ms or less than zero were discarded (1.35% of the data). Mean RTs for correct responses were analyzed using a MANOVA for repeated measures, with two within-subjects factors: cue (tone vs. chord) and target (same-1, same-2, different-related, and different-unrelated).

As shown in Figure 3, a significant effect of target was found, $F(3, 17) = 3.43$, $p = .041$, RTs for same targets were faster than those for different targets, $F(1, 19) = 7.00$, $p = .016$, and there was a marginally significant trend for RTs for same-1 to be longer than those for same-2, $F(1, 19) = 4.06$, $p = .058$. The effects of cue and the Cue × Target interaction were not significant, *F*s = 0.85 and 1.69, respectively, both *p*s > .20.

Accuracy rates for each condition are also displayed in Figure 3. There was a main effect of target type, $F(3, 17) = 8.38$, $p = .001$. As in our previous experiments, accuracy was better for same targets than for different targets, $F(1, 19) = 5.71$, $p = .027$; accuracy was also better for different-unrelated targets than for different-related targets, $F(1, 19) = 24.93$, $p < .001$. The effects of cue and the Cue × Target interaction

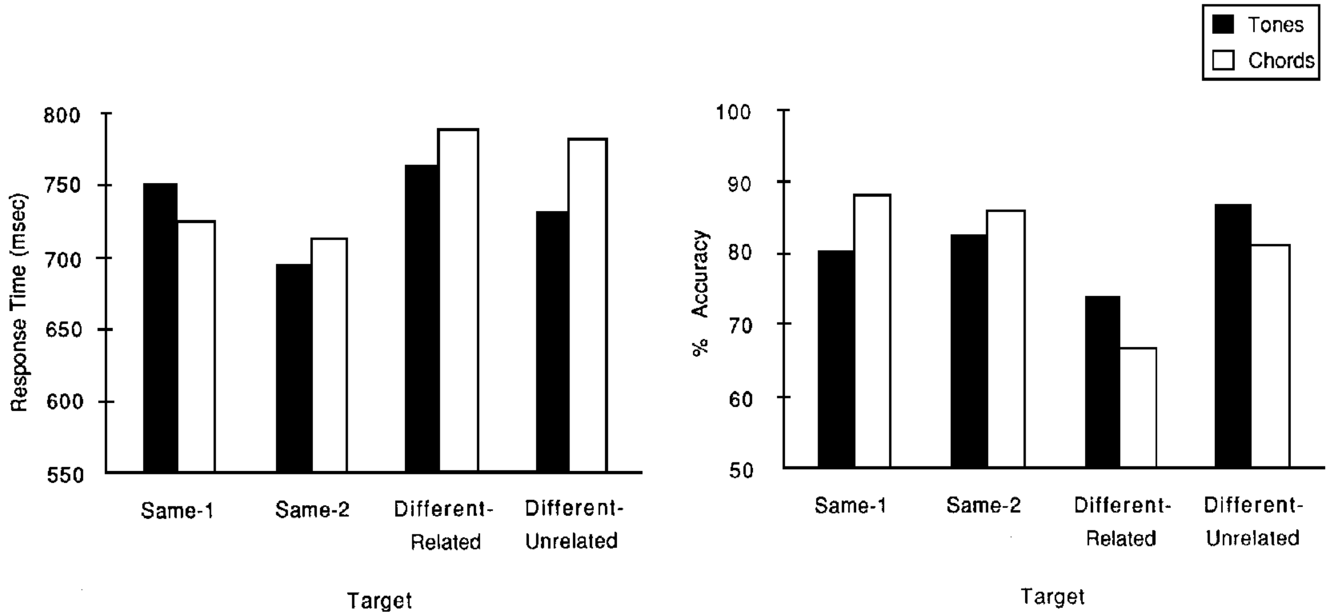

Figure 3. Response times and accuracy rates for same/different judgments in Experiment 4.

were not significant, *F*s = 0.03 and 1.22, respectively, both *p*s > .33.

As in Experiment 1, we reanalyzed RTs and accuracy rates with degree of musical training as an additional factor, using a median split to classify subjects as either highly trained or untrained. Again, we found a trend for highly trained subjects to have quicker RTs and greater accuracy rates, but there were no significant differences (all *F*s < 3.75, all *p*s > .06).

Discussion

The pattern of results is very similar to that obtained in Experiments 1 and 3. On the surface, these results suggest that imagery may not be a necessary component of this particular task. However, postexperimental debriefing of the subjects revealed that all of them had opted for an imagery strategy in the absence of specific instructions to the contrary. Some of them had "tried to hear the cue in my head," whereas others tried to "imagine what the probe might sound like." All subjects reported that they felt their accuracy was higher during the trials in which they used imagery. Despite our attempt to exclude imagery from the experiment, it appears that subjects used imagery as a default strategy and considered it the best method available.

The use of imagery by subjects in Experiment 4 may initially appear to be incongruous with the lack of imagery use by subjects in Experiment 2. In the earlier experiment subjects merely had to predict how long it would take other subjects to generate images of tones and chords. We stressed that they were not to actually image, but to guess. In Experiment 4, however, subjects were required to actively compare the pitches of the probe with those of the cue, and no instructions either for or against imagery use were given. The earlier experiment did not make any decisional demands requiring use of the harmonic properties inherent in an image, so subjects had no need of an image. In the latter experiment, however, subjects were required to make decisions that subjects in Experiment 2 did not have to make, that is, to decide whether the probe was one step in pitch higher than the cue. We suspect that it is only for this additional step, the same/different judgment, that subjects in Experiment 4 needed and formed an image.

A comparison of the response times and accuracy rates from Experiments 1, 3, and 4 showed that there were, in fact, differences among the three experiments [$F(2, 59) = 20.10$, $p < .001$, for response times; $F(2, 59) = 16.22$, $p < .001$, for accuracy]. Further analysis suggests that the presence of an interfering sequence (mask) in Experiment 3 resulted in a slowing of response time and a decrease in accuracy from Experiment 1 (Newman-Keuls, $p < .05$). Not surprisingly, the mask appears to disrupt image formation (as suggested by the longer IFTs in Experiment 3), resulting in a decrease in speed and accuracy. Experiment 4 did not differ in overall response time or accuracy rate from Experiments 1 or 3.

General Discussion

One of the primary findings of Experiments 1 and 3 is that subjects require more time to form images of chords, the physically more complex stimuli, than of tones, the less complex stimuli. Moreover, in both experiments the ratio of chord to tone image formation time was approximately the same, even though the absolute image formation times were longer in Experiment 3. The additional time required for chords, however, is not sufficient to suggest that subjects form images of chords by imaging each constituent tone separately and then serially combining the separate elements (as suggested by the results of Kosslyn et al., 1983). Although the greater amount of time required for chords is possibly due to their more complex structure, it is not clear how the imaging of chords proceeds. Chord images may be formed by imaging each of the constituent tones in parallel, or they may be formed by activating previously formed representations of entire chords. Imaging individual tones in parallel to form a chord may take more time than imaging a single tone because of the more complex nature of a chord representation or because additional time is needed to fuse the individual tone representations into a single complex musical entity. Had subjects formed images of simultaneously played tones that are not commonly played in combination, image formation times might have more closely resembled the results of Kosslyn et al. (1983), although it is difficult to understand why a chord might be a more well-formed whole than a picture of, for example, a mountain lion. Future studies may attempt to disentangle these issues by having subjects form images of single tones, two tones simultaneously, three tones simultaneously, and so forth and having tone combinations that either do or do not form commonly heard chords.

The context established by a tone or chord image influences the perception of subsequently presented tones and chords in much the same way as a context established by a tone or chord percept. In musical recognition studies, in which a prior context is induced with perceptual stimuli, accuracy is greater for same targets than for different targets, and accuracy for different-unrelated targets is greater than for different-related targets (e.g., Bharucha, 1984; Bharucha & Krumhansl, 1983; Cuddy et al., 1979; Dewar et al., 1977; Dowling, 1978); similar patterns were found in Experiments 1, 3, and 4. The similarity of the response patterns obtained using image stimuli with those obtained in musical recognition studies (perceptual stimuli) indicates that images are capable of establishing harmonic context. The ability of an image to establish such a context is consistent with the notion that the processes involved in imaging musical stimuli are very similar, if not identical, to those involved in the perception of such stimuli. In addition, after the formation of the image the presence of a mask was found to interfere with the comparison of the image to a subsequently presented tone or chord, a result paralleling that of earlier studies on memory for tones (Deutsch, 1982).

These results are also consistent with Farah and Smith's (1983) findings, in that imaging tones and chords appears to facilitate subsequent perception of the same tones or chords, as reflected by the overall greater accuracies and faster decision times for correct same responses relative to correct different responses. The results of Segal and Fusella (1970), Brooks (1968), and others who have found that images that occur simultaneously with to-be-perceived stimuli in the same modality interfere with recognition of those stimuli are not problematical because subjects were not required to maintain their images throughout the presentation of the probe.

The results from Experiment 2 did not support the argument that these results were simply due to demand characteristics or knowledge on the part of subjects. When asked to predict outcomes of imaging tones and chords, naive control subjects predicted no difference in the time to image tones and chords. Furthermore, the predicted image formation times were over three times larger than the actual image formation times found in Experiment 1; this difference supports the claim that the results in Experiment 1 were due to subjects' use of imagery and not some other nonimaginal strategy. Subjects were also not able to accurately predict the rank order of decision times in the different conditions. Whereas subjects were able to accurately predict faster responses for same targets, they did not accurately predict the order of response times for different targets, and their predictions for chords were less accurate than their predictions for tones and were statistically unreliable. In short, the predictions of nonimagery control subjects were quantitatively and qualitatively different from the performances of imagery subjects, which should not have happened if our imagery subjects were using tacit knowledge, as Pylyshyn (1981, 1984) might argue, to predict their results.

Additional support for the claim that subjects in Experiments 1 and 3 used imagery was garnered in Experiment 4, in which subjects used an imagery strategy even when not explicitly instructed to do so. Indeed, subjects reported that they could think of no other way to perform the task. It is not surprising, then, that their results look very much like those found in the preceding imagery experiments. It is difficult, however, to conceive of a memory store for pitch that is not closely tied to imagery. Although it is possible that some people, perhaps highly trained musicians, are able to encode and store musical stimuli propositionally, the majority of musically illiterate or subliterate people probably cannot. For these people, the only way to store and retrieve musical representations might be through the use of auditory imagery. It is not clear how to separate the differences, if any, between the use of imagery and the use of some other form of stored pitch memory (cf. Anderson, 1978). Similar difficulties are encountered in the attempt to separate visual imagery and memory (Hubbard, Kall, & Baird, in press).

In conclusion, our data support a functional role for musical imagery in cognition (cf. Marschark, Richman, Yuille, & Hunt, 1987) and suggest effects of structural complexity on image formation. Subjects are able to form and use images of musical tones and chords. The time required to form musical images is related to the complexity of the images—that is, complex chord images require more time than simple tone images. Furthermore, imaged tones and chords are governed by the same internalized regularities of music—that is, harmonic relationships—that constrain perception of music. The patterns found in the accuracy rates mirror the patterns found in analogous music perception tasks, and the patterns of response times are consistent with the accuracy rates and with predictions made from the music perception literature. This similarity of response patterns is consistent with the claims that imagery uses processing mechanisms similar, if not identical, to those used in perception (Finke, 1985; Neisser, 1976; Shepard & Podgorny, 1978); specifically, music imagery invokes the same processing mechanisms used in music perception.

References

Anderson, J. R. (1978). Arguments concerning representations for mental imagery. *Psychological Review, 85*, 249–277.

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

Bharucha, J. J. (1987). MUSACT: A connectionist model of musical harmony. In *Proceedings of the Ninth Annual Meeting of the Cognitive Science Society*. Hillsdale, NJ: Erlbaum.

Bharucha, J. J., & Krumhansl, C. L. (1983). The representation of harmonic structure in music: Hierarchies of stability as a function of context. *Cognition, 13*, 63–102.

Bharucha, J. J., & Stoeckig, K. (1986). Reaction time and musical expectancy: Priming of chords. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 403–410.

Bharucha, J. J., & Stoeckig, K. (1987a). *Harmonic priming as a function of harmonic relatedness*. Manuscript in preparation.

Bharucha, J. J., & Stoeckig, K. (1987b). Priming of chords: Spreading activation or overlapping frequency spectra? *Perception & Psychophysics, 41*, 519–524.

Block, N. (Ed.). (1981). *Imagery*. Cambridge, MA: MIT Press.

Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Psychology, 22*, 349–368.

Cuddy, L. L., Cohen, A. J., & Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology, 33*, 148–157.

Denis, M., & Carfantan, M. (1985). People's knowledge about images. *Cognition, 20*, 49–60.

Deutsch, D. (1982). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (pp. 271–316). New York: Academic Press.

Deutsch, D., & Feroe, J. (1981). The internal representation of pitch sequences in tonal music. *Psychological Review, 88*, 503–522.

Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. K. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory, 3*, 60–67.

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memories for melodies. *Psychological Review, 85*, 341–354.

Dowling, W. J., & Harwood, D. L. (1986). *Music cognition*. New York: Academic Press.

Farah, M. J. (1985). Psychophysical evidence for a shared representational medium for mental images and percepts. *Journal of Experimental Psychology: General, 114*, 91–103.

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

Finke, R. A. (1985). Theories relating mental imagery to perception. *Psychological Bulletin, 98*, 236–259.

Hubbard, T. L., Kall, D., & Baird, J. C. (in press). Memory, imagery, and size-distance invariance. *Memory & Cognition*.

Intons-Peterson, M. J. (1980). The role of loudness in auditory imagery. *Memory & Cognition, 8*, 385–393.

Kosslyn, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.

Kosslyn, S. M., Reiser, B. J., Farah, M. J., & Fliegel, S. L. (1983).

Generating visual images: Units and relations. *Journal of Experimental Psychology: General, 112,* 278–303.

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

Krumhansl, C. L., Bharucha, J. J., & Kessler, E. J. (1982). Perceived harmonic structure of chords in three related musical keys. *Journal of Experimental Psychology: Human Perception and Performance, 8,* 24–36.

Marschark, M., Richman, C. L., Yuille, J. C., & Hunt, R. R. (1987). The role of imagery in memory: On shared and distinctive information. *Psychological Bulletin, 102,* 28–41.

Mitchell, D. B., & Richman, C. L. (1980). Confirmed reservations: Mental travel. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 58–66.

Neisser, U. (1976). *Cognition and reality.* New York: W. H. Freeman.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316–333.

Pylyshyn, Z. W. (1981). The imagery debate: Analogue media versus tacit knowledge. *Psychological Review, 88,* 16–45.

Pylyshyn, Z. W. (1984). *Computation and cognition.* Cambridge, MA: MIT Press.

Segal, S. J., & Fusella, V. (1970). Influence of imaged pictures and sounds on the detection of visual and auditory signals. *Journal of Experimental Psychology, 83,* 458–464.

Shepard, R. N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America, 36,* 2346–2353.

Shepard, R. N. (1982). Structural representations of musical pitch. In D. Deutsch (Ed.), *The psychology of music* (pp. 343–390). New York: Academic Press.

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

Shepard, R. N., & Podgorny, P. (1978). Cognitive processes that resemble perceptual processes. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5, pp. 189–237). Hillsdale, NJ: Erlbaum.

Siegel, S. (1956). *Nonparametric methods for the behavioral sciences.* New York: McGraw-Hill.

Weber, R. J., & Brown, S. (1986). Musical imagery. *Music Perception, 3,* 411–426.

Weber, R. J., & Harnish, R. (1974). Visual imagery for words: The Hebb Test. *Journal of Experimental Psychology, 102,* 409–414.

Received September 10, 1987
Revision received December 14, 1987
Accepted January 4, 1988 ■

1989 APA Convention "Call for Programs"

The "Call for Programs" for the 1989 annual APA convention will be included in the October issue of the *APA Monitor.* The 1989 convention will be in New Orleans, Louisiana, from August 11 through 15. Deadline for submission of program and presentation proposals is December 15, 1988. This earlier deadline is required because many university and college campuses will close for the holidays in mid-December and because the convention is in mid-August. Additional copies of the "Call" will be available from the APA Convention Office in October.