1988

# Mental Scanning in Auditory Imagery for Songs

Andrea R. Halpern
*Bucknell University*, ahalpern@bucknell.edu

## Recommended Citation

Halpern, Andrea R.. "Mental Scanning in Auditory Imagery for Songs." *Journal of Experimental Psychology: Learning, Memory, and Cognition* (1988) : 434-443.

Journal of Experimental Psychology:
Learning, Memory, and Cognition
1988, Vol. 14, No. 3, 434–443

# Mental Scanning in Auditory Imagery for Songs

Andrea R. Halpern
Bucknell University

Four experiments examined how people operate on memory representations of familiar songs. The tasks were similar to those used in studies of visual imagery. In one task, subjects saw a one-word lyric from a song and then saw a second lyric; then they had to say if the second lyric was from the same song as the first. In a second task, subjects mentally compared pitches of notes corresponding to song lyrics. In both tasks, reaction time increased as a function of the distance in beats between the two lyrics in the actual song, and in some conditions reaction time increased with the starting beat of the earlier lyric. Imagery instructions modified the main results somewhat in the first task, but not in the second, much harder task. The results suggest that song representations have temporal-like characteristics.

How is it that regardless of musical training, people can remember thousands of songs for many years? Researchers in the psychology of music have offered many answers: We remember tunes by the scale system, contour (Dowling, 1978), harmony (Bharucha & Krumhansl, 1983), interval configurations (Cuddy & Cohen, 1976), or by other abstract schemes (Chew, Larkey, Soli, Blount, & Jenkins, 1982). These studies take the perspective that tune representations are schematized to some degree.

The goal of the current research is to emphasize the literalness of tune representations, much in the spirit of research on visual imagery. Compelling evidence has been presented by the previously mentioned researchers and others (reviewed by Shepard & Cooper, 1982) that visual images have some characteristics of visual perception. Mental imaging takes real time (Kosslyn, Ball, & Reiser, 1978), is subject to adaptation effects (Finke, 1979), and so forth. The present research extends some of the approaches and methods of visual imagery to auditory imagery. Not much attention has heretofore been paid to the image-like characteristics of songs, speech, or other auditory experiences. In fact, "mental imagery" has become synonymous with "visual imagery." The current research attempts to validate the common impression that music "plays" inside our heads, with a tempo and seriality similar to that of actual songs.

Although little previous research has been carried out on auditory images for songs, some studies have looked at imagery of other kinds of auditory stimuli. For example, Intons-Peterson (1980) looked at the loudness of imagined sounds and found some effects noted in the visual imagery literature, such as a symbolic distance effect, but she failed to find other effects of loudness that she predicted if loudness were analogous in imagery and perception. For instance, subjects did not take longer to generate louder versus softer imagined sounds. Intons-Peterson concluded that loudness is optionally represented in an auditory image. (Note that visual imagery theorists would find it difficult to conclude that even separable dimensions such as size and shape could be optionally represented in a visual image.) Farah and Smith (1983) found that imagining a tone helped subjects to detect it in a subsequent signal detection task. They concluded that imagery and perception have similar representations of frequency. Weber and Brown (1986), who did look at auditory imagery for songs, found that describing the pitch contour of melodies took about the same amount of time and produced comparable error rates whether the tune was perceived or imagined.

These studies are somewhat equivocal in clarifying the similarity between auditory imagery and perception. The current article is an initial attempt to establish several auditory imagery phenomena that could be predicted from the assumption that auditory imagery will share characteristics with auditory perception. Although music is only one example of an auditory stimulus, it was chosen both because people claim to have imagery experiences with music and also because research exists establishing some of the parameters operating when people actually listen to music.

Although representations of visual and auditory stimuli will reflect modality-specific differences, we can still examine some analogies between the two domains. Intrinsic to any visual stimulus or representation is its spatial extent or spatial-like qualities. Hence, quite a few studies in the visual imagery literature attempt to show behavior during imaging that implies use of a spatial-like representation. The scanning study of Kosslyn et al. (1978) is a good example. In a series of experiments, subjects performed tasks comparing separate locations on a memorized array, map, or other visual input. Reaction time to complete the tasks increased with increasing distance of the locations in the actual stimulus, as long as subjects were instructed to use visual images.

Analogous to vision, intrinsic to any auditory stimulus is

---

This research was partially supported by National Science Foundation Grant BNS-8607405.

I am grateful to the following students for their help during this project: Melissa Connor, Kathleen Darrow, Kathy Kron, Susan Milton, Lari Roberts, Maureen Schmitter, and Karralynn Schreck. Helpful comments on previous versions of this article were provided by Jay Dowling, Bill Prinzmetal, Marilyn Smith, Robert Greene, Robert Crowder, and several anonymous reviewers.

Experiment 1 was presented at the meeting of the Eastern Psychological Association, April 1986, in New York.

Correspondence concerning this article should be addressed to Andrea R. Halpern, Psychology Department, Bucknell University, Lewisburg, Pennsylvania 17837.

its temporal extent and ordering. For example, tunes obviously have a beginning, middle, and end. One can easily experience the difficulty of trying to violate this intrinsic ordering. Try to start singing the song "Jingle Bells" from the second phrase ("O'er the fields we go . . .") *without* thinking of or running through the first phrase of the song. Similarly, try to hum any song from the last note toward the first. People find it difficult even to recognize the relationship between a tune and its backward version (Dowling, 1972).

If an auditory representation is analogous to an auditory stimulus, then we may hope to find evidence for temporal extent in the representation. The approach taken here is modeled on the Kosslyn et al. (1978) visual scanning study. Subjects were presented with two words from familiar songs and then asked to verify that the second lyric is from the same song as the first (Experiments 1 and 2) or asked to compare the pitches of the notes referred to by the lyrics (Experiments 3 and 4). (In this article, *lyric* refers to one word of a song.) The crucial variable is the distance between the two lyrics in a trial. In this case, *distance* is defined as the number of beats separating the lyrics in the actual song. If people process parts of the song representation between the two designated lyrics, then time to perform the task should increase with distance.

A secondary variable is the starting point of the first lyric of the pair. If the song representation is strongly ordered, then instructions to enter the representation from, say, Beat 5, may be difficult to carry out. In that case, we may see increasing reaction times as a function of starting point, as if the subject would need to "run through" the preceding information, as in the "Jingle Bells" example.

Another variable of interest is the extent to which imagery instructions will affect performance in the task. If subjects appear to be using imagery when so instructed, will they also use it spontaneously? In Kosslyn et al.'s (1978) study, the increase of reaction time with distance only appeared when subjects were instructed to imagine "a little black speck zipping in the shortest straight line from the first object to the second" (Kosslyn et al., 1978, p. 52). Subjects seemed to use a nonimagery representation when imagery instructions were omitted. That is, the time to say whether a second object was on an imagined map after focusing on an initial object was approximately a constant, independent of the distance between the objects. In both tasks used here, the points in the songs that subjects were required to attend to were indexed by single-word lyrics that could potentially be randomly accessed. Random access is also compatible with an imagery representation, as Friedman (1983) found for representations for the days of the week. As is evident in what follows, the tasks used here can be performed—and probably more efficiently—if subjects can randomly access song elements. However, if reaction time nevertheless increases with increasing distance between lyrics, then the analogue nature of song processing appears to be more obligatory.

Finally, attempts were made to correlate auditory imagery ability and musical training with performance on the tasks. Although the studies were designed primarily to investigate the internal musical experience of ordinary people, musicians might show differential performance as a result of superior training and/or inborn abilities.

To summarize, Experiments 1 and 2 were similar to the Kosslyn et al. (1978) scanning task, but in the auditory mode. Experiments 3 and 4 used a mental pitch comparison task, which provided a stricter test of whether subjects would display a response pattern suggestive of auditory imagery.

## Experiment 1

This experiment was the initial attempt to establish whether a mental scanning task made sense using songs. Subjects initially saw one song lyric. After a second lyric appeared, they were to mentally scan the song (if in fact the second lyric was from the same song) and to press a button when they "arrived." Half the subjects received imagery instructions and half did not.

### Method

*Subjects.* Forty-one Bucknell University undergraduates served as subjects for course credit. In this and the following experiments, subjects were required to have been raised in the United States in order to maximize the chances of their being familiar with the stimulus songs.

*Materials.* Lyrics to the beginning of three songs familiar to this population served as stimuli: "Do Re Mi" (from *The Sound of Music*), "Hark the Herald Angels Sing," and "The Star-Spangled Banner" (the American national anthem). These were chosen because each had the beginning or only syllable of a lyric falling unambiguously on Beats 1, 3, 5, 7, 9, 11, and 13 of the song. For instance, Beats 1, 3, and 5 for "The Star-Spangled Banner" fell on the words *oh*, *can*, and *see*, respectively. In addition, each song had unique words in its first phrase, so that a particular lyric referenced only one place in the song. These requirements severely limit the stimulus set, because even the word "the" cannot be used in a trial if it appears more than once in the relevant phrase of the song. "Twinkle Twinkle Little Star" served as a practice song.

For each song, lyrics beginning on Beats 1, 3, and 5 (the variable referred to as *startpoint*) were paired with Lyrics 2, 4, 6, or 8 beats away (*stepsize*). The three startpoints, four stepsizes, and three songs made for 36 true trials. During the experimental session, each trial was repeated once. In addition, a false trial was yoked to each true trial by replacing the second lyric in the pair by one resembling the first lyric. For example, the yoked trial to "resun" was "remoon." In no studies did the song title also appear as a lyric.

*Procedure.* Presentation of stimuli and recording of responses were controlled by an Apple II Plus computer. Each trial consisted of presentation of the name of one of the three songs centered on a video monitor, followed 1 s later by the first lyric of a pair and, after 750 ms, by the second lyric. In the imagery condition, subjects were instructed to first mentally focus on the first lyric. Then, if the second lyric was not a lyric in the song, they were to press the "false" button on a response board as quickly and accurately as possible. If the second lyric was indeed in the song, subjects followed imagery instructions to start at the first lyric and mentally play the song until they arrived at the second lyric. At that time, they were to press the "true" button. In the nonimagery condition, subjects were told to mentally focus on the first lyric, and when the second lyric appeared, to indicate whether it was or was not an actual song lyric by pressing the appropriate button as quickly and accurately as possible. Subjects initiated each trial by pressing a third button.

Subjects used their dominant hand for responding, and the left/right position of the response buttons was reversed for the half the subjects. Both speed and accuracy were stressed.

Each session began with the experimenter showing a list of the stimulus songs to the subjects and asking if the songs were familiar. Only subjects indicating familiarity with the songs participated. The experimenter then showed them the lyrics of the beginning part of each song. Subjects were asked to clap out the beats while reciting the lyrics in order to verify that the experimenter and subjects agreed on beat placement. Six practice trials with "Twinkle Twinkle" ensued, each with feedback. The 108 experimental trials, without feedback, followed in a different random order for each subject.

### *Results*

The dependent variable of most interest was the mean reaction time for each correct "true" trial, recorded from the onset of the second lyric. For all studies, both the raw reaction times and log transformations of reaction times were analyzed. Only the raw data analyses are reported, because the transformations made no difference in results. If all observations for a cell were unavailable as a result of errors, the missing cell value was always estimated using multiple regression. The overall error rate in this experiment was approximately 6%.

Reaction times were collapsed over melodies and analyzed via a three-way analysis of variance (ANOVA): Groups was the between factor, and stepsize and startpoint were within factors. As seen in Figure 1, there was a main effect for group: the imagery group ($M = 4640$ ms) took longer to respond than did the control group, $M = 4098$ ms, $F(1, 39) = 6.2$, $p < .05$. Reaction time increased with stepsize, $F(3, 117) = 28.0$, $p < .001$. The pattern of increase had a large linear component, $F(1, 39) = 51.7$, $p < .001$, and a smaller quadratic component, $F(1, 39) = 10.8$, $p < .01$. Examining the figure, we see that reaction time increases more quickly with larger stepsizes. Reaction time also increased with startpoint, $F(2, 78) = 10.5$, $p < .001$, qualified by a Startpoint × Stepsize interaction, $F(6, 234) = 6.1$, $p < .001$. Stepsize 2 trials seem to be the source of the interaction; reaction time increases with startpoint in quite an orderly fashion for the other stepsizes. This was true for both groups, given that the Group × Startpoint × Stepsize interaction was not significant.

The one remaining significant effect was a Group × Startpoint interaction, $F(2, 78) = 3.5$, $p < .01$. As seen in the figure, reaction time clearly increases with startpoint under imagery instructions. Little effect of startpoint is apparent when imagery instructions are omitted.

### *Discussion*

With or without imagery instructions, reaction time increased with an increasing number of beats in the real stimulus song. This finding suggests that subjects were operating on an analogous mental representation of the song in this task by mentally playing the songs when the second lyric was recognized as being part of the song. This increase was not uniform: Lyrics close together seemed to be scanned proportionally faster than did those farther away. The increase in reaction time with startpoint in the imagery group implies that subjects are slower to process parts of the song that occur farther from the beginning—as if they need to run through the beginning part of the song before beginning the trial. Note that the imagery instructions made no mention of starting the song from the beginning. The lack of a startpoint effect in the control group suggests that subjects do not always need to

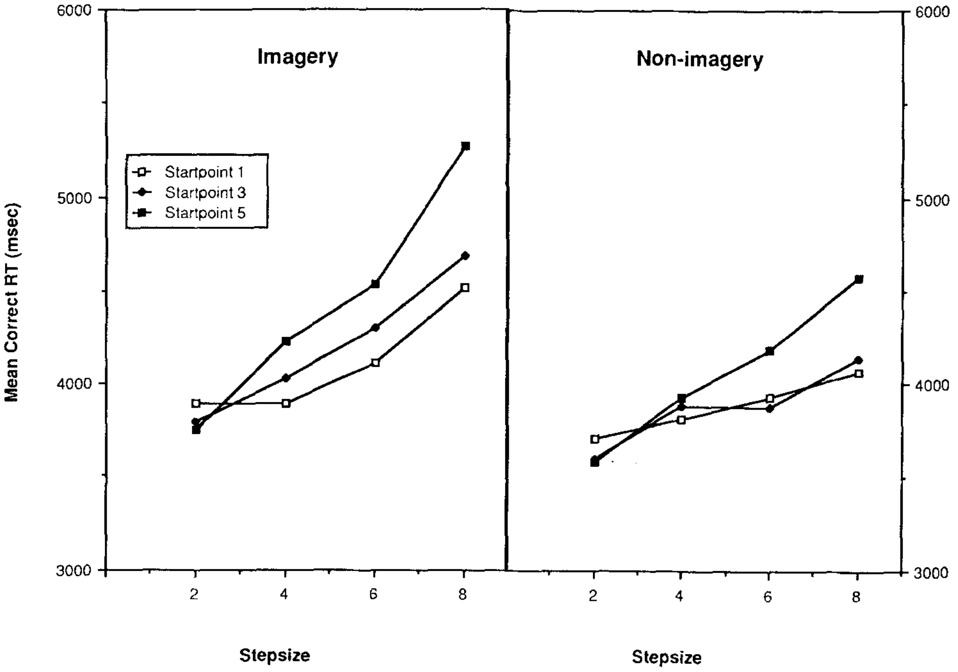

*Figure 1.* Reaction time (RT) as a function of startpoint, stepsize, and instruction group in Experiment 1.

process the beginning of a song when mental work is required on a middle portion of it.

### Replications

The results of Experiment 1 replicated those of a preliminary study that only used the imagery condition of Experiment 1.

Although the lack of a Group × Stepsize interaction in Experiment 1 implies that both groups showed an increase of reaction time with stepsize, the increase was clearer for the imagery group. Typically in the visual imagery literature, subjects are interviewed postexperimentally to ascertain compliance with imagery or nonimagery instructions. Perhaps some subjects in the control group were using imagery and some were not. Thus, a second replication of Experiment 1 was run that included an instruction-compliance questionnaire. Another subject characteristic that may affect the results of interest is auditory imagery ability. In the visual imagery literature, assessed vividness of imagery has sometimes affected experimental results (e.g., Finke & Kosslyn, 1980). Because no suitable questionnaire about auditory imagery ability was readily available, one was developed for this study. In order to determine if reaction-time patterns could be the result of cooperation with the experimenter the questionnaire also asked subjects whether they knew the purpose of the study.

Subjects were 40 Bucknell undergraduates. The procedure for the main part of the experiment was identical to that in Experiment 1. In order not to emphasize imagery experiences to the control subjects, subjects were given an auditory-imagery-ability questionnaire after the experimental session. This questionnaire consisted of brief descriptions of three scenarios: a classroom, a visit to a friend, and a baseball game. For each scenario, subjects were asked to imagine two typical sounds and, using a 1 to 10 scale, to rate each on clarity (*not clear at all* to *very clear*) and ease of imagining (*very easy* to *difficult*). For instance, in the third scenario, respondents were asked to imagine the roar of the crowd and the crack of the bat during a hit.

The imagery-use questionnaire simply asked subjects whether they "mentally played the song from Lyric 1 to Lyric 2" and what percentage of the time they estimated they did this. They were also asked to guess the hypothesis of the experiment.

Most results of Experiment 1 were replicated, with the following exceptions: The imagery group took longer to complete the task than did the control group only when outlier reaction times were removed, and there was no Group × Startpoint interaction.

No subject guessed the hypothesis of the experiment in enough detail to warrant being disqualified. The control group comprised a wide range of reported imagery use, with modes of 40% and approximately 75%. The mean of reported imagery use was 73% for the imagery group and 63% for the control group; this was not a significant difference.

Even if reported use of imagery was widespread in both groups, perhaps we would see a group difference if we regrouped the subjects by amount of reported imagery. Therefore, subjects were divided into two new groups: (a) Those who reported they used imagery less than 70% of the time (approximately the mean of reported imagery use, $n = 14$) and (b) those reporting imagery 70% of the time or greater ($n = 7$), regardless of assignment to the original instruction group. Neither the (new) group main effect nor interactions involving group were significant. In other words, subjects were behaving similarly (i.e., showing imagery-like effects) regardless of what they reported doing. It is our argument here that the strategy of choice was accessing an analogue representation of the song, where points between the lyrics of interest were processed or scanned.

The reliability of the auditory-imagery questionnaire proved to be acceptable, $\alpha = .75$. Overall item means (with the scores on the difficulty scale reversed) ranged from 7.72 (clarity of *bat cracking*) to 9.22 (clarity of *a car door slamming shut*) out of 10.

Each subject's imagery score was the average of the six scenario items. Imagery scores did not differ between the imagery ($M = 8.83$) and control ($M = 8.76$) instruction groups. Neither did subjects' imagery scores correlate with their overall reaction time in the main task.

In summary, then, subjects appear to be using imagery strategies even when not instructed to do so. This is supported both by self-reports and the similarity of reaction-time patterns between imagery and nonimagery groups.

## Experiment 2

In all versions of the scanning task reported so far, startpoint interacted with stepsize. That is, for lyrics only separated by two beats, the starting beat of the first lyric had little effect on reaction times. For lyric pairs separated by a greater number of beats, startpoint had an increasing effect. The reaction times reported so far comprised the interval between the onset of the second lyric and the button press. They do not directly assess the time taken to reach the first lyric. This interval from the onset of the first lyric to an acknowledgement of having mentally focused on the first lyric (Reaction Time 1 [RT1]) is measured directly in this experiment to demonstrate the startpoint effect less ambiguously. The second lyric is not presented until the subject claims to be focusing on the first lyric. Then, time between mentally focusing on the first lyric and scanning to the second lyric is measured (Reaction Time 2 [RT2]). This time interval should be unaffected by startpoint (which in a sense the subject is controlling for), but should be strongly affected by stepsize.

To summarize, this experiment differs from Experiment 1 in that reaction time is decomposed into two components. In this way, the contribution of startpoint and stepsize can be assessed separately.

### Method

*Subjects.* Thirty Bucknell University students served as subjects, half in the imagery group and half in the nonimagery control group.

*Procedure.* Materials and procedures were the same as in Experiment 1, except for the following changes: All subjects were instructed to mentally focus on the first lyric, as before. Once they had comfortably focused on this lyric they were instructed to press a "focus" button on the response board. Pressing this button caused the second lyric to appear on the monitor. Subjects in the nonimagery group were instructed to press as quickly and accurately as possible the "no" button if the second lyric was not in the tune and the "yes" button if the second lyric was in the tune. Subjects following the imagery instructions were told to continue with the lyric they had been focusing on and mentally play the rest of the tune until they arrived at the second lyric. At this time they were to press the yes button. If the lyric was not in the song they were instructed to press the no button as quickly as possible. All subjects initiated each trial by pressing a start button.

## *Results*

Separate analyses were performed on RT1 and RT2 because each measure was calculated from a different group of trials. RT1 was based on all trials, because at the time the first lyric was presented, subjects had no way of knowing whether the trial would be a true or a false one. RT2 was based only on correct true trials, as before. In a small percentage of trials (less than 2%), subjects displayed abnormally short RT1 values of less than 900 ms, which suggested they had a false start when they pressed the focus button (overall $M = 2,348$ ms). Subjects also reported that they occasionally pressed the focus button prematurely. These data were eliminated from analysis. The error rate of the remaining trials was about 5%.

*RT1.* Only startpoint and group were defined for this dependent measure, because stepsize is only defined after presentation of the second lyric. A two-way ANOVA was performed with startpoint as the within factor and group as the between factor. The only significant effect was the predicted increase of RT1 with startpoint, $F(2, 56) = 11.6, p < .001$ (see Figure 2). The apparent group difference suggesting more rapid task performance in the nonimagery group was not reliable. The lack of a Group × Startpoint interaction is understandable in that the two groups were not given differential instructions about how to perform the task during the interval of focusing on the first lyric.

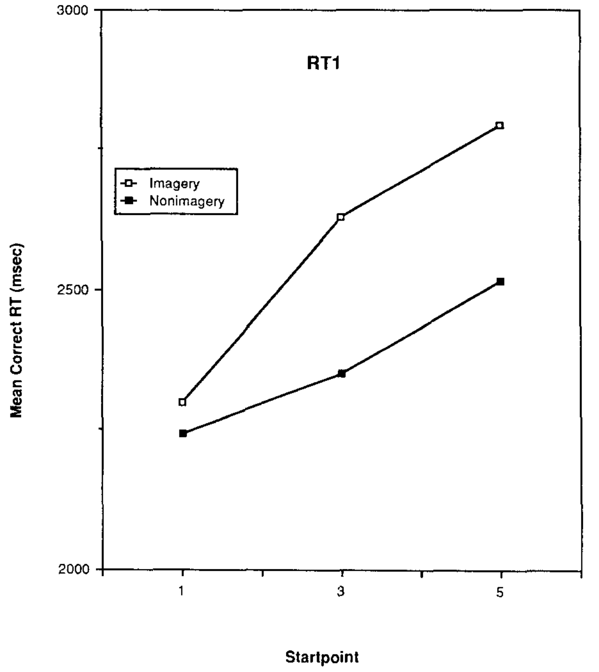

*Figure 2.* Reaction Time 1 (RT1) as a function of startpoint in Experiment 2 for each group.

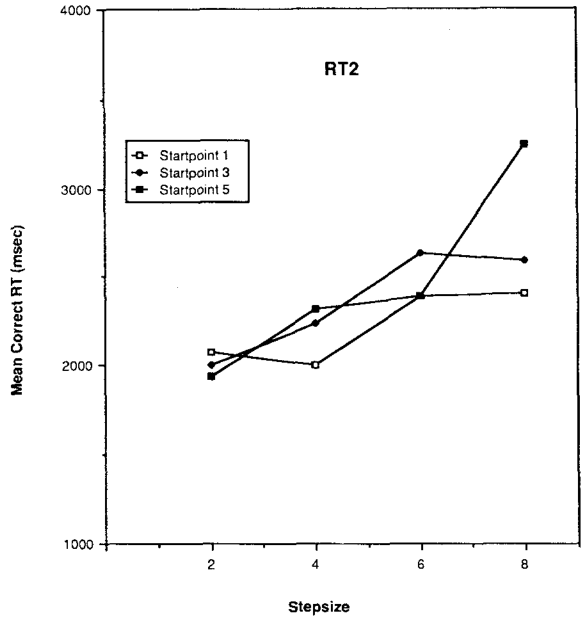

*Figure 3.* Reaction Time 2 (RT2) as a function of stepsize and startpoint in Experiment 2, collapsed across groups.

*RT2.* Startpoint and stepsize were analyzed as within factors, and group once again was analyzed as a between factor. As predicted, RT2 increased with stepsize, $F(2, 56) = 31.1, p < .001$. This increase was linear, as only the linear orthogonal contrast was significant, $F(1, 84) = 92.3, p < .001$. As with RT1, no main effect of group emerged nor did any interactions with group.

Contrary to prediction, RT2 increased with startpoint, $F(2, 56) = 5.8\ p < .01$. An interaction of startpoint and stepsize, $F(6, 168) = 6.9, p < .001$, is illustrated in Figure 3. Except for the anomalous data point at Stepsize 6, Startpoint 5, the figure is similar to Figure 1. Little effect of startpoint is apparent at Stepsize 2, whereas the effect is more noticeable at larger stepsizes.

## *Discussion*

Experiment 2 confirmed that the startpoint effect can be measured directly. That is, when told to focus on the first lyric, subjects took more time if the lyric is farther into the song. One might argue that the increase in time is solely a reflection of the familiarity of the lyric: First lyrics are more familiar than third lyrics, and so forth. This explanation loses plausibility when we consider that only lyrics on Beats 1, 3, and 5 were probed. These are all words in the first phrases of very familiar songs. In addition, subjects were reminded of the song lyrics in the familiarization phase. Over the course of the experiment, each first lyric was presented an additional 36 times. Thus, familiarity differences seem unlikely. The reaction time increases of over 100 ms for each of the startpoints seem more likely due to the tendency to process parts of the song between the lyrics.

The increase of RT2 with stepsize shows that more time is taken to scan between more distant lyric pairs than between closer lyric pairs. The effect of startpoint on RT2 suggests that subjects were not always scanning directly from the first lyric to the second. We may infer that at least on some trials, subjects returned to the beginning of the song before rendering an answer. Alternatively, familiarity may have played more of a role in determination of RT2 than seemed logical for RT1. The thirteenth beat of a tune may indeed be substantially less familiar than earlier beats. As one example, the thirteenth beat of the "The Star Spangled Banner" is "What [so proudly . . .]," which is the first word of a phrase and may be hard to comprehend as an isolated word. The interaction of startpoint and stepsize suggests that these two factors may both have been operating, such that at large stepsizes, subjects have more need to double-check their answers by returning to the beginning of the song.

It should be noted that the results in the scanning task do not imply a specifically musical representation. Subjects could have performed the task by serially accessing a list of the correct lyrics or perhaps by using a list of visual representations of lyrics or notes. The next experiment probes specifically for a musical representation of the songs by using a mental pitch comparison task. The pitch comparison task also *requires* subjects to attend to and process the first lyric in each trial, whereas previously, a subject could produce the correct answer without following instructions to mentally focus on the first lyric.

## Experiment 3

In addition to being a musical task, Experiment 3 differs from the first two experiments by using an expanded set of songs. Also, the amount of time that subjects see the first lyric was reduced from 750 ms to 500 ms in an effort to elicit the startpoint effect at Stepsize 2. Perhaps lyrics only two beats away from each other are so strongly associated (by simple proximity, or by semantic or syntactic relations) that a long viewing time allows direct access to closely related lyrics. A shorter viewing time might require subjects to access each lyric in turn and thus reflect differences in processing time.

### *Method*

*Subjects.* Thirty-one Bucknell University undergraduates participated for course credit. All subjects indicated familiarity with the stimulus songs. Data from an additional seven participants were discarded because of high error rates, as well as data from one other student because she sang the "imagined" songs during the session.

*Materials.* Eight songs were selected that fulfilled the requirements of familiarity and beat placement noted in the previous experiments. These included the following songs: "Rudolf the Red-Nosed Reindeer," "Puff the Magic Dragon," "Raindrops Keep Fallin' On My Head," "White Christmas," "Somewhere Over the Rainbow," "Battle Hymn of the Republic," "I'm Looking Over a Four-Leaf Clover," and "The Star Spangled Banner." Six trials were constructed for each of the 12 trial types (three startpoints and four stepsizes). For each trial, the two component lyrics represented different pitches in the actual song. It proved impossible to have every song represented once and only once in each cell (hence the need for a pool of eight songs). However, each song was used between seven and ten times within the 72 trials. Because some songs had many more or fewer second-lyric-higher trials than did the other type as a result of the song's pitch contour, eight unscored trials were added so that, for instance, "White Christmas" was not associated with a correct answer of "higher" in seven out of eight trials. Overall, there were 45 higher and 35 lower trials.

*Procedure.* The apparatus was the same as used in Experiments 1 and 2. For each trial, subjects saw a song title, followed by the first lyric. After 500 ms, the second lyric appeared. Subjects in the imagery group were told to "begin with the first lyric and play through the song in your mind until you reach the second lyric." Both groups were told to compare the pitch of Lyric 2 with that of Lyric 1 and press either the higher or lower button on a response board. As before, subjects initiated each trial, and the position of the response buttons was reversed for half the subjects. Both speed and accuracy were stressed.

Each session began as before, with familiarization of each song's lyrics and beat placement. Subjects were also asked if they understood what "higher or lower in pitch" meant, and they were asked which of two pitches sung by the experimenter was higher. All subjects seemed comfortable with the concept of pitch height.

After all instructions had been given, eight practice trials with feedback ensued. This was followed by the 80 experimental trials in a different random order for each subject. After the experiment, subjects were asked about their musical background.

### *Results*

One trial was eliminated from analysis ("Raindrops," Startpoint 5, Stepsize 8), because inadvertently the familiarization phase did not present the song as far as the second lyric, and nearly all subjects erred. Thus a total of 71 trials was analyzed.

This experiment differed dramatically from the previous ones in its difficulty. As already noted, data from 7 subjects were discarded because of error rates of 40% or over. The mean error rate of the remaining subjects was 24%. This substantial error rate allowed analysis of two dependent measures: mean correct reaction time for each cell and error rate in each cell. Because reaction time and error rate were uncorrelated in this and the following experiments, univariate rather than multivariate analyses are reported.

A three-way ANOVA on mean correct reaction time examined the stepsize, startpoint, and group effects. No group main effect nor any interaction involving group reached significance. The trend in the group means was similar to that found previously (imagery $M = 7{,}551$ ms; control = 6,303 ms); however, this apparent difference was not reliable. Both means and standard deviations were approximately twice as large as in Experiments 1 and 2. Figure 4 displays reaction times collapsed across groups.

Once again, reaction time increased as a function of stepsize, $F(3, 87) = 36.5$, $p < .001$, and startpoint, $F(2, 58) = 28.0$, $p < .001$. The usual Startpoint × Stepsize interaction, $F(6, 174) = 3.0$, $p < .01$, can be seen even more clearly than in the previous experiments: The startpoint separation emerges only for stepsizes greater than 2. The increase of reaction time with stepsize was largely linear, $F(1, 29) = 64.5$,

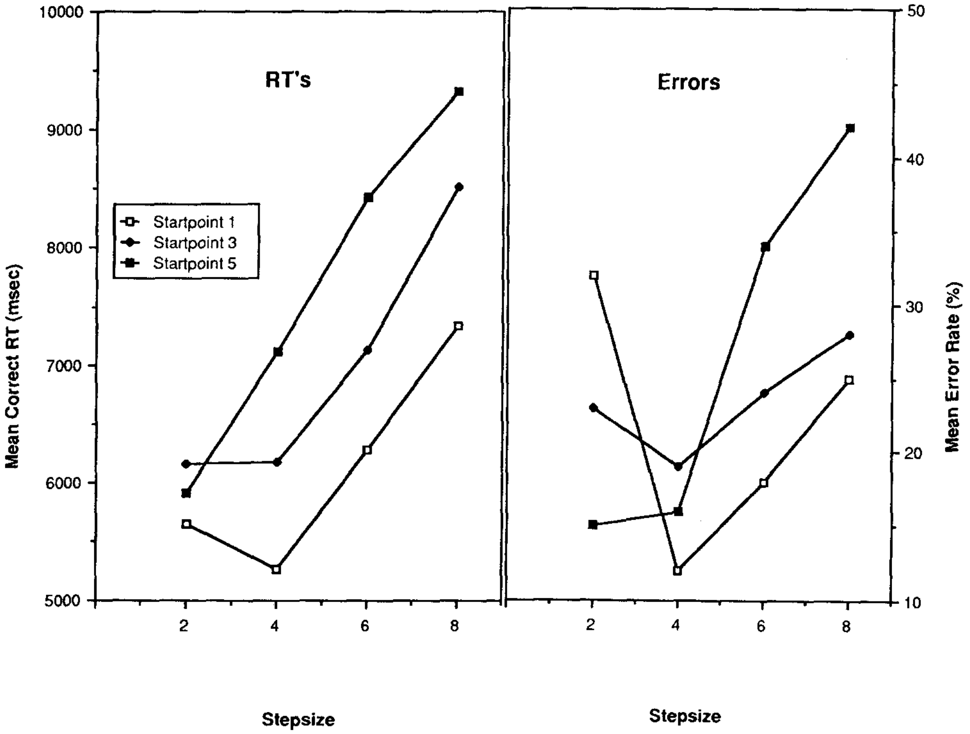

*Figure 4.* Reaction time (RT) and error rate as a function of startpoint and stepsize, collapsed across groups, in Experiment 3.

$p < .001$, but also included a quadratic component, $F(1, 29) = 8.4, p < .001$.

The pattern of results for error rates can also be seen in Figure 4. As with reaction times, no main effect nor interaction involving group emerged, so the display is collapsed across groups. Error rate changed as a function of stepsize, $F(3, 87) = 12.8, p < .001$, but not of startpoint. There was a Startpoint × Stepsize interaction, $F(6, 174) = 7.2, p < .001$, which appeared to show a decrease of errors at Stepsize 4. Post hoc inspection of the trials revealed that the stepsize categories may have differed in the average pitch difference between the two lyrics. In many experimental paradigms, wider separation of the stimuli on some dimension allows for easier comparisons or discrimination. Consequently, a secondary analysis in terms of pitch separation was carried out to see if the easier trials were those where the to-be-compared notes were maximally different in pitch.

*Pitch separation.* The average pitch separations for Stepsizes 2, 4, 6, and 8 were 4.17, 5.74, 5.29, and 5.35 semitones (musical half steps), respectively. The largest pitch separation appeared to be coincident with the lowest error rate. However, a one-way ANOVA failed to confirm any differences in pitch separation.

Nevertheless, perhaps the pattern of errors and/or reaction times could be associated with different pitch separations, regardless of the original uncontrolled distribution of pitch separation over stepsizes. Accordingly, all the trials were regrouped into five pitch-separation categories to allow for a reasonable number of observations in each group: 1 or 2 semitones; 3 or 4 semitones; 5 (there were no trials with 6 semitones' separation) semitones; 7 or 8 semitones; and 9 or 10 semitones. These categories were then considered to be five levels of one factor (pitch separation) for an ANOVA. A second factor was instruction group: imagery and nonimagery.

Both errors and reaction times were analyzed. For neither was the group factor significant. For both, the pitch separation factor was significant: $F(4, 116) = 7.0, p < .001$ for reaction times and $F(4, 116) = 6.8, p < .001$ for errors. Neither main effect was modified by a Pitch Separation × Group interaction. Means are shown in Table 1.

Newman-Keuls' tests on the combined groups revealed that the reaction time for Category 3,4 was longer than for any other except Category 7,8; Category 7,8 also differed from Category 9,10 ($p < .05$). Thus, although the shortest reaction time did occur in the widest pitch separation category, the

Table 1

*Mean Errors and Reaction Times (RTs) for Each Pitch Separation Category in Experiments 3 and 4 Collapsed Across Groups*

| Pitch separation | Errors (%) | | RT (ms) | |
|---|---|---|---|---|
| | Exp. 3 | Exp. 4 | Exp. 3 | Exp. 4 |
| 1,2 | $26_a$ | $31_c$ | $6271_{e,g}$ | $5910_h$ |
| 3,4 | $25_a$ | $21_d$ | $7583_f$ | $5855_h$ |
| 5[a] | $25_a$ | $15_d$ | $6537_{e,g}$ | $5831_h$ |
| 7,8 | $19_a$ | $21_d$ | $6976_{f,g}$ | $5400_h$ |
| 9,10 | $11_b$ | - | $6121_e$ | — |

*Note.* Means having the same subscript are not significantly different at $p < .05$ using a Newman-Keuls' test. Pitch separation is measured in semitones. Exp. = Experiment.

[a] No lyrics were separated by six semitones in Experiment 3.

remaining ordering does not suggest that subjects are generally faster with wider pitch separations.

Error rates did follow a logical sequence. As seen in Table 1, error rates increase with decreasing pitch separation. However, the Newman-Keuls' test only confirmed that Category 9,10 was more accurate than the other categories, which did not differ among themselves ($p < .05$).

Analyses of covariance (ANCOVAS) were also performed for each dependent measure, with pitch separation as the covariate, but the ANCOVAS did not allow any different interpretations of the data.

*Musical training.* Experiment 3 was a much harder and arguably more musically demanding task than the previous scanning task. Thus, musicians may have had advantages over nonmusicians in performing the task. To investigate this, the subjects were divided into musicians and nonmusicians by a median split of their years of musical training (*mdn* = 5.5 years). Musicianship was added as a factor in an ANOVA of reaction times. Overall, musicians did not perform differently than did nonmusicians, nor did this factor interact with instruction group, stepsize, startpoint, or any combination thereof.

Because the pattern of errors as a function of stepsize and startpoint was not particularly informative in the original analysis, a simple comparison was made between the error rates of the two groups. Musicians made fewer errors ($M$ = 19.2%) than did nonmusicians ($M$ = 29.0%), $t(29) = 3.03$, $p < .01$. In summary, although musicians made fewer errors than did nonmusicians, they were no faster on the trials that both groups answered correctly.

### *Discussion*

Changing the task so that a musical judgment was required considerably increased the difficulty level but did not substantially change the reaction-time patterns from the previous experiments. Reaction times increased with greater distance between the to-be-judged notes and also increased with the starting point of the first lyric, which suggested that intermediate notes were being processed.

The reaction-time pattern was identical for imagery and nonimagery instructed groups. This implies that the difficulty or some other aspect of the task strongly encouraged participants to use an analogue representation without being told to do so. The ideal control group for this comparison is a logical impossibility—that is, a group that is told *not* to use an imagery representation (which would be equivalent to saying "don't think of an elephant"). Perhaps a control group could be instructed specifically to use a nonimagery strategy. Nevertheless, the current results show that subjects spontaneously adopt a strategy similar to that of imagery-instructed people.

The apparent inevitability of this strategy is even more interesting considering that intervening tones can have a large disruptive effect in actual pitch comparison tasks (Deutsch, 1970, 1972). Blackburn (cited in Dowling & Harwood, 1986) found similar effects when the target tone was imagined. Efficient performance on this task would seem to include bypassing the intervening tones in memory.

Musicians had some advantage over nonmusicians in task accuracy. We might expect the same result, however, had both groups been given an actual pitch-comparison task. Qualitatively, musicians did not differ from nonmusicians in reaction-time patterns. Thus, we may conclude that training (or endowment) does not increase or lessen the need for processing the intermediate tones. Musicians might show a different pattern if unfamiliar and/or complex tunes were used as stimuli. However, the goal of the current research is to investigate the representation of music that is well-established in memory.

One procedural change in the experiment was shortening the presentation time of the first lyric to try to clarify the startpoint effect at the smaller stepsizes. In fact, startpoint differences were again minimal at Stepsize 2. Lyrics only two beats distant from one another may simply be so associated as to eliminate the need for processing the one intervening note.

The increase in errors with decreasing pitch separation of the two tones deserves further study. A pitch-separation effect for actually presented tones was shown by Dewar, Cuddy, and Mewhort (1977). The effect was weak here, perhaps because of the post hoc nature of the analysis. The next study explicitly varied different pitch separations while keeping other factors constant.

## Experiment 4

This experiment probed explicitly for the pitch-separation effect intimated by Experiment 3. A new variable of pitch separation was added to the design. Reaction time and/or errors were expected to decrease with increasing distance in frequency between the notes referred to in the trial.

### *Method*

*Subjects.* Twenty Bucknell University undergraduates participated in this experiment to earn course credit.

*Materials.* In order to accommodate the new variable of pitch separation, the pool of stimulus songs was expanded to 15 familiar songs (as assessed by informal polling of undergraduates). In order to keep the stimulus set at a reasonable size (we disqualified subjects if they claimed ignorance of even one song), only Stepsizes 4, 6, and 8 were used. In addition, pilot work suggested that startpoint would not interact in interesting ways with pitch separation. Eliminating startpoint as a variable again permitted a smaller stimulus set than would otherwise be needed.

Specifically, the design tested three stepsizes (4, 6, or 8 beats) and four categories of pitch separation: 1 or 2 semitones (musical half-steps); 3 or 4 semitones; 5 or 6 semitones; and 7 or 8 semitones. The four pitch-separation categories and three stepsizes resulted in 12 trial types. Three observations for each trial type were generated by including one trial each with Startpoints 1, 3, and 5 for the first lyric whenever possible (it proved impossible to adhere to this scheme strictly), which resulted in a total of 36 trials. Because Experiment 3 suggested that error rates might be high in this task, the number of observations was further increased by repeating each of the trials once, for a total of six observations per cell. Finally, 15 unscored trials were added to prevent subjects from basing their answers on their memory of the first presentation of the trial, making a grand total of 87 trials.

Each song from the stimulus pool was used from two to five times, and for half the trials, the second pitch was higher than the first; for half it was lower.

*Procedure.* The instructions and task were executed as in Experiment 3. Because imagery instructions did not interact with any factors in Experiment 3, only a nonimagery instruction group was used. At the end of the session, subjects were asked how they thought they were accomplishing the task.

### *Results*

Data from two subjects with error rates over 40% were discarded. In the postexperimental interview, these two subjects admitted to guessing most of the time, whereas all the rest used wording indicative of an imagery experiencing ("singing the tune inside my head"). The mean error rate of the remaining subjects was 22%. Separate two-way ANOVAS, with stepsize and pitch separation as within factors, were conducted on the mean correct reaction time and error rate for each cell.

Considering reaction time first, time to answer did increase as a function of stepsize, $F(2, 38) = 31.7$, $p < .001$. Means were 4,787, 5,974, and 6,486 ms for Stepsizes 4, 6, and 8, respectively. Means for the pitch-separation categories were ordered in the predicted fashion, with reaction times of 5,910, 5,855, 5,831, and 5,400 ms for the smallest to the largest pitch separations, respectively (see Table 1). However, these were not significantly different, $p > .05$. A significant Stepsize × Pitch Separation interaction, $F(6, 114) = 4.98$, $p < .001$, was not clearly interpretable.

Mean error rates for the smallest to the largest pitch separation categories were 31%, 21%, 15%, and 21%, respectively, which differed significantly, $F(3, 57) = 9.08$, $p < .001$. A Neuman-Keuls test revealed that the latter three means were indistinguishable, but that people were erring more on pitch separations of 1 or 2 semitones. Error rate overall did not change as a function of stepsize, $F(2, 38) = 3.21$, $p = .05$. However, a significant Stepsize × Pitch Separation interaction, $F(6, 114) = 3.22$, $p < .01$, was due to an increase in error rate as a function of stepsize only for Pitch Separation Category 1 or 2.

### *Discussion*

The increase of reaction time with increasing stepsize was again replicated here with a completely different set of trials than had been used in the previous experiments. The predicted effect of pitch separation was modestly supported, in that errors were most frequent for the smallest pitch-separation category. The mean reaction times were ordered in the predicted direction, but there were no reliable differences among them.

Although every effort was made to balance startpoint and song over trial types, we could not do so completely, especially for song. In the preliminary study to Experiment 1 already mentioned, where song was balanced over startpoint and stepsize, a main effect and an interaction involving song were found ("Star Spangled Banner" produced a faster overall reaction time but slower scanning rate than did the other two songs). Thus, we may expect that particular songs will have an effect on at least some variables. On the other hand, the pitch-separation variable simply did not allow for the desired balancing. Recall that the completely balanced stimulus set in Experiments 1 and 2 confined us to testing three songs. In working with preexperimentally known material, we may be forced into this tradeoff between balancing and generalizability.

## General Discussion

To summarize, four experiments show that requiring subjects to attend to two points in a familiar song probably involves processing of intermediate notes. This was shown by the increasing, and usually linear, function relating reaction time to stepsize. If parts of the song could have been randomly accessed, no systematic relation of reaction time to stepsize should have occurred.

The increase of reaction time with startpoint is particularly interesting because not even imagery subjects were instructed to run through an early section of the song before answering a question about a later part. Although the startpoint effect without imagery instructions is somewhat equivocal in the scanning task, it is quite clear in the pitch-comparison task. Had the titles presented in each trial always been the first lyric of the song, a startpoint effect would not have been surprising. That is, subjects would have thereby been encouraged to think about each song from its beginning. However, titles were never first lyrics for just that reason. Titles were lyrics taken from later in the song ("Rainbow" for "Somewhere Over the Rainbow") or did not appear as song lyrics at all ("Battle" for "Battle Hymn of the Republic").

Taken together, the results of these experiments support the claim that auditory imagery is not only a strong subjective experience (experimental subjects never objected to carrying out auditory imagery tasks) but is also at least partly amenable to quantification. People indeed behave as if they were running songs through their heads. This should not be taken as a claim that doing so is always obligatory. Music is not a continuous stream of notes. It has sections both large (movements) and small (phrases). To whatever extent these sections can be symbolically coded, we might expect that people can avoid processing them in analogue fashion. Thus, it is not reasonable to assume that we need to mentally play the first three movements of Beethoven's *Ninth Symphony* in order to think about the final "Ode to Joy" movement. However, at least within small musical units (the stimuli in these experiments only rarely exceeded a phrase), analogue processing is apparently the strategy of choice.

The current work also has little to say about the comparison of mental scanning to some perceptual version of the tasks, as has been the practice in the visual domain (Kosslyn, 1978). Research underway in our laboratory looks at other characteristics of mental representation of tunes, such as actual (imagined) tempo and pitch. We should not be surprised, for instance, if we can imagine tunes faster than we usually hear them. However, there may be limits on how fast or slow the songs can be imagined before the coherence of the stream breaks down (e.g., Bregman & Campbell, 1971). The current experiments provide a start in assessing the similarities of auditory imagination to all the extensively studied phenomena in auditory perception.

## References

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

Bregman, A. S., & Campbell, J. (1971). Primary auditory stream segregation and perception of order in rapid sequences of tones. *Journal of Experimental Psychology, 89,* 244–249.

Chew, S. L., Larkey, L. S., Soli, S. D., Blount, J., & Jenkins, J. J. (1982). The abstraction of musical ideas. *Memory & Cognition, 10,* 413–423.

Cuddy, L. L., & Cohen, A. J. (1976). Recognition of transposed melodic sequences. *Quarterly Journal of Experimental Psychology, 28,* 255–270.

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

Deutsch, D. (1972). Mapping of interactions in pitch memory store. *Science, 175,* 1020–1022.

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. (1972). Recognition of melodic transformations: Inversion, retrograde, and retrograde inversion. *Perception & Psychophysics, 5,* 417–421.

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

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

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

Finke, R. A. (1979). The functional equivalence of mental images and errors of movement. *Cognitive Psychology, 11,* 235–264.

Finke, R. A., & Kosslyn, S. M. (1980). Mental imagery acuity in the peripheral visual field. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 126–139.

Friedman, W. J. (1983). Image and verbal processes in reasoning about the months of the year. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 126–132.

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

Kosslyn, S. M. (1978). Measuring the visual angle of the mind's eye. *Cognitive Psychology, 10,* 356–389.

Kosslyn, S. M., Ball, T. M., & Reiser, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 47–60.

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

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

Received April 6, 1987
Revision received July 31, 1987
Accepted August 6, 1987 ■

---

## Call for Nominations for Editor of *Behavioral Neuroscience*

The Publications and Communications Board has opened nominations for the editorship of *Behavioral Neuroscience* for the years 1990–1995. Richard F. Thompson is the incumbent editor. Candidates must be members of APA and should be available to start receiving manuscripts in early 1989 to prepare for issues published in 1990. Please note that the P&C Board encourages more participation by women and ethnic minority men and women in the publication process and would particularly welcome such nominees. Submit nominations no later than August 1, 1988 to

Martha Storandt
Department of Psychology
Washington University
St. Louis, Missouri 63130.

Other members of the search committee are Byron Campbell, Mortimer Mishkin, Mark Rosenzweig, and Shepard Siegel.