The speed of pitch resolution in a musical context

R. Ranvaud
Departamento de Fisiologia e Biofisíca, Instituto de Ciências Biomédicas, Universidade de São Paulo, Avenida Lineu Prestes 1524, São Paulo, SP 05508-900, Brazil

W. F. Thompson[a]
Department of Psychology, Atkinson College, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

L. Silveira-Moriyama
Departamento de Fisiologia e Biofisíca, Instituto de Ciências Biomédicas, Universidade de São Paulo, Avenida Lineu Prestes 1524, São Paulo, SP 05508-900, Brazil

L.-L. Balkwill
Department of Psychology, Atkinson College, York University, 4700 Keele Street, Toronto, Ontario M3J 1P3, Canada

(Received 19 April 2000; revised 5 January 2001; accepted 5 March 2001)

In five experiments, we investigated the speed of pitch resolution in a musical context. In experiments 1–3, listeners were presented an incomplete scale (doh, re, mi, fa, sol, la, ti) and then a probe tone. Listeners were instructed to make a rapid key-press response to probe tones that were relatively proximal in pitch to the last note of the scale (valid trials), and to ignore other probe tones (invalid trials). Reaction times were slower if the pitch of the probe tone was dissonant with the expected pitch (i.e., the completion of the scale, or doh) or if the probe tone was nondiatonic to the key implied by the scale. In experiments 4 and 5, listeners were presented a two-octave incomplete arpeggio, and then a probe tone. In this case, listeners were asked to make a rapid key-press response to probe tones that were relatively distant in pitch from the last note of the arpeggio. Under these conditions, registral direction and pitch proximity were the dominant influences on reaction time. Results are discussed in view of research on auditory attention and models of musical pitch.
[DOI: 10.1121/1.1367254]

PACS numbers: 43.75.Cd [RDA]

I. INTRODUCTION

Attention is of paramount importance in guiding the behavior of human beings. In recent years, empirical research has focused on attention in the visual system. Visual attention must rapidly and continuously shift between different spatial locations in response to environmentally significant events. Shifts in visual attention are not dependent on eye movements, but they occur in real time (Posner, 1985). In particular, there is evidence that the time taken to shift visual attention from one spatial location to another is proportional to the angular separation of these two points (Tsal, 1983; Downing and Pinker, 1985). This finding suggests that visual attention moves within an analog representation, and that spatial relations and distance information are directly represented by the structure of that representation.

Similar principles apply to auditory attention. Rhodes (1987) found that the time taken to shift auditory attention in three dimensional space was a linearly increasing function of the angular distance moved, at least for distances up to 90 degrees. This finding suggests that both visual and auditory spatial information are represented analogically, and that similar constraints apply to movements of visual and auditory attention within these representations. Another location-based characteristic known in visual attention, inhibition of return (Posner and Cohen, 1984), was recently demonstrated for auditory attention by Mondor *et al.* (1998) and by Spence and Driver (1998), further illustrating parallel features in visual and auditory attentional shifts.

In this investigation, we further examined shifts in auditory attention. However, rather than examine shifts from one spatial location to another, we examined shifts in attention between different pitches. Our investigation was motivated by two theoretical considerations. First, the analysis of pitch by the auditory system is, in an important sense, analogous to spatial analysis by the visual system. The analysis of visual space occurs early in processing. Responses at different parts of the retina are mapped retinotopically onto the primary visual cortex. In contrast, the analysis of spatial information may occur quite late in auditory processing, only after information from the two ears is combined. Rather than spatial location, it is frequency that is analyzed at early stages of auditory processing. Different acoustic frequencies create vibrations at different segments of the basilar membrane, leading to a tonotopic organization of pitch in the primary auditory cortex.

A second consideration was that there is an extensive body of research concerned with musical pitch resulting in several models describing psychological relationships between pitches as they occur in musical contexts (e.g., Deut-

---

[a] Author to whom correspondence should be addressed. Electronic mail: billt@yorku.ca

sch, 1982; Krumhansl, 1990). These models, and the empirical work supporting them, provided a solid foundation upon which to make predictions about the speed with which pitches can be resolved in a musical context. In particular, when attention is focused on one pitch (the attended pitch), the speed with which a new pitch is resolved may be influenced by the psychological distance between the attended pitch and the new pitch.

One early model of pitch—the helical model—considers two dimensions: pitch height, which is proportional to the logarithm of the fundamental frequency of a complex tone, and pitch chroma, which treats as equivalent those tones whose fundamental frequencies are separated by an octave. More recent models of pitch have acknowledged the importance of other dimensions of pitch relationships, such as pitch contour, and tonal function (for reviews, see Dowling and Harwood, 1986; Krumhansl, 1990).

It is widely agreed that proximity in pitch height—or ''pitch proximity''—has a significant influence on perceived relationships between pitches. Tones that are proximal in pitch height are perceived to be highly related, listeners expect melodic patterns in which sequential tones are proximal in pitch height, and proximal tones tend to form perceptual groups when played consecutively. Pitch proximity may have a greater influence than spatial location on auditory grouping. Deutsch (1975) presented musical materials to listeners in a manner such that location information suggested one grouping of sounds, while pitch proximity suggested a conflicting grouping of sounds. She presented two major scales simultaneously to listeners, one ascending and the other descending. Successive tones in each of the two scales were presented alternately to the left and right ears, such that when a tone from the ascending scale was presented to the right ear, a tone from the descending scale was presented to the left ear. Listeners did not accurately perceive the location of pitches. Rather, listeners grouped the tones in a way that preserved pitch proximity. This effect, known as the ''scale illusion,'' indicates that pitch proximity may override spatial location in auditory grouping.

Musical context may introduce a number of other properties that may influence perceived relationships between pitches, such as scale membership, contour, harmony, and tonality. Sensitivity to these properties arises through exposure to music organized along discrete sets of pitches known as scales. Research suggests that listeners readily abstract underlying scales (Cohen, 1991), and that perceived relationships among pitches in a musical context are influenced by whether or not those pitches are members of the scale (Krumhansl, 1979). Tones whose pitches are members of the scale are called diatonic tones. Tones whose pitches are not members of the scale are called nondiatonic tones. The most common scale in Western music is the major scale, which has seven pitches per octave (doh, re, mi, fa, sol, la, ti). The first note of the scale is referred to as the tonic, and serves as a point of perceived stability in music. Other notes vary in their perceived stability, resulting in a *tonal hierarchy* (Krumhansl and Kessler, 1982). Sensitivity to the scale and the tonal hierarchy will be collectively referred to as *tonality*.

Melodic contour, which we refer to as *registral direction* (the up/down pattern of notes in a melody), also may influence the speed of pitch resolution. If a melody involves part of an ascending scale, for example, listeners may expect that ascending motion to continue. The latter prediction is based on the Gestalt principle of ''good continuation,'' which has been examined extensively in studies of music (e.g., Deutsch, 1999; Narmour, 1990). Scale and contour are thought to be among the most significant influences on memory for melodies (Dowling, 1978).

Harmony also may influence perceived pitch relationships. The perception of harmony is influenced by sensory dissonance, but is also shaped by long term knowledge of music (Bharucha and Stoeckig, 1986, 1987; Tekman and Bharucha, 1992). Only the sensory aspect of harmony was addressed in this investigation. When two or more tones are sounded simultaneously, their combination may result in physical beating among partials. An excess of such beating patterns is perceived as dissonance; an absence of such beating is perceived as consonance (Plomp and Levelt, 1965).[1] Pitches that are consonant with each other may be perceived as psychologically related, even when they are presented in sequence (Krumhansl, 1995; Thompson and Stainton, 1998). The latter effect may reflect sensitivity to the degree of spectral overlap among such tones, or result from an evaluation of sensory interactions that arise when the memory trace of one tone is combined with physically occurring properties of a subsequent tone.

To summarize, psychological relationships between pitches are influenced by several factors including pitch proximity, tonality, registral direction, and dissonance. In this study, we examined whether these factors affect the time it takes to shift auditory attention between pitches. Listeners were induced to focus on one pitch, and were then presented a probe tone. They were asked to respond rapidly to certain probe tones (''valid'' trials) while ignoring others (''invalid'' trials). We focused on four factors that may influence psychological pitch distance, and, therefore, the time required to shift auditory attention through pitch space. Specifically, we predicted that the time taken to resolve the pitch of a probe tone would be more rapid when

(1) the pitch of the probe tone is proximal to the expected pitch (proximity);
(2) the probe tone is not dissonant with the expected tone (dissonance);
(3) the probe tone is stable within the key defined by the preceding context (tonality); and
(4) the probe tone continues the registral direction suggested by the preceding context (registral direction).

There is precedent for using reaction time to evaluate pitch perception in a musical context. Janata and Reisberg (1988) examined the reaction time of listeners who were instructed to indicate as quickly as possible whether probe tones were diatonic to the scale defined by a preceding musical context. Reaction times were significantly faster for more stable tones within the key. The authors argued that differences in reaction time reflected listeners' internal representation of the tonal hierarchy. One question addressed in our investigation is whether such an influence is observed when responses to pitch are not contingent on an explicit (conscious) assessment of tonality.

We report the results of five experiments. The design of all experiments was adapted from studies of visual attention, in which participants responded rapidly to visual stimuli presented at varying distances from an initial fixation point (Umiltá, 1988; Hikosaka *et al.*, 1996; Egeth and Yantis, 1997; Steinman and Steinman, 1998). In experiments 1–3, listeners were presented an incomplete major scale, which directed their attention to the note that would complete that scale. Reaction times were obtained for different probe tones. In experiments 4 and 5, an incomplete arpeggio was used to direct the attention of listeners to a certain note. Responses were obtained for the same set of probe tones used in experiments 2 and 3. The results for scale and arpeggio contexts were compared to explore how psychological distances between pitches are affected by musical context.

II. EXPERIMENT 1

In experiment 1, we examined the speed with which listeners can resolve pitches following an incomplete major scale. Participants were first exposed to an incomplete major scale (doh, re, mi, fa, sol, la, ti) in order to direct their attention to a specific pitch (the completion of the scale, or ''doh''). A major scale was adopted because it is well established that this context generates strong expectations for the pitch that completes the scale, i.e., doh (Krumhansl and Shepard, 1979; Krumhansl, 1990). The incomplete scale was followed by one of eight probe tones, and participants were instructed to make a key-press response as soon as they detected the pitch of the probe tone.

To ensure that participants responded to the pitch of the probe tone, rather than to other dimensions of the probe tone (e.g., onset), a go–no-go procedure was adopted in which the response criterion was specifically determined by pitch. Listeners were instructed that if the probe tone was less than three semitones from the tone that would complete the scale, they should respond as quickly as possible (valid probe tones), but if the probe tone was seven or more semitones from the tone that would complete the scale, they should not respond (invalid probe tones). Thus, the decision to respond was contingent on an analysis of pitch, and reaction times should be affected by the psychological distance between the expected pitch (next tone on the major scale) and the pitch of the (valid) probe tone.

A. Method

1. Participants

Fourteen participants (nine males, five females) with at least three years of musical training were recruited for experiment 1 among the staff and student community at the Institute of Biomedical Sciences at the University of São Paulo. All participants were volunteers, their ages ranged from 19 to 50 (mean=26.4, s.d.=9), and all reported normal hearing.

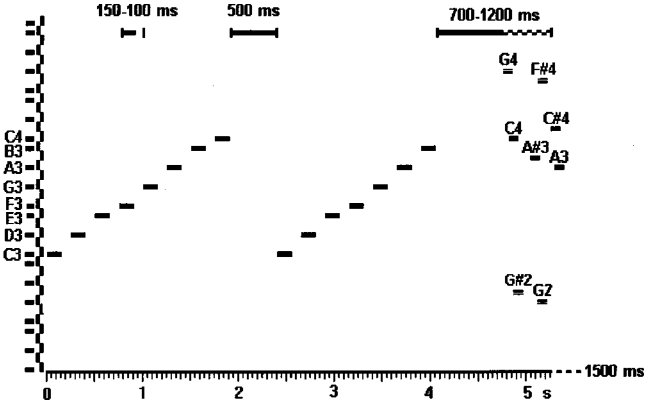

FIG. 1. The experimental design of experiment 1. The vertical axis represents frequencies on a logarithmic scale; the horizontal axis represents time (s). The letters at the right indicate the probe tones for a C major scale beginning at C3. Valid probe tones are shown as a solid line and invalid probe tones are shown as a double line. Horizontal bars at the top indicate the duration of tones and intervals between tones.

2. Materials and tone construction

The tones to which participants were exposed were generated using MEL Professional V2.0 software acquired from Psychology Software Tools (PST, Inc., Pittsburgh; see Schneider, 1988), on an IBM-PC compatible Pentium computer. Tones were sine waves, and were each 150 ms in length including 30 ms of onset and 30 ms of offset. Participants heard the stimuli through high-quality speakers at a comfortable listening level (approximately 70 dB) in a sound-attenuated booth with reduced lighting. Participants were tested individually, and responded by pressing a button on a joystick with their thumb, directly entering the data into the computer.

3. Conditions and procedure

Details of the experimental design are shown, to scale, in Fig. 1. To avoid habituation, the first pitch of the scale was one of four pitches, determined randomly for each trial: C3 (523 Hz), A2 (440 Hz), G2 (391 Hz) or F2 (349 Hz). There were four valid probe tones for the C3 scale: C4, C♯4, A3 and A♯3. There were four invalid probe tones for the C3 scale: F♯4, G4, G2 and G♯2. Valid and invalid probe tones of the A2, F2, and G2 scales were transpositions of the example given.

After each response, participants were provided feedback on their performance, as shown in Table I. If errors were committed in the course of a block, new tests were presented to the participants at the end of the block, such that the number of correct responses was the same for all conditions. Altogether, each participant had to provide correct re-

TABLE I. Feedback (in italics) given to participants in experiment 1, according to their response in the different conditions.

Reaction time	Valid probe tone	Invalid probe tone
RT<130	*Anticipated*	*Error & anticipated*
130<RT<1000	*RT displayed*	*Error*
RT>1000	*Slow*	*Error*
No response	*Slow*	*Correct*

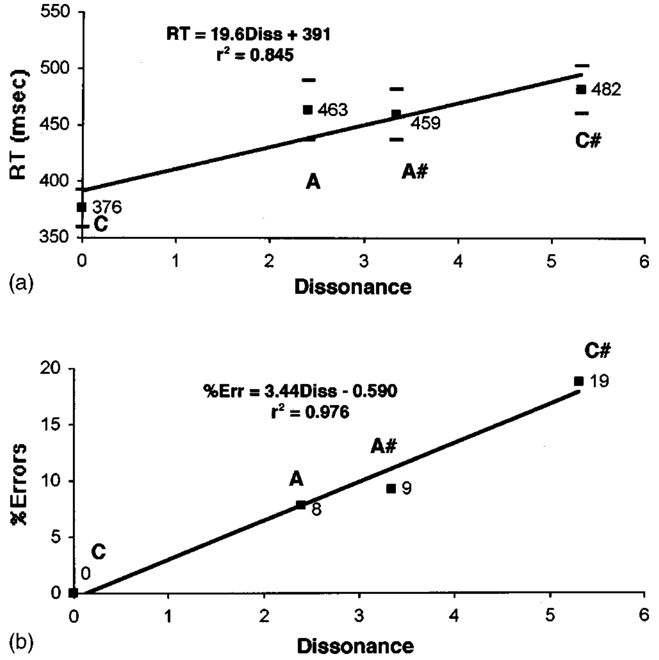

FIG. 2. (a) Mean reaction times (RT) for the four valid probe tones of experiment 1, plotted as a function of dissonance relative to the expected probe tone. According to Plomp and Levelt (1965), the values of dissonance are as follows: C4=0, A3=2.40, A♯3=3.34, C♯4=5.32. The bars indicate standard errors. Regression equations and squared correlations are shown. (b) Percentage of errors committed by subjects in response to valid probe tones in experiment 1, plotted as a function of dissonance relative to the expected probe tone. Regression equations and squared correlations are shown.

sponses to 192 tests (16 blocks of 12 tests each), of which half were valid, and half were invalid. The results presented were thus based on 96 reaction times per participant, 24 for each one of the four valid probe tones. The number of errors committed was also recorded.

Participants were tested on three separate experimental sessions. On the first day, they were given instructions for the task, and participated in a training session in which they were informed ahead of time whether the probe tone would be valid or invalid. After completing the training session, all participants took part in a practice experiment, the reaction times of which were discarded. On the second and third days, participants went through another practice experiment (the data being again discarded) before participating in the experiment (8 random blocks of 12 tests each, 6 valid and 6 invalid, plus repeats to make up for mistaken responses, with pauses after the third block and the sixth block to avoid excessive fatigue).

B. Results and discussion

The mean reaction times for the four valid probe tones are shown in Fig. 2(a). A one-way analysis of variance showed a significant effect of probe tone, $F(3,39)=28.41$, $p<0.001$). Planned comparisons were conducted to assess the predictions. First, reaction times for the expected probe tone, C4 (mean=376 ms), were significantly faster than reaction times across other valid probe tones (mean=468 ms), $F(1,39)=81.35$, $p<0.001$. This result suggests that the scale context indeed focused auditory attention on the pitch of C4, so that no shift in auditory attention was required to resolve that pitch, leading to shortest reaction times for this probe tone.

A second planned comparison revealed that reaction times for the probe tone C♯4 (mean=482 ms) were significantly longer than reaction times across other probe tones (mean=433 ms), $F(1,39)=23.51$, $p<0.001$. This finding is consistent with predictions based on dissonance and tonality, as this pitch is perceived as dissonant with the expected pitch (C4) and it is not diatonic to the scale. Yet this finding is inconsistent with predictions based on proximity and registral direction: the pitch of C♯4 is only one semitone away from the expected pitch and is in the expected registral direction. According to the analysis of Plomp and Levelt (1965), C♯4 has a greater degree of perceived dissonance with the expected pitch, C4, than any other probe tone. The quantitative values of dissonance obtained by these authors were used to plot values of the abscissa in Fig. 2(a).

The remaining two probe tones, A3 and A♯3, showed intermediate reaction times, which were statistically indistinguishable (463 and 459 ms, respectively). These two probe tones have similar levels of dissonance with the expected probe tone at C4. The tones also have similar tonal hierarchy values as quantified by Krumhansl and Kessler (1982), even though only one of these probe tones (A3) is diatonic to the established key. As shown in Table II, regressions of reaction times as a function of dissonance [shown in Fig. 2(a)] and tonality both show a high correlation coefficient ($r^2=0.845$ and 0.895 respectively, $p<0.05$).

In addition to analyzing reaction times, we also examined the proportion of incorrect responses for each valid probe tone. This analysis was performed to assess whether psychological distance influenced the accuracy with which

TABLE II. Squared correlations (r^2) between mean reaction times and % errors as a function of dissonance, pitch proximity, and tonality. An asterisk indicates statistical significance at the 0.05 level.

	Reaction time			% Errors		
Experiment No.	Dissonance	Pitch proximity	Tonality	Dissonance	Pitch proximity	Tonality
1	**0.845***	0.429	**0.895***	**0.976***	0.053	**0.747**
2	**0.949***	0.101	**0.729**	**0.886***	0.000	**0.576**
3	**0.710**	0.055	**0.469**	**0.989***	0.067	**0.789**
4	0.331	**0.667**	0.005	0.016	**0.980***	0.161
5	0.001	**0.968***	0.276	0.009	**0.997***	0.780

listeners classified probe tones. The distribution of incorrect responses among the four valid probe tones was consistent with the reaction time data, as shown in Fig. 2(b). Errors were highly nonrandom (chi squared $(3)=66.9$, 3 $p<0.001$). The pattern of errors suggests that processing the expected tone, C4, was a much easier task than processing tones at A3 or A♯3, and C♯4 (in the order). As shown in Table II, the percentage of correct responses was strongly related to dissonance ($r^2=0.976$, $p<0.01$).

III. EXPERIMENTS 2 AND 3

Experiment 1 indicated that listeners responded more quickly to a probe tone with an expected pitch than to a probe tone with an unexpected pitch. In addition, reaction times increased linearly with dissonance relative to the expected probe tone, and decreased with the tonal stability of the probe tone relative to the key established by the preceding scale. The slowest reaction time, observed for the probe tone at C♯4, was expected on the basis that it is psychologically distant from the expected note of C4. Although C4 and C♯4 are proximal in pitch (separated by only one semitone), they are highly dissonant with each other, and psychologically distant from each other in a C major context (Krumhansl, 1990). Overall, dissonance and tonality were strong predictors of the reaction time.

Experiments 2 and 3 were conducted for three reasons. First, listeners in experiment 1 were extensively trained to respond to valid probe tones within 1 s, and several responses were measured for each probe tone. It is possible that such extensive training may have encouraged listeners to develop explicit strategies of responding that are not typically invoked when shifting auditory attention between pitches under normal conditions of music listening, and which might mask the automatic pitch encoding processes under investigation. Experiments 2 and 3 were designed to verify whether the results of experiment 1 would be obtained without elaborate training and exposure to the experimental protocol. Second, the tones used in experiment 1 were sine waves, but pitch perception in most musical contexts involves more familiar timbres. We considered the possibility that pitch resolution for sine tones may be influenced by somewhat different factors than pitch resolution for more familiar timbres. Thus, the stimuli used in experiments 2 and 3 consisted of sampled piano tones.

Third, it is possible that the large reaction time observed for the probe tone at C♯4 was partially related to the asymmetric distribution of valid and invalid probe tones (see method section, experiment 1). Valid probe tones were more distant in pitch from invalid probe tones in the lower register (between 13 and 18 semitones) than from invalid probe tones in the upper register (between 5 and 10 semitones). This asymmetry may have made it more difficult to distinguish between higher pitched valid and invalid probe tones than between lower pitched valid and invalid probe tones. Thus, the large reaction time at C♯4 may have resulted, in part, from a difficulty in discriminating this pitch from upper invalid pitches. Experiments 2 and 3 eliminated this asymmetry.

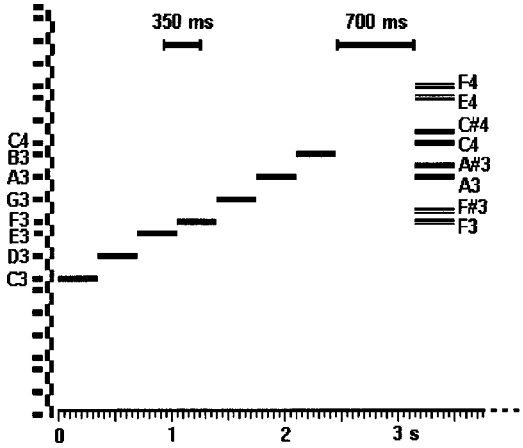

FIG. 3. The experimental design of experiments 2 and 3.

We noted that reaction times for the first two responses to each probe tone in experiment 1 illustrated the same trend as the mean reaction time for the full set of 24 responses (386, 448, 474 and 490 ms versus 376, 463, 460 and 482 ms, respectively, for probe tones C4, A3, A♯3, and C♯4). Thus, experiments 2 and 3 limited the number of responses to two.

A. Method

1. Participants

Twelve women and eight men from the York University community, ranging in age from 19 to 31 (mean=24.3) participated in experiment 2. Ten received undergraduate course credit for their participation, and ten were volunteers. Twelve women and three men, aged between 18 and 54 (mean 23.1) participated in experiment 3. Twelve received undergraduate credit for their participation and three were volunteers. The number of years of formal musical training of participants ranged from zero to greater than ten (mean =6.8 years) in experiment 2 and from zero to ten (mean=3.4 years) in experiment 3. All participants reported normal hearing.

2. Materials and tone construction

Tones were taken from the internal timbre set of a Power Macintosh computer. The timbre of all tones was the sampled piano sound of the Roland piano (for an acoustic description, see Fletcher and Rossing, 1991), with fundamental frequencies set to equal temperament tuning. Stimuli were presented to listeners in a sound attenuated booth through Sennheiser HD-480 headphones. Responses were entered directly onto a computer.

3. Conditions and procedure

The stimuli used for experiments 2 and 3 were identical and are illustrated in Fig. 3. Listeners were presented an incomplete C major scale, followed by a pause of 700 ms, and then a probe tone in each trial. The seven notes of the incomplete scale, C3, D3, E3, F3, G3, A3, B3, were each 350 ms in duration, as was the duration of the probe tone. There were eight probe tones.

Participants were instructed to make a rapid key-press response if the probe tone was two or fewer semitones from the last tone of the incomplete scale (ti); otherwise they should make no response, and wait for the next trial. The four valid probe tones were A3, A♯3, C4, and C♯4. For both experiments 2 and 3, the maximum reaction time acceptable was set to 2000 ms. The four invalid probe tones were F3, F♯3, E4, and F4. Upper and lower probe tones were symmetric in pitch distance from the reference tone: valid probe tones were exactly 1 or 2 semitones above or below the reference tone (ti), and invalid probe tones were 5 or 6 semitones above or below the reference tone.

In experiment 2 each probe tone was presented just twice, regardless of whether correct responses were obtained, thus minimizing participant exposure to the experimental procedure. In experiment 3, each invalid probe tone was presented twice, but each valid probe tone was presented as many times as needed to obtain two valid key-press responses (i.e., responses under 2000 ms). Thus, if no response to a valid probe tone was made by 2000 ms, that trial was repeated at a randomly determined position later in the experimental session. The order in which probe tones were presented was randomized independently for each participant.

Experiments 2 and 3 also differed in the following ways. In experiment 2, listeners were given minimal training, with an emphasis on accuracy, and proceeded to the experiment as soon as they succeeded in correctly classifying (unspeeded) 16 probe tones in a row (i.e., each possible probe tone, twice). Once the experiment began, there were no trial replacements for mistaken responses, and therefore the number of correct responses was not always the same for all probe tones or participants.

In experiment 3, participants were trained with one set of practice trials involving all valid probe tones, and another set of practice trials involving all invalid probe tones. Participants were given the opportunity to switch back and forth between these two practice sets until they felt comfortable with the distinction between valid and invalid probe tones. Emphasis was on speed, and participants were given feedback on their performance during the practice trials (reaction time in ms).

B. Results and discussion

The mean reaction times obtained in experiments 2 and 3, shown in Fig. 4(a), illustrate the same trend observed in experiment 1. Figures 4(a) and (b) illustrate linear fits of mean reaction times and error-rates for valid probe tones, respectively, as a function of dissonance. These analyses generally suggest that longer reaction times and higher error rates occurred for probe tones that were more dissonant with the expected probe tone. Table II illustrates that reaction times in experiment 2 also supported the importance of tonality. Asymmetry in pitch distance of valid and invalid probe tones about the reference tone, or the development of explicit strategies following extensive training, were thus not likely to have been essential to the results of experiment 1. Details of the experimental design, the instructions, and the training procedures appeared to influence absolute values of reaction times and error rates, but the essential pattern of reaction times, and how they related to dissonance, was preserved. In all three experimental contexts, reaction times were more strongly related to dissonance and tonality than to pitch proximity and registral direction.

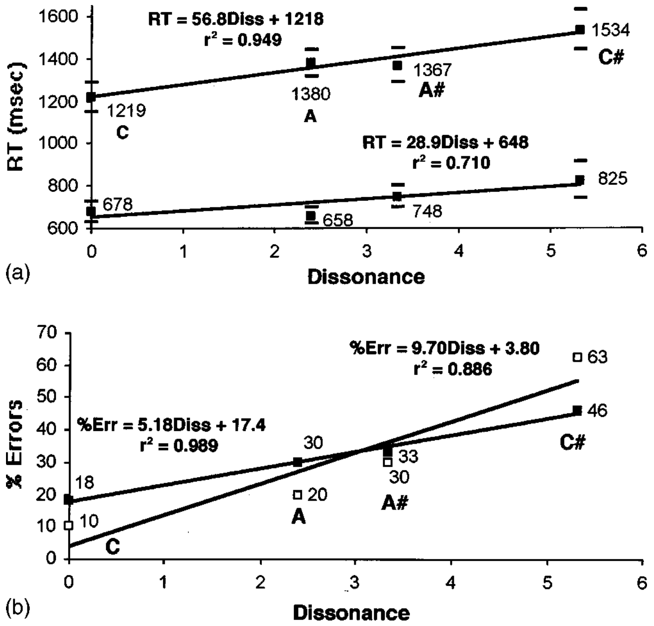

FIG. 4. (a) Mean reaction times (RT) for the four valid probe tones of experiments 2 (upper line) and 3 (lower line). The values of dissonance are as follows: C4=0, A3=2.40, A♯3=3.34, C♯4=5.32. (b) Percentage of errors committed by subjects in response to valid probe tones in experiments 2 (unfilled squares) and 3 (filled squares), plotted as a function of dissonance relative to the expected probe tone. Regression equations and squared correlations are shown.

The design of experiment 2 meant that responses were not always obtained when a valid probe tone was presented, and different probe tones elicted a different number of responses (0, 1, or 2). The number of responses made to each of the eight probe tones of experiment 2 was recorded as 0, 1, or 2, and a repeated measures ANOVA was conducted on the resultant data. As expected, there were significantly more responses made to valid probe tones (mean=1.54 responses per probe tone) than to invalid probe tones (mean=0.39 responses per probe tone), $F(1,19)=128.04$, $p<0.0001$. This finding indicates that participants were sensitive to the difference between valid and invalid probe tones.

Among valid probe tones, there were significantly more responses made to the expected note, C4, than to other valid probe tones, $F(1,19)=15.17$, $p<0.001$. Among unexpected valid probe tones, there were significantly more responses made to A3 than to A♯3 or C♯4, $F(1,19)=16.05$, $p<0.001$. Finally, there were significantly more responses made to A♯3 than to C♯4, $F(1,19)=16.38$, $p<0.001$. These findings parallel the reaction time data obtained in experiment 1, and suggest that as the psychological distance between probe tones and the expected note increased, participants tended to overestimate the distance in semitones between the probe tone and the last note of the scale.

Reaction times obtained in experiment 3 were assessed in a two-way ANOVA, with repeated measures on probe tone (four valid probe tones) and example (two responses for each probe tone). As expected, there was a main effect of probe tone, $F(3,42)=2.8508$, $p=0.05$, but no main effect of example, $F(1,14)=1.0547$, ns, and no interaction between probe tone and example, $F(3,42)=1.3413$, ns. These findings, summarized in Fig. 4(a), illustrate that the time taken to respond to different probe tones depended on the pitch of those probe tones.

As in experiment 1, the mean reaction time for the expected note, C4 (mean=678.37 ms), was lower than the mean reaction time for other probe tones (mean=743.84 ms), but in this case the difference was not significant, $F(1,42)=1.60$, ns. However, the mean reaction time for diatonic notes (C4 and A3, mean=668.32 ms) was significantly lower than the mean reaction time for nondiatonic notes (A♯3 and C♯4, mean=786.62 ms), $F(1,42)=6.95$, $p=0.02$. This finding supports the importance of one aspect of tonality in determining reaction times: scale membership. In addition, the mean reaction time for the probe tone at C♯4 (mean=825.47 ms), the most dissonant note with respect to the expected note, was significantly longer than the mean reaction time for all other notes (mean=694.80 ms), $F(1,42)=6.36$, $p=0.02$. This finding supports the importance of dissonance in determining reaction times. Taken together, the results again suggest that the time taken to shift attention from one pitch to another is proportional to the psychological distance between pitches, and that this distance is influenced by tonality and the degree of dissonance between pitches.

IV. EXPERIMENTS 4 AND 5

Experiments 1–3 indicated that listeners respond more slowly to unexpected notes than to expected notes following an incomplete major scale. Experiments 4 and 5 were conducted to examine the effect for another musical context—one that generates an expected note that is not proximal to the last note of the context. Listeners were presented an incomplete two-octave arpeggio (E2, G♯2, B2, E3, G♯, B3) followed by a probe tone. In this case, listeners were instructed to make a rapid key-press response to probe tones that were greater than three semitones from the last tone (the reference tone) of the arpeggio. Thus, the distance of valid and invalid probe tones from the reference tone was inverted. The arpeggio was used to direct the attention of participants to the next expected note of the arpeggio, E4. Again two experimental approaches were taken, paralleling the approaches of experiments 2 and 3. The approach of experiment 4 was to minimize training (preventing the development of complex strategies in responding to the probe tones) and to emphasize accuracy. The approach of experiment 5 was to increase training and emphasize speed of response.

A. Method

1. Participants

Twelve women and eight men from the York University community, ranging in age from 19 to 31 (mean=24.3), participated in experiment 2. Ten received undergraduate course credit for their participation, and ten were volunteers. Sixteen women and three men, aged between 18 and 54 (mean 23.0) participated in experiment 3 thirteen received undergraduate credit for their participation and six were volunteers. The number of years of formal musical training of participants ranged from zero to greater than ten (mean=6.8 years) in experiment 2 and from zero to ten (mean=2.9 years) in experiment 3. All but four of these participants also took part in experiments 2 and 3. All participants reported normal hearing.

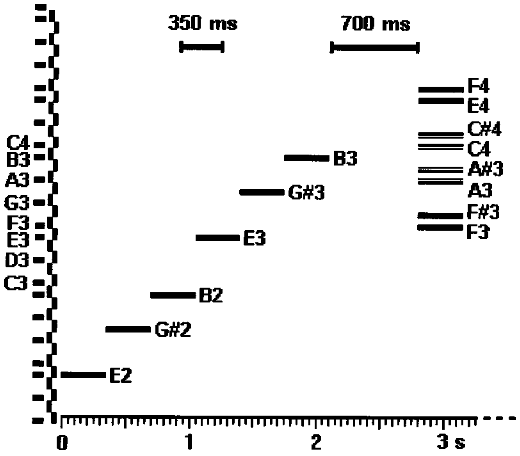

FIG. 5. The experimental design of experiments 4 and 5.

2. Materials and tone construction

The materials and tone construction were the same as those used in experiments 2 and 3.

3. Conditions and procedure

The stimuli used for experiments 4 and 5 are illustrated in Fig. 5. Listeners were presented an incomplete E major arpeggio, followed by a pause of 700 ms, and then a probe tone. The six notes of the arpeggio, E2, G♯2, B2, E3, G♯3, B3, were each 350 ms in duration, as was the duration of the probe tone. There were eight probe tones. Participants were instructed to make a rapid key-press response if the probe tone was three or more semitones from the last note of the arpeggio; otherwise they should make no response, and wait for the next trial. Thus, the response criterion was reversed with respect to that used in experiments 2 and 3. That is, participants in experiments 3 and 4 were instructed that they should not respond to probe tones whose pitch was proximal to the final tone of the arpeggio.

The four valid probe tones were F3, F♯3, E4, and F4. For both experiments 4 and 5, the maximum reaction time acceptable was set to 2000 ms. The four invalid probe tones were A3, A♯3, C4, and C♯4. As in experiments 2 and 3, upper and lower probe tones were symmetric in pitch distance from the reference tone: valid probe tones were exactly 5 and 6 semitones above or below the reference tone, and invalid probe tones were 1 and 2 semitones above or below the reference tone.

In experiment 4, each probe tone was presented twice, regardless of whether correct responses were obtained. In experiment 5, each invalid probe tone was presented twice, but each valid probe tone was presented as many times as needed to obtain two valid key-press responses (i.e., responses under 2000 ms). Thus, if no response to a valid

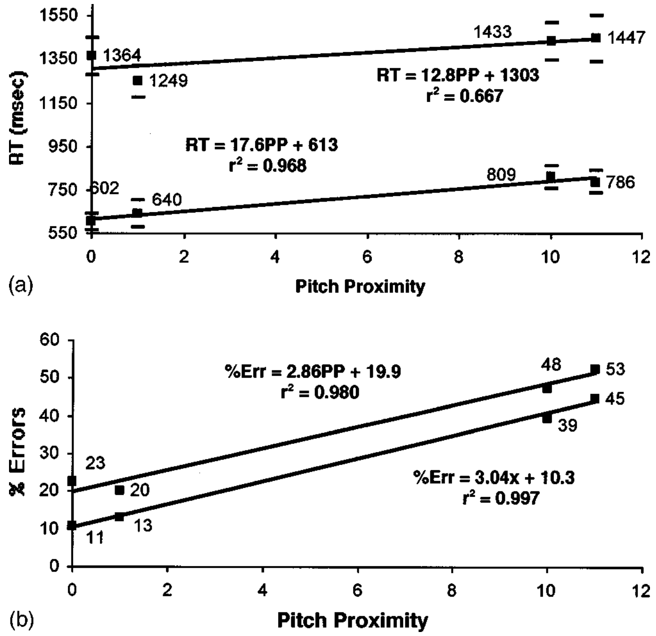

FIG. 6. (a) Mean reaction times (RT) for the four valid probe tones of experiments 4 (upper line) and 5 (lower line) plotted as a function of pitch proximity relative to the expected probe tone, E4. Values of pitch proximity were E4=0, F4=1, F♯3=10, F3=11; the units are semitones. (b) Percentage of errors committed by subjects in response to valid probe tones in experiments 4 (upper line) and 5 (lower line) as a function of pitch proximity.

probe tone was made by 2000 ms, that trial was repeated later in the experimental session. The order in which probe tones were presented was randomized independently for each participant. Other differences between experiments 4 and 5 were the same as those described for experiments 2 and 3.

B. Results and discussion

Mean reaction times and the percentage of errors obtained in experiments 4 and 5 are shown in Figs. 6(a) and (b), plotted as a function of pitch proximity relative to the expected note. Table II summarizes regression analyses conducted on both mean reaction times and error-rates for valid probe tones. These analyses suggest that reaction times and error-rates were not strongly related to dissonance and tonality, but rather, were more strongly related to pitch proximity and registral direction.

As in experiment 2, different probe tones elicted a different number of responses in experiment 4 (0, 1, or 2). Significantly more responses were made to valid probe tones (mean=1.46 responses per probe tone) than to invalid probe tones (mean=0.70 responses per probe tone), $F(1,19)=23.26$, $p<0.0001$. This finding indicates that, as with the scale context, participants were sensitive to the difference between valid and invalid probe tones. Among valid probe tones, there were significantly more responses made to probe tones in the expected registral direction (E4 and F4) than to probe tones in the unexpected registral direction, $F(1,19)=18.67$, $p<0.001$. This finding suggests that the arpeggio, unlike the scale, directed attention in an upward direction. There were no other significant effects.

Reaction time data obtained in experiment 5 were assessed in a two-way ANOVA, with repeated measures on probe pone (four valid probe tones) and example (two examples). As expected, there was a main effect of probe tone, $F(3,54)=6.43$, $p<0.001$, but no main effect of example, $F(1,18)<1.0$, ns, and no interaction between probe tone and example, $F(3,54)<1.0$, ns. Planned comparisons showed that the mean reaction time was lower for the expected probe tone, E4 (mean=601.82 ms), than for the other three probe tones (mean=745.03 ms), $F(1,54)=9.20$, $p<0.005$. This result suggests that the arpeggio context focused auditory attention on the pitch of E4, so that no shift in auditory attention was required to resolve that pitch, leading to shortest reaction times for this probe tone.

The mean reaction time was also significantly lower for probe tones in the same registral direction (E4 and F4, mean =620.86 ms) than for probe tones in the opposite registral direction (F3 and F♯3, mean=797.61 ms), $F(1,54)=18.69$, $p<0.0001$. Finally, among the three unexpected probe tones (F3, F♯3, F4), the two extreme notes (furthest away from the last note of the context, F3 and F4, mean=712.92 ms) had marginally lower reaction times than the note that was closer in pitch (F♯3, mean=809.26 ms), $F(1,54)=3.70$, $p<0.06$. The latter finding suggests that pitch proximity may influence reaction times following an incomplete arpeggio.

Overall, the results of experiments 4 and 5 support the notion that the time taken to resolve pitch depends on the psychological distance between the attended pitch and a presented pitch. In this case, however, registral direction and pitch proximity appeared to influence the psychological distance between pitches.

V. GENERAL DISCUSSION

The results of this investigation indicate that the time taken to respond to probe tones varies significantly depending on the pitch of that probe tone. In particular, pitch resolution in a musical context was determined by the psychological distance between the expected and actual pitch of the probe tone. This finding is analogous to effects reported for the visual system: response times for different visual cues are proportional to the distance between the location of the initial visual focus and the location of the target cue. In this case, the focus of auditory attention was established by generating a strong expectation for the completion of a pitch pattern (scale or arpeggio), which is a location in pitch space. Delays in responding to different pitches are interpreted as the time taken for participants to shift auditory attention from the expected pitch to the pitch of the probe tone.

The experimental designs illustrated in Figs. 1, 3, and 5 allowed an evaluation of four influences on pitch distance. Pitch proximity was defined in terms of a log frequency scale (number of semitones). Measures of dissonance and tonality were drawn from the work of Plomp and Levelt (1965) and Krumhansl (1990), respectively. Regression curves of reaction times as a function of these factors were determined and evaluated for their goodness of fit (Table II). Another parameter, registral direction, was assessed by comparing reaction times for probe tones in the expected registral direction with reaction times for probe tones in the opposite registral direction.

The process of shifting attention between pitches appears to be constrained by different factors depending on whether the preceding context is an incomplete major scale or an arpeggio. Perceived dissonance and tonality played an important role in the scale context, while neither of these parameters were significant factors in the arpeggio context. For the arpeggio context, reaction times were significantly faster for probe tones that continued the registral direction of the arpeggio than for probe tones that reversed the registral direction of the arpeggio. There was also moderate support for an influence by pitch proximity following the arpeggio context.

One interpretation of these results is that expectancies following the arpeggio context are most strongly associated with general (gestalt) principles of pitch relationships, whereas expectancies following the scale context are more strongly associated with harmony and tonality. Thus, the scale context may have invoked a more musical representation of pitch distance, which then determined reaction times. The arpeggio, in contrast, did not appear to invoke a rich musical representation of pitch distance, and judgments were therefore based on more general auditory principles (see Bregman, 1990; Deutsch, 1999).

This interpretation may appear to conflict with research conducted by Krumhansl and her associates, who found that harmonic materials induce a strong sense of tonality (Krumhansl, 1990). Two aspects of the harmonic stimuli used in those studies should be borne in mind, however. First, those studies involved minimizing effects of pitch height by using circular tones (Shepard, 1964). In contrast, we used noncircular tones (sine tones or piano tones), and thus did not inhibit influences from pitch proximity. Second, the harmonic materials in studies reviewed by Krumhansl (1990) were chord sequences, whereas we used an ascending arpeggio that extended across two octaves.

There are other possible explanations for the differences observed for the scale and arpeggio contexts. First, whereas the scale context involved stepwise melodic motion (i.e., intervals no greater than two semitones), the arpeggio involved leaps between successive notes, and also extended across a greater range of frequencies. The involvement of melodic leaps and the large extent of movement in the arpeggio context may have conveyed a strong sense of direction, distance, and momentum, thus overwhelming any influences by harmony and key.

Another factor may be the distance between the expected pitch and the valid probe tones. In the arpeggio context, the expected pitch (Fig. 5) was either the same or 1 semitone removed from the higher pitched valid probe tones, but 10 and 11 semitones away from the lower pitched valid probe tones. In the scale context, this difference was considerably reduced (0 and 1 semitone away from the higher pitched valid probe tones; 2 and 3 semitones away from the lower pitched valid probe tones). Thus, any effects of registral direction for the arpeggio context may have been bolstered by proximity to the attended pitch. That is, low-pitched valid and invalid probe tones may have been difficult to discriminate because both categories are distant in pitch from the expected pitch.

A notable feature of the scale results is a proportional increase of mean reaction times over experiments 1–3. Taking into account the details of the experimental designs, these three experiments might be classified as easiest (experiment 1), of intermediate difficulty (experiment 3), and most difficult (experiment 2). Mean reaction times increase accordingly at a ratio of 1:1.6:3.1. Interestingly, the slopes of the fits of reaction times as a function of dissonance increase across these experiments in the same way, the coefficient-ratios being 1:1.5:2.9. Moreover, these linear fits, when extrapolated backwards, all cross the abscissa at about the same point (reaction $\text{time}=0$, $\text{dissonance}=-20$, approximately). This finding suggests that the cognitive processes involved in recognizing the expected valid probe tone (doh), which take longer in more difficult situations, continue at a constant rate in the search and analysis of the remaining valid probe tones.

VI. CONCLUSIONS

The organization of pitch space and of shifts of auditory attention in this space depend on musical context, which shapes this space and directs auditory attention to a given position within the space. A familiar scale produces a pitch space in which yardsticks calibrated with dissonance and tonality are the best instruments to measure psychological distance. An arpeggio, in contrast, creates a pitch space in which psychological distances are determined more by registral direction and proximity.

ACKNOWLEDGMENTS

We are grateful to Luiz Eduardo Ribeiro do Valle, Marcus Vinícius C. Baldo and Luiz Renato Carreiro for assistance with software and several fruitful discussions especially concerning experiment 1. Tais Silveira Moriyama helped perform experiment 1 and Doug Gifford provided technical support for experiments 2–5. The research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada, awarded to the second and fourth authors. The third author was awarded a scholarship from the Conselho Nacional de Desenvolvimento Científico e Tecnológico and travel grants from the Fundação Faculdade de Medicina da Universidade de São Paulo and the Association of Universities and Colleges of Canada in the course of this work. Part of this research was presented at the annual meeting of the Canadian Society for Brain, Behavior and Cognitive Science, Ottawa, June 1998.

---

[1] Dissonance, as described by Plomp and Levelt (1965), is a complex concept that involves perceptual judgment, sensory organization of the cochlea and physical beats between harmonics of complex tones. Quantitative values of dissonance between complex tones were obtained by these authors by summing the separate contributions of dissonance between the closest pairs of the pure tones that constituted the first six harmonics of each complex tone. The values of dissonance for pure tones were extracted from measurement of the percentage of listeners who described two pure tones as consonant.

Bharucha, J. J., and Stoeckig, K. (**1986**). ''Reaction time and musical expectancy: Priming of chords,'' J. Exp. Psychol. **12**(4), 403–410.

Bharucha, J. J., and Stoeckig, K. (**1987**). ''Priming of chords: Spreading activation or overlapping frequency spectra?'' Percept. Psychophys. **41**(6), 519–524.

Bregman, A. S. (**1990**). *Auditory Scene Analysis* (MIT, Cambridge, MA).

Cohen, A. J. (**1991**). ''Tonality and perception: Musical scales primed by excerpts from 'The Well Tempered Clavier' of J. S. Bach,'' Psychol. Res. **53**, 305–314.

Deutsch, D. (**1975**). ''Two-channel listening to musical scales,'' J. Acoust. Soc. Am. **57**, 1156–1160.

Deutsch, D. (**1982**). ''The internal representation of information in the form of hierarchies,'' Percept. Psychophys. **31**, 596–598.

Deutsch, D. (**1999**). ''Grouping mechanisms in music,'' in *The Psychology of Music*, 2nd ed., edited by D. Deutsch (Academic, New York), pp. 299–348.

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

Dowling, W. J., and Harwood, D. L. (**1986**). *Music Cognition* (Academic, Orlando, FL).

Downing, C. J., and Pinker, S. (**1985**). ''The spatial structure of visual attention,'' in *Attention and Performance XI*, edited by M. I. Posner and O. S. M. Marin (Erlbaum, Hillsdale, NJ).

Egeth, H. E., and Yantis, S. (**1997**). ''Visual attention: Control, representation and time course,'' Annu. Rev. Psychol. **48**, 269–297.

Fletcher, N. H., and Rossing, T. D. (**1991**). *The Physics of Musical Instruments* (Springer-Verlag, New York).

Hikosaka, O., Miyauchi, S., and Shimojo, S. (**1996**). ''Orienting of spatial attention—its reflexive, compensatory, and voluntary mechanisms,'' Cognitive Brain Research **5**, 1–9.

Janata, P., and Reisberg, D. (**1988**). ''Response-time measures as a means of exploring tonal hierarchies,'' Music Perception **6**(2), 161–172.

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

Krumhansl, C. L. (**1990**). *Cognitive Foundations of Musical Pitch* (Oxford U.P., Oxford, England).

Krumhansl, C. L. (**1995**). ''Effects of musical context on similarity and expectancy,'' Systematische Musikwissenschaft (Systematic Musiology) **3**(2), 211–250.

Krumhansl, C. L., and Kessler, E. (**1982**). ''Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys,'' Psychol. Rev. **89**, 334–368.

Krumhansl, C. L., and Shepard, R. N. (**1979**). ''Quantification of the hierarchy of tonal functions within a diatonic context,'' J. Exp. Psychol. **5**, 579–594.

Mondor, T. A., Breau, L. M., and Milliken, B. (**1998**). ''Inhibitory processes in auditory selective attention: Evidence of location-based and frequency-based inhibition of return,'' Percept. Psychophys. **60**(2), 296–302.

Narmour, E. (**1990**). *The Analysis and Cognition of Basic Melodic Structures* (Univ. of Chicago, Chicago).

Plomp, R., and Levelt, W. J. M. (**1965**). ''Tonal consonance and critical bandwidth,'' J. Acoust. Soc. Am. **38**, 548–560.

Posner, M. I. (**1985**). *Chronometric Explorations of the Mind* (Oxford U.P., New York).

Posner, M. I., and Cohen, Y. (**1984**). ''Components of visual orienting,'' in *Attention and Performance X*, edited by H. Bouma and D. G. Bowhuis (Erlbaum, Hillsdale, NJ), pp. 531–555.

Rhodes, G. (**1987**). ''Auditory attention and the representation of spatial information,'' Percept. Psychophys. **42**(1), 1–14.

Schneider, W. (**1988**). ''Micro Experimental Laboratory: An integrated system for IBM-PC compatibles,'' Behav. Res. Methods Instrum. Comput. **20**, 206–217.

Shepard, R. N. (**1964**). ''Circularity in judgments of relative pitch,'' J. Acoust. Soc. Am. **36**, 2346–2353.

Spence, C., and Driver, J. (**1998**). ''Auditory and audiovisual inhibition of return,'' Percept. Psychophys. **60**(1), 125–139.

Steinman, S. B., and Steinman, B. (**1998**). ''Vision and Attention. I: Current Models of Visual Attention,'' Optom. Vision Sci. **75**(2), 146–155.

Tekman, H. G., and Bharucha, J. J. (**1992**). ''Time course of chord priming,'' Percept. Psychophys. **51**(1), 33–39.

Thompson, W. F., and Stainton, M. (**1998**). ''Expectancy in Bohemian folk song melodies: Evaluation of Implicative Principles for implicative and closure intervals,'' Music Perception **15**, 231–252.

Tsal, Y. (**1983**). ''Movements of attention across the visual field,'' J. Exp. Psychol. **9**, 523–530.

Umiltá, C. (**1988**). ''Orienting of attention,'' in *Handbook of Neuropsychology*, edited by F. Boller and J. Grafman, Vol. 1, pp. 175–193.