

Studying expertise in music reading: Use of a pattern-matching paradigm

ANDREW J. WATERS

University College London, London, England

GEOFFREY UNDERWOOD

University of Nottingham, Nottingham, England

and

JOHN M. FINDLAY

University of Durham, Durham, England

Two experiments are described that make use of a pattern-matching paradigm to investigate perceptual processing of music notation. In Experiment 1, it is reported that the speed of comparing two visually presented musical sequences is related to the sight-reading skill of the subjects. The effect of the temporal and pitch structure of the comparison stimuli is also assessed. In Experiment 2, eye-movement recordings were taken as subjects performed the task. These data demonstrated that more experienced musicians are able to perform the comparisons with fewer, and shorter, glances between the patterns. These and other findings suggest that skilled sight-reading is associated with an ability to rapidly perceive notes or groups of notes in the score, and confirm that the pattern-matching paradigm is a useful tool in examining expertise in music reading.

What are the critical psychological processes that mediate skilled music reading? Although the last 20 years have seen an immense amount of research into skilled performance in knowledge-rich domains, such as chess, scientific problem solving, and computing (Ericsson & Smith, 1991), there have been relatively few studies investigating expertise in music reading (see Sloboda, 1984, for the most recent review). The purpose of the present paper is to describe a new paradigm in music-reading research that can be used to elucidate some of the perceptual processes underlying skilled music reading.

Ericsson and Smith (1991) have outlined a three-stage research approach, the "expertise approach," that has been adopted by researchers in different domains. First, expert-novice differences are observed in the laboratory in a collection of carefully constructed tasks that elicit such differences. Second, the cognitive processes critical to the production of the performance on such tasks are described and analyzed, using the full range of methods of analysis available in cognitive psychology. Finally, the critical cognitive processes are examined, and learning mechanisms are proposed to account for their acquisition. Using this approach, many studies have been conducted on the cognitive abilities of chess experts (e.g., Chase & Simon, 1973), bridge players (e.g., Charness, 1979), Go

and Gomoku players (Reitman, 1976), crossword solvers (Underwood, Deihim, & Batt, 1994), expert musicians (Sloboda, 1984), and computer programmers (McKeithen, Reitman, Rueter, & Hirtle, 1981), to name but a few domains. From investigations into chess expertise, for example, it has been suggested that master chess players have acquired a large vocabulary of chess patterns, or chunks, that are associated with plausible moves or evaluations (e.g., Chase & Simon, 1973; de Groot, 1978).

From a musical perspective, musical sight-reading represents a useful skill for all musicians and a critical skill for many (Sloboda, 1978b). In many real-life and examination situations (e.g., Associated Board examinations¹), musicians are required to perform a novel passage at first sight. From a cognitive psychological perspective, sight-reading represents a complex transcription task (Sloboda, 1982). Other examples of transcription tasks, such as copy typing, have been intensively studied (e.g., Gentner, 1987; Salthouse, 1984). In sight-reading, the reader has an additional burden of trying to make the performance "musical," as well as producing an accurate transcription (Sloboda, 1985b).

Given that sight-reading represents a complex transcription task, it is apparent that many different types of processing ability must underlie sight-reading expertise. First, one can hypothesize that the expert reader acquires some kind of perceptual or *pattern recognition* skill that allows for rapid and relatively automatic perceptual processing of commonly occurring musical structures within the musical text (Wolf, 1976). Skilled readers seem able to recognize groups or musically meaningful "chunks" of notes more rapidly than less skilled readers (Bean, 1938;

The authors acknowledge the Science and Engineering Research Council for providing the funding for the research, and Robin Walker and Bob Kentridge for help in the experimental work. Correspondence should be directed to A. J. Waters, Departments of Psychology/Epidemiology, H.B.U., Brook House, 2-16 Torrington Pl., London WC1H 0LP, England (e-mail: rmjdajw@ucl.ac.uk).

Salis, 1980; Sloboda, 1978a). It is possible that expert sight readers have acquired some kind of "dictionary" of musical patterns that are engaged during music reading to facilitate the processing of musical material (Sloboda, 1974, p. 241). Skilled readers may also be able to code and store musical information in short-term memory more effectively (Clifton, 1986; Halpern & Bower, 1982; Sloboda, 1976b; Thompson, 1987).

Second, one can also expect skilled sight readers to have acquired general knowledge of music structure (e.g., knowledge of harmony, counterpoint, and form; Sloboda, 1978b). Such grammatical knowledge may facilitate the reading process in a number of ways; for example, musicians may learn to use context to anticipate the continuations of musical passages, thus decreasing the information-processing load. The strength of such expectations for tonal music was illustrated by Sloboda (1976a), who showed that skilled pianists could be induced to make "proofreader's errors" when various alterations were introduced in a tonal passage. Sloboda (1977, 1984) has further argued, from use of the eye-hand span task, that skilled sight readers are more sensitive than less skilled readers to "superordinate" musical structures within notation. Skilled musicians are also likely to have developed an expressive expertise or style, which may be brought to bear in sight-reading performance (Sloboda, 1983, 1985a). Finally, it is clear that sight readers will need to have acquired adequate motoric skills to perform a piece accurately and expressively. A number of studies have examined the nature of motor programming in piano performance (e.g., MacKenzie & van Eerd, 1990; Shaffer, 1981).

The present paper focuses on perceptual processing skills in music reading. This choice was motivated by two considerations. First, the fact that musicians can have similar performance abilities but very different sight-reading abilities (Wolf, 1976) indicates that such musicians are likely to have similar cognitive and motoric abilities in the musical domain, but may differ widely in perceptual abilities. Second, there is good evidence already cited that skilled and less skilled readers do differ in perceptual processing skills (e.g., Salis, 1980).

More specifically, it has been suggested that skilled reading is dependent on the recognition of groups, or "chunks," of notes in the score, rather than on processing items in a note-by-note manner (Fasanaro, Spitaleri, & Valiani, 1990; Wolf, 1976). According to this view, at least *part* of sight-reading expertise is dependent on the acquisition of a vocabulary of commonly occurring note groups, or chunks, that can be rapidly encoded and processed in reading. Skilled readers may possess a larger vocabulary with larger chunks (cf. chess studies) and/or they may simply be able to process the chunks more rapidly. Furthermore, Goolsby (1994a, 1994b) has presented evidence from eye-movement recordings (as subjects vocalized melodies) that skilled readers do not fixate each note for efficient performance. The less skilled readers do, however, "fixate on as much of the notation as the time allows (i.e., looking at notes/rests in proportion to rhythmic values), progress note by note, and per-

form with numerous errors" (Goolsby, 1994b, p. 121). Goolsby suggests that some system of chunking may be used by skilled readers to grasp more than one note, or item of visual detail, during a single fixation.

There are three aims to this paper. First, a pattern-matching paradigm is introduced that can be used to directly test the hypothesis that skilled sight-reading is associated with an ability to rapidly process groups of notes. Second, some response-time and eye-movement data are presented with the use of such a paradigm. Third, suggestions are made as to how the paradigm can be developed for future investigations into perceptual processing skills in music reading.

The Pattern-Matching Task

The first stage of the expertise approach is to devise a variety of tasks in which expert/novice differences can be observed and measured. The paradigm adopted in this paper is a "same-different" matching task, which has been utilized in expertise research (e.g., Ellis, 1973) and music psychology (e.g., Beal, 1985). In this task, the subject is shown two (visually presented) musical strings presented on the screen of a computer monitor, as shown in Figure 1.

The subject is required to compare the stimuli as quickly but as accurately as possible, and to press one of two response buttons to register his/her decision (same/different). The stimulus shown in Figure 1 represents a *same* trial, since the two musical strings are identical in pitch and temporal structure. On *different* trials, there is a small alteration in either the pitch of one of the notes or the duration of one of the notes. The subjects are thus required to process both the pitch and the durational structure of the stimuli. In the version of the task used in this paper, the stimuli were all composed of 10 notes (2 bars of 5 notes each) since this provided sequences of sufficient length to allow for bar-by-bar matching strategies, while not being so long as to result in very long response times (normally associated with large variances). In principle, however, the pattern-matching task can be interesting with stimuli of *any* length, and indeed the number of notes might prove to be an interesting variable to manipulate.

There are three advantages to adopting this task. First, the task can lend itself to an investigation of the chunking processes of subjects of different levels of expertise. For example, Ellis (1973) required chess players of different standards to report whether two simultaneously presented quarter-board chess positions were the same



Figure 1. Example of two sequences for comparison.

or different. The slope of reaction time as a function of number of pieces in the two diagrams (three to seven pieces) was virtually flat for strong players but showed a strong linear relationship for weak players. This result supported the idea that for strong players a large cluster of pieces is the basic unit for matching, whereas for weaker players single pieces may be the basic unit.

Second, the same-different matching paradigm is ideally suited for investigating performance in conditions in which the *structure* of the material to be compared can be manipulated in a systematic fashion. As will become clearer later, stimuli in the present studies were written in one condition to possess "coherent" temporal and pitch sequences. In the spirit of other expertise research (e.g., Chase & Simon, 1973), stimuli in other conditions were generated simply by randomizing the position of pitch and durational elements within a stimulus, and these stimuli were said to possess "randomized" temporal and pitch sequences. The time taken for subjects to compare stimuli of different structural characteristics can thus be determined.

Third, eye-movement recordings can be taken as subjects perform the comparisons. This approach was taken in Experiment 2. The purpose of this is to give some indication as to *how* subjects of different skill levels perform the task. In particular, eye-movement recordings can give some indication of the sizes of units used in the comparisons and reveal how comparison behavior is affected by the structure of the stimuli being compared.

EXPERIMENT 1: PILOT

An initial pilot study was conducted to determine the skill sensitivity of the pattern-matching paradigm. It was predicted that more skilled readers would perform better in speed and/or accuracy at the pattern-matching task than would less skilled readers. If, however, performance at the task turned out not to be skill sensitive, then the task would be of only very limited interest, since this would probably indicate that expert and novice subjects were performing the task like some kind of speeded visual search task (i.e., without necessarily engaging musical knowledge). A further aim of this experiment was to investigate the effect of the structure of the stimuli on task performance. In this pilot study, performance at the matching task was measured for four different types of stimuli: temporally coherent, pitch coherent; temporally coherent, pitch randomized; temporally randomized, pitch coherent; and temporally randomized, pitch randomized (see Figure 2).

There is much evidence from the music psychology literature that pitch structure affects ease of music processing with both auditory (Deutsch, 1980) and visual musical material (Halpern & Bower, 1982). Sloboda (1978a) demonstrated that musicians can perceive and recall the contour of simple musical patterns presented very briefly (100 msec), while nonmusicians are less able to retain the contour at such presentation times. Deutsch (1980) further reported that the rhythmic struc-



(a) Temporally coherent, Pitch coherent stimulus



(b) Temporally coherent, Pitch randomized stimulus



(c) Temporally randomized, Pitch coherent stimulus



(d) Temporally randomized, Pitch randomized stimulus

Figure 2. Examples of stimuli in each condition of Experiment 1.

ture of groups of notes was found to have a substantial effect on recall performance in a melodic dictation task. Results indicated that information was processed in temporally defined chunks. This result is consistent with other studies which have demonstrated the importance of temporal grouping in the processing of musical and nonmusical information (Bower & Springston, 1970; Dowling, 1973; Handel, 1973; Restle, 1972).

Since construction of the materials and the procedure were very similar to those of the main study (Experiment 2), this is now discussed in some detail.

Method

Materials

Thirty melodies (henceforth termed *originals*) were composed² in accordance with the following constraints. They were all 10 notes long, written in 3/4 or 4/4, consisting of 2 bars of 5 notes. None contained a key signature, there were no accidentals, and each melody contained 5–8 different pitches and 3–5 values of different durations. They were all written so as to be temporally and pitch coherent. An example is shown in Figure 2a.

The 30 original melodies were randomly assigned to the *same*, *pitch different*, and *duration different* condition. The 10 original melodies allocated to the *same* condition constituted the temporally coherent, pitch coherent stimuli. For each of these stimuli, three further structural configurations were generated. For the temporally coherent, pitch randomized stimuli, the temporal structures of the original stimuli were maintained, while the pitch structures were disrupted by randomizing the sequence of pitches of the originals. This transformation disrupted the contour of the melody, resulting in large pitch jumps and multiple contour changes (while still being composed of the original pitch elements). On the rare occasions when the randomization exactly or almost exactly matched the pitch structure of the original, the process was repeated to regenerate a new randomized pitch pattern. For Figure 2a, the transformation yielded the following stimulus shown in Figure 2b. For the temporally randomized, pitch coherent stimuli, the temporal structures

were disrupted by randomizing the durational elements while keeping constant the pitch sequence of the originals (see Figure 2c). This was done under the constraint that the number of beats in each bar be maintained; if the randomization process exactly mirrored the temporal structure of the original, the process was repeated to regenerate a new randomized temporal pattern. Finally, the temporally randomized, pitch randomized stimuli were generated by pairing together the temporally randomized and pitch randomized sequences, producing Figure 2d.

For each of 40 *same* stimuli, a duplicate was produced, which served as the comparison stimulus. This was prepared so that there was a gap of 2.5 cm between the sequences on presentation. For the pitch- and duration-*different* conditions, the 10 originals were treated in an analogous fashion to create 40 stimuli, with 10 in each of the four different structural conditions. For the pitch-*different* trials, the comparison sequences were not, of course, precise duplicates but sequences in which one of the notes differed in pitch. The temporal structure was identical in the two comparison stimuli, only one of the pitches differed in the comparison stimulus, and the altered pitch differed by one "notch" (or "second") on the staff upwards or downwards (e.g., from G to A or from E to D), thus avoiding a large visual difference between the two stimuli. Each note position (1–10, henceforth termed *serial position*) differed once for each of the four structural conditions. An example of a pitch-different trial is shown in Figure 3 (the sixth note differs).

For the duration-*different* trials, the comparison stimuli were sequences in which one of the durations of the notes differed. The same conditions applied as in the generation of pitch-*different* trials except that the altered comparison duration was half or double the original duration; the pitch sequence was of course unaltered. Figure 4 illustrates a duration-*different* trial (the seventh note differs).

Two important consequences of this manipulation should be highlighted. First, in the comparison stimulus, the second bar does not contain the requisite number of beats and thus violates one of the conventions of music. Second, the manipulation violates the space-duration relationship in music whereby longer notes generally have longer gaps between notes (indeed, the space between notes can be considered an important cue to the duration of the note; see Sloboda, 1981). Subjects were informed prior to experimentation of the effect of this manipulation.

Procedure

Three groups of subjects participated in the experiment: two "expert" groups (full time music students who had achieved Associated Board Grade VIII in a monophonic instrument) and a novice group (no novice had progressed beyond Associated Board Grade II, although they were all familiar with the names of the notes). Subjects in one of the expert groups had made high scores in the Grade VIII sight-reading examination and had given high self-rating scores for sight-reading (henceforth termed Group 1, $n = 11$); subjects in the other expert group had scored lower in the sight-reading test and also reported lower self-rating scores (henceforth termed Group 2, $n = 11$). Each subject was tested in a cubicle containing an Apple Macintosh computer and keyboard. They were correctly informed by the experimenter that one third of the



Figure 3. Examples of a pitch-difference trial.



Figure 4. Example of a duration-difference trial.

trials would be of the *same* type, the others being pitch- and duration-*different* trials in equal proportions. The subjects were instructed to press one button with the right hand for a *same* response and another button with the left hand for a *different* response. They were correctly informed that any note position (1–10) could be the locus of a difference, that each position was equally likely to differ, and that there would be only one difference (pitch or duration) for each *different* trial. They were told to make their responses as quickly and as accurately as possible, and that the practice trials would give them some idea of how fast they could make their comparisons. Stimuli were presented using the software package Psychlab, and all experimental parameters and timing of responses were controlled and recorded by Psychlab.

The subjects positioned themselves at a distance from the screen at which they felt most comfortable (around 50 cm). Each trial was initiated by the presentation of a fixation cross for 2 sec to the left of center of the screen. The fixation cross functioned both to alert the subject to the onset of a trial and to ensure that the subject was fixating the appropriate position on the screen. Thereafter, a stimulus was presented whereby the clef sign of the upper sequence was positioned where the fixation cross had been presented. The stimulus remained on the screen until the subject had made his/her response. After each response, the stimulus disappeared, and there was an intertrial interval of 3 sec, after which the next trial was initiated by the presentation of the fixation cross. The subjects were not given any feedback as to the accuracy of their responses. After the practice trials, the subjects completed two experimental blocks of 60 trials each, separated by a break of 10 min. Presentation of stimuli was randomized for each subject. Finally, the subjects were debriefed as to the nature of the experiment and asked a few informal questions about the task.

Results and Discussion

The critical results were that there were clear quantitative and qualitative differences in task performance between the three groups. As predicted, there were large effects of expertise in reaction times in this pilot study but no large differences in overall accuracy of task performance (Group 1, 77.7% accuracy; Group 2, 81.5%; novices, 79.6%). For example, in the same trials, Group 1 performed the comparisons with a mean of 3,830 msec, Group 2 did so with a mean of 4,720 msec, and the novices with a mean of 8,043 msec. All these differences were statistically significant (Group 1 vs. novices, $t = -7.44, p < .01$; Group 2 vs. novices, $t = -5.58, p < .01$; Group 1 vs. Group 2, $t = -2.18, p < .05$). There were also large expertise differences in response times on *different* trials, with Group 1 performing the comparisons with a mean of 2,646 msec, Group 2 with a mean of 3,150 msec, and the novices with a mean of 5,362 msec.

Table 1 gives the mean reaction times from the *same* trials.

As can be seen, both expert groups were slower to respond to temporally randomized material than to temporally coherent material in the *same* trials, whereas the novices showed no such sensitivity to temporal structure, and the expected expertise \times temporal structure interaction was significant [$F(2,30) = 5.55, p < .01$]. This finding is consistent with many other findings in the expertise literature, which have demonstrated a decline of performance in experts, but not in novices, when material is randomized (e.g., Chase & Simon, 1973). The effect size was, however, rather small (around 300 msec), when compared with the expertise effect. Furthermore, Group 2, though not Group 1 or the novices, was slower to respond to temporally randomized material than to temporally coherent material in the *different* trials. These data suggest that the temporal structure of the material generally has an effect on the comparison behavior of the experts. For example, temporal randomization may affect chunking strategies in task performance. Although overall reaction times were slower for pitch-randomized material than for pitch-coherent material, there was no evidence of an expertise \times pitch structure interaction.

Furthermore, although there was no difference in overall accuracy in task performance between the three groups, there was an interesting difference in the *types* of errors made. Both expert groups made more errors on duration-*different* trials than on pitch-*different* trials, whereas the novices showed a slight effect in the opposite direction. Although not predicted, this effect is understandable if one assumes that experts have difficulty in detecting duration differences because of the violation of the space-duration relationship that occurs as a result of producing duration alterations. Experts may code duration not only from the shape and color (white/black) of the note but also from the amount of space separating it from its neighbor; if they use the latter code in this task, this will produce poor performance. The novices, on the other hand, presumably coded the durations simply from the shape and color of the note, and thus were less likely to overlook duration differences.

It is apparent that this pilot study has thus gone some distance toward validating the use of the pattern-matching paradigm. The aim of Experiment 2 was to determine more precisely the nature of the experts' advantage at this task by taking *eye-movement* recordings as subjects performed the comparisons. In particular, is the experts' advantage due to their processing larger units in the comparison pro-

cess, or do all readers use similar-sized units, which the experts simply process more rapidly? Another aim of Experiment 2 was to determine how the effect of temporal randomization affected the experts' comparison behavior.

EXPERIMENT 2

There were three aims in this experiment. First, and most important, it was hoped that the eye-movement recordings might provide an explanation for the expertise effect in the pilot study by shedding light on the comparison behavior of the subjects in this task. Moreover, the relationship between expertise and eye-movement parameters, such as durations of fixations, could be scrutinized. Second, the experiment could investigate the effect of temporal randomization of stimuli on the comparison behavior of the experts. Last, we sought to determine the effect on performance of a new factor—pattern separation. This factor referred to the distance between the two musical patterns presented on the screen. There were two levels, "near" (patterns separated by 2.5 cm, the same as in Experiment 1) and "far" (patterns separated by 7.5 cm). The idea of the pattern-separation manipulation was to determine whether subjects could be induced to adopt different-sized units in performing the comparisons.

Since the eye-movement recordings are critical to the study, a few words are required on eye movements and music reading. Although sophisticated apparatus is now available for measuring precisely the location and duration of fixations on a passage of musical text (e.g., Goolsby, 1989, 1994a, 1994b; Reulen et al., 1988), there has still been very little research conducted on music reading (compared with text reading). In most of the previous studies, subjects have been required to produce a motoric output (i.e., vocal or instrumental performance) while reading a musical sequence (Goolsby, 1994a, 1994b; Halverson, 1974; Schmidt, 1981; Van Nuys & Weaver, 1943; York, 1952). In the study reported here, the subjects were clearly doing something "cognitively" and "motorically" different from what they do in the normal reading situation. Thus, the eye-movement parameters reported in this study might differ radically from previously observed findings. Nonetheless, the previous literature may supply a tentative source of hypotheses as to the relationship between eye-movement parameters and reading skill.

In this area, Goolsby (1994a, 1994b) and Kinsler and Carpenter (1995) have completed the most extensive and technologically sophisticated studies conducted to date, although Kinsler and Carpenter were not concerned with the effect of expertise on eye-movement behavior. Using an SRI Eyetracker, Goolsby (1994a, 1994b) recorded eye-movement behavior as skilled and less skilled music readers read melodies of varying levels of notational complexity. Results indicated that the skilled readers used more, but shorter (i.e., briefer), fixations than did the less skilled readers. In a second paper, Goolsby (1994b) provided a detailed comparison of a skilled and a less skilled reader. This indicated that the skilled reader was characterized

Table 1
Mean Reaction Times (in Milliseconds)
on *Same* Trials, Experiment 1

Group	Temporally Coherent		Temporally Randomized	
	Pitch Coherent	Pitch Randomized	Pitch Coherent	Pitch Randomized
1	3,540	3,838	3,920	4,020
2	4,438	4,682	4,876	4,884
Novices	7,927	8,215	7,918	8,110

by *not* fixating on every note, whereas the less skilled reader did fixate on virtually every note and most rests.

From these studies, it was possible to predict that skilled readers would use briefer fixations than less skilled readers. It was also possible to predict that more skilled readers would use fewer fixations to compare the patterns, since they appeared to be more efficient in the number of fixations required to process material.

Method

Subjects

Three groups of subjects were recruited to take part in the experiment. Group 1 was composed of 8 full-time music students from the Department of Music, University of Durham. Each subject had passed an Associated Board Grade VIII examination in at least one instrument associated with the treble clef register. Group 2 was composed of 8 psychology students, who had all passed an Associated Board Grade IV, V, VI, or VII examination in a monophonic instrument. The third group consisted of 8 nonmusicians. These subjects possessed little musical experience, and none had passed an Associated Board examination in any instrument. The subjects' ages ranged from 18 to 29.

Apparatus

The experiment was run in the perception laboratory at the Department of Psychology, Durham University. Stimuli, generated by the software package Deluxe Music, were presented on an Apple Macintosh Quadra. All experimental parameters and timing of responses were controlled and recorded by the presentation package Display and Record.³ This package presents a stimulus by pre-drawing it invisibly in the screen background color during the fixation period and then changing the color look-up table to present it within a single refresh frame, at the same time initiating the eye-movement recording. Eye movements were recorded using the differential limbus reflection technique and utilizing a binocular infrared system with the Skalar IRIS system (described by Reulen et al., 1988). This was set up to record the horizontal signal from one eye and the vertical signal from the other eye. The resolution (minimum detectable saccade) was about 0.1° and accuracy around 0.4° horizontally and 0.8° vertically. The analogue signal was sampled every 5 msec for as long as the subject took to respond to each trial. Digital records were stored on disk for later examination using a semiautomatic saccade-detection program. A chinrest was used to limit head movements throughout the test session.

Materials

Sixty melodies were generated in the same way as in the pilot experiment, and were randomly divided into two sets, A and B. For each melody, a temporally randomized counterpart was generated in the same way as in the pilot study. The 60 stimuli in each stimulus set were randomly split into three groups of 20 each, corresponding to the *same*, *pitch-different*, and *duration-different* trials. Two trials were produced per stimulus, one for the near condition and one for the far condition. For each of the *pitch-different* and *duration-different* stimuli, a comparison stimulus was produced according to the same constraints as in Experiment 1.

Four experimental blocks of 60 trials each were created: Set A–near and Set A–far; Set B–near and Set B–far. Two practice blocks of 12 trials each were also generated for the near and far trials. In each subject group, 4 subjects performed Set A–near and Set B–far, and the remaining 4 performed Set B–near and Set A–far. This arrangement meant that a subject did not see the same stimulus twice (in the near and far conditions). Prior to performance of a near or far experimental block, the subjects were run on the near or far prac-

tice block. Order of completion of near and far blocks and order of presentation of trials were determined randomly for each subject.

Design

The experiment had a three-factor mixed design, with one between-subjects factor, expertise, with three levels (Group 1 vs. Group 2 vs. nonmusicians), and two within-subjects factors; pattern separation, with two levels (near vs. far), and temporal structure, with two levels (temporally coherent vs. temporally randomized). The dependent variables were the response times and response choice for the comparisons and the eye-movement parameters associated with comparison behavior.

Procedure

Each subject first received an instruction sheet, informing him/her on the details of the experiment, and then underwent the calibration procedure. The subject was seated about 50 cm from the presentation monitor, with his/her head positioned on the chinrest set at an appropriate height for his/her build. The lightweight infrared transducer was inserted on the subject's head and tightened such that it was secure but comfortable. The transducers were then adjusted independently to gain an optimal recording position for each eye. The subject was instructed to serially fixate nine dots, arranged as a square matrix on the presentation monitor, so as to calibrate them on both the horizontal and vertical dimensions. There were two calibrations per subject, one before the near trials and one before the far trials. The subject was encouraged to keep his/her head still throughout the experiment. The procedure for each trial was the same as in Experiment 1. Each subject completed two practice blocks and two experimental blocks. The whole experimental session lasted about 40 min, after which the subject was debriefed and thanked for his/her participation.

Results

This section is divided into two parts. The first part presents the analyses of response data (reaction times and accuracy measures); the second part presents analyses of eye-movement measures.

Analysis of Reaction Time and Accuracy Data

Median reaction times were computed for each subject from correct responses in each condition. The number of errors was also aggregated for each subject in each condition. Overall, Group 1 performed at 80.3% accuracy ($SE = 3.54$), Group 2 performed at 83.8% ($SE = 2.43$), and the nonmusicians at 86.9% ($SE = 2.77$); a one-way ANOVA did not produce a main effect of expertise [$F(2,21) = 1.24, p > .10$]. As in Experiment 1, both Group 1 and Group 2 were much faster in performing the comparisons than was the nonmusician group, and a natural log transform was used to reduce the disparity in variance between the groups for all analyses involving reaction-time data.

Analysis of *same* trials. Table 2 shows the means in milliseconds for each subject group in each condition.

The pattern of data indicated that, in common with the pilot study, there were large effects of expertise together with much smaller effects of temporal structure. In addition, there appears to be an expertise \times temporal structure interaction similar to that reported in the pilot study, and faster responses for near than for far trials. A three-way ANOVA (expertise \times pattern separation \times tempo-

Table 2
Mean Reaction Times (in Milliseconds)
on Same Trials, Experiment 2

Group	Near		Far	
	Temporally Coherent	Temporally Randomized	Temporally Coherent	Temporally Randomized
1	3,685	3,932	4,146	4,478
2	4,979	5,250	5,141	5,388
Nonmusicians	10,373	10,229	10,647	10,797

ral structure) performed on the transformed means from the same trials confirmed this appraisal. The analysis produced a significant expertise × temporal structure interaction [$F(2,21) = 4.30, p < .05$]. Simple main effects computed for this interaction indicated that Group 1 and Group 2 were significantly faster with the coherent material than with the randomized material [Group 1, $F(1,21) = 16.4, p < .01$; Group 2, $F(1,21) = 8.19, p < .01$], but the nonmusicians did not show any significant difference between these conditions [$F(1,21) = 0.01, p > .1$]. As expected, there was a main effect of expertise [$F(2,21) = 24.2, p < .01$] and planned comparisons between the three groups revealed that the Group 1–nonmusicians [$t(21) = -6.70, p < .01$] and Group 2–nonmusicians comparisons [$t(21) = -4.96, p < .01$] were statistically significant; the Group 1 versus Group 2 comparison showed a trend toward significance [$t(21) = -1.74, p < .10$]. Finally, there was a main effect of pattern separation [$F(1,21) = 12.4, p < .01$], indicating that near trials were responded to more quickly than far trials (5,855 vs. 6,242 msec).

There were few errors (<2%) made on same trials (i.e., responding “different” on same trials), and hence they were not analyzed. More errors were made on different trials (responding “same” on different trials), and these are presented below.

Analysis of different trials. The percentage of errors made in each condition is given in Table 3.

This pattern of data is very similar to that documented in the pilot study, with the musician groups making more errors on the duration-different trials than did the nonmusicians, and vice versa for the pitch-different trials. A four-way ANOVA (expertise × trial type × pattern separation × temporal structure) was performed on the number of errors made on the different trials. Trial type refers to pitch- or duration-different trials. The analysis produced the expected expertise × trial type interaction [$F(2,21) = 10.4, p < .01$]. Simple main effects computed for this interaction revealed that Group 1 [$F(1,21) = 19.4, p < .01$] and Group 2 [$F(1,21) = 9.41, p < .01$] made more errors on duration-different trials than on pitch-different trials. There was a trend in the opposite direction for the nonmusicians [$F(1,21) = 3.02, p = .097$]. There was, however, no main effect of expertise [$F(2,21) = 1.58, p > .10$], indicating that there were no reliable differences in accuracy of task performance between the three groups on the different trials (making the possibility of a speed–accuracy tradeoff between groups unlikely). The within-group correlations between speed and

accuracy on different trials were for Group 1, $r = 0.21$, Group 2, $r = 0.17$, and nonmusicians, $r = 0.10$, indicating no substantial speed–accuracy tradeoffs within each group. Finally, there was also a main effect of pattern separation [$F(1,21) = 5.69, p < .05$], with subjects making more errors on near trials (25.6%) than far trials (21.5%).

The mean reaction times (calculated from correct responses) in each condition are shown in Table 4.

The reaction-time data again shows large expertise differences, and also indicates that both Group 1 and Group 2 are faster to respond to temporally coherent material than to temporally randomized material, although the nonmusicians in this case also show a slight sensitivity to temporal structure. A four-way ANOVA (expertise × trial type × pattern separation × temporal structure) performed on transformed reaction-time scores on the different trials produced a significant main effect of expertise [$F(2,21) = 17.2, p < .01$], and planned pairwise comparisons demonstrated that Group 1 and Group 2 were not significantly different [$t(21) = -1.26, p > .10$], but both Group 1 [$t(21) = -5.57, p < .01$] and Group 2 [$t(21) = -4.30, p < .01$] were significantly faster than the nonmusician group. In contrast with the same trial analysis, none of the interactions reached statistical significance. For example, the expertise × temporal structure interaction, which was significant for the same trials, failed to reach significance [$F(2,21) = 1.07, p > .10$]. Despite this, temporal structure did have a larger overall effect on Group 1 (209 msec) and Group 2 (184 msec) than on the nonmusicians (115 msec).

Analysis of Eye-Movement Data

A great deal of data were generated from the eye-movement recordings. On some trials, for example, the nonmusicians were using as many as 15 macrosaccades between the patterns (i.e., large eye movements from one pattern to the other), and as many as 30 fixations to effect their comparisons. To cut down data analysis to manageable proportions, we analyzed the eye-movement behavior only from correct responses to the same trials.

Subjects generally performed the comparisons by progressing from left to right, comparing small passages of music (henceforth termed comparison units) en route. A number of macrosaccades between the patterns (henceforth termed flips) were executed in each trial. A unit was defined as the amount of time that elapsed between macrosaccades. Some units consisted of just one fixation; others contained more than one fixation. The following vari-

Table 3
Error Rates (in Percent) on Different Trials, Experiment 2

Group	Pitch Different				Duration Different			
	Near		Far		Near		Far	
	TC	TR	TC	TR	TC	TR	TC	TR
1	15.0	18.8	16.3	13.8	42.5	41.3	36.3	38.8
2	20.0	16.3	12.5	13.8	35.0	41.3	20.0	32.5
Nonmusicians	22.5	22.5	23.5	21.3	17.5	15.0	11.3	17.5

Note—TC = temporally coherent, TR = temporally randomized.

Table 4
Mean Reaction Times (in Milliseconds)
on Different Trials, Experiment 2

Group	Pitch Different				Duration Different			
	Near		Far		Near		Far	
	TC	TR	TC	TR	TC	TR	TC	TR
1	2,844	2,838	2,867	3,053	2,636	3,002	2,830	3,121
2	3,226	3,328	3,405	3,652	3,509	3,530	3,285	3,544
Nonmusicians	6,088	6,399	6,503	6,681	5,914	5,791	5,785	5,878

Note—TC = temporally coherent, TR = temporally randomized.

ables were extracted for analysis: (1) *Number of flips*, or macrosaccades from one pattern to the other. It was assumed that the number of flips would be roughly inversely proportional to the sizes of units adopted in comparison (i.e., the fewer the flips, the larger the size of the comparison unit). (2) *Unit viewing time* referred to the mean viewing time of all units involved in a comparison for a trial. This can be thought of as being similar to the concept of “gaze duration” in the language-reading literature. (3) *Number of fixations per unit*. There was often just one fixation per unit, but these could be as many as four or five. A mean was taken from all the units involved per trial. (4) *Fixation durations*. A mean fixation duration was calculated from all the fixations per trial.

Finally, the number of “checking” saccades (regressive saccades executed clearly associated with checking procedures) was also aggregated for each subject. In general, however, the subjects heeded the experimenters’ instruction that they should perform the comparisons as quickly as possible, and hence tended not to indulge in extensive checking processes. There were few checking saccades, and they were not statistically analyzed.

Condition means were computed from all the trials within a particular condition, for all subjects. Table 5 presents the means of all the variables for the three groups.

Statistical Analyses

Number of flips. Group 1 clearly used fewer flips than did Group 2, who in turn use fewer flips than did the nonmusicians, with the musician groups using slightly more flips for temporally randomized material than for temporally coherent material. A three-way ANOVA (expertise × pattern separation × temporal structure) performed on the transformed means of number of flips yielded a

main effect of expertise [$F(2,21) = 14.9, p < .01$]. As for the reaction-time analyses, a natural logarithm was used to reduce the disparity in variance between groups for analysis of the eye-movement measures. Planned comparisons made between the three levels of expertise indicated that Group 1 [$t(42) = -5.37, p < .01$] and Group 2 [$t(42) = -3.59, p < .01$] used significantly fewer flips than did the nonmusicians, with a trend toward significance for the Group 1–Group 2 comparison [$t(42) = -2.01, p < .10$]. There was also a significant expertise × temporal structure interaction [$F(2,21) = 4.37, p < .05$], illustrated in Figure 5. None of the other effects approached significance.

Simple main effects were computed for this interaction. Group 1 musicians used more flips for temporally randomized material [$F(1,21) = 19.2, p < .01$], and Group 2 showed a trend in this direction [$F(1,21) = 3.50, p < .10$]. There was no effect of temporal structure for the nonmusicians [$F(1,21) = 0.05, p > .10$].

Mean unit viewing time. Once again, there were large expertise effects for the unit viewing times. A three-way ANOVA (expertise × pattern separation × temporal structure) performed on the transformed means of unit viewing times produced a main effect of expertise [$F(2,21) = 14.3, p < .01$]. Planned comparisons made between the three levels of expertise indicated that Group 1 [$t(21) = -5.26, p < .01$] and Group 2 [$t(21) = -3.56, p < .01$] used significantly shorter unit times than did the nonmusicians. The Group 1–Group 2 comparison approached significance [$t(21) = -1.76, p < .10$]. There was also a main effect of pattern separation [$F(1,21) = 12.1, p < .05$], indicating that subjects used longer unit viewing times in the far trials than in the near trials (means: near, 539 msec; far, 569 msec). None of the other effects reached significance.

Number of fixations per unit. As with the previous analyses, there were large effects of expertise. A three-way ANOVA (expertise × pattern separation × temporal structure) performed on the transformed means of number of fixations per unit produced a main effect of expertise [$F(2,21) = 10.3, p < .01$]. Planned comparisons made between the three levels of expertise indicated that Group 1 [$t(21) = -4.48, p < .01$] and Group 2 [$t(21) = -2.82, p < .05$] used significantly fewer fixations per unit than did the nonmusicians. The Group 1–Group 2 comparison, however, did not reach significance [$t(21) =$

Table 5
Summary Statistics of Eye-Movement Variables on Same Trials

Group	Temporal Structure	Near				Far			
		Number of Flips	Unit Viewing Time	Number of Fixations/Unit	Fixation Duration	Number of Flips	Unit Viewing Time	Number of Fixations/Unit	Fixation Duration
1	Coherent	7.45	422	1.90	220	7.53	467	1.97	231
	Randomized	8.20	423	1.98	213	8.03	475	2.09	228
2	Coherent	9.11	510	2.10	240	8.86	537	2.17	238
	Randomized	9.28	538	2.16	245	9.34	545	2.25	243
Nonmusicians	Coherent	12.46	748	2.54	270	12.38	765	2.59	273
	Randomized	12.40	742	2.46	270	12.53	762	2.60	265

Note—Durations are measured in milliseconds.

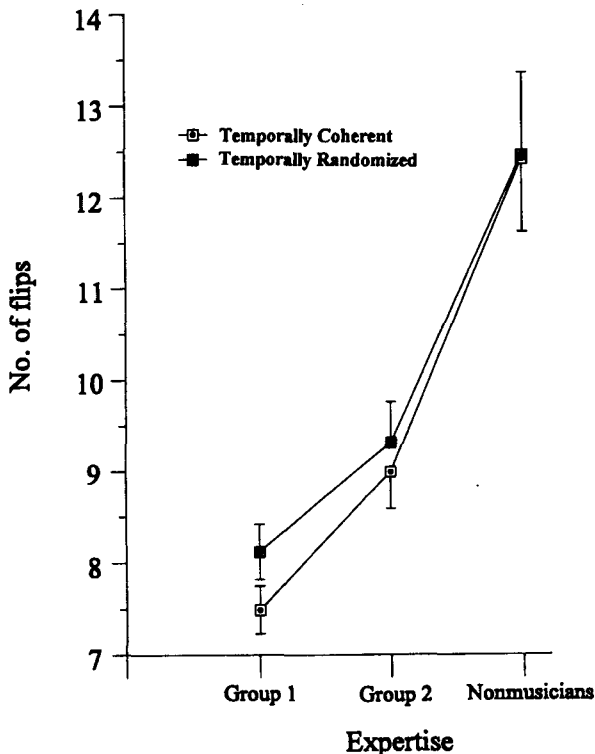


Figure 5. Expertise by temporal structure interaction for number of flips.

$-1.66, p > .10$). There was also a main effect of pattern separation [$F(1,21) = 7.51, p < .05$], indicating that subjects used fewer fixations per unit in the near trials than in the far trials (means: near, 2.19 fixations; far, 2.28 fixations). None of the other effects reached significance.

Fixation durations. Once again, an ANOVA revealed a main effect of expertise [$F(2,21) = 5.46, p < .05$]. Planned comparisons made between the three levels of expertise indicated that Group 1 used significantly briefer fixation durations than did the nonmusicians [$t(21) = -3.28, p < .01$]. The Group 2–nonmusician comparison approached statistical significance [$t(21) = -1.98, p < .10$], and the Group 1–Group 2 comparison did not reach significance [$t(21) = -1.30, p > .10$]. None of the other effects approached significance.

Discussion

The principal aim of this experiment was to investigate the comparison processes associated with the expertise effect observed in the pilot study. In Experiment 2, in common with Experiment 1, there were no reliable differences in the overall accuracy of task performance between the groups but large differences in reaction times for comparisons, with Groups 1 and 2 performing the comparisons much more rapidly than the nonmusicians. The eye-movement measures gave some indication as to the basis of this expertise effect. First, Groups 1 and 2 used fewer flips than the novices in effecting the

comparisons. It is thus assumed that subjects in Groups 1 and 2 used larger units to compare the stimuli in the *same* trials. Second, Groups 1 and 2 required less viewing time than the novices to process the comparison units. Third, Groups 1 and 2 used fewer fixations than the novices to process such comparison units. In sum, the more experienced musicians used larger units to compare the stimuli and processed those units with fewer fixations and in less time. There were also similar, though weaker, differences between Group 1 and Group 2.

As regards the fixation durations used by the groups, it was clear that Group 1 used significantly briefer fixations than the nonmusician group, and there was weaker statistical evidence that Group 2 used briefer fixations than the nonmusician group. This is consistent with previous data documenting an association between reading skill and fixation durations (e.g., Goolsby, 1994a, 1994b). It is noteworthy that the mean fixation durations reported here (around 220 msec for Group 1, 240 msec for Group 2, and 270 msec for the novices) were much shorter than those quoted by Goolsby (around 370 msec for skilled readers and 470 msec for less skilled readers). This difference is presumably due to task demands. The present task requires that subjects scan as quickly as possible, whereas in real music-reading situations, often there are times when the reader can just sit on a note, “biding time” (Goolsby, 1989, p. 288). Nonetheless, Experiment 2 has demonstrated that the association between reading skill and fixation durations is apparent with the nonmotoric pattern-matching task used here.

Experiment 2 also replicated the finding from Experiment 1 that subjects from Groups 1 and 2 make more errors on duration-*different* trials than on pitch-*different* trials, while the nonmusicians show a trend in the opposite direction. As stated earlier, the violation of the space-duration relationship in generating duration differences probably results in the tendency of the musician subjects to overlook the duration “misprints.”

The second aim of the experiment was to determine the effect of randomization of temporal structure on task performance. In common with the pilot study, the reaction-time data in Experiment 2 demonstrated that Group 1 and Group 2 showed more sensitivity to temporal structure than did the nonmusician group. The eye-movement data demonstrated that Group 1 used significantly fewer flips to compare coherent material than to compare randomized material (and Group 2 showed a trend in this direction), while the nonmusicians showed no difference in the number of flips required to compare coherent and randomized material. Hence, Group 1 and Group 2 appear to use slightly larger units to compare the nonrandomized material, although the effect sizes are rather small.

The final aim of the experiment was to determine the effect of pattern separation. This turned out to be rather more difficult to assess, though potentially interesting. Subjects compared the near trials more quickly than the far trials, although more errors were made in the near trials, which would appear to indicate a speed-accuracy trade-

off between the near/far conditions. The eye-movement data showed that subjects used longer unit viewing times in the far trials than in the near trials, and used more fixations to encode those units. The simplest explanation for this pattern of data is that subjects adopted a "speed-not-accuracy" response strategy for the near trials, possibly because they perceived the task to be easier. Alternatively, it is possible that the proximity of the other pattern in the near condition causes reduced unit viewing times by "demanding" an eye movement, since the production of saccades is a balance between an active mechanism for maintaining fixation and processes in peripheral vision demanding a move. This may give rise indirectly to increased error rates in the near condition. What is clear is that there was no evidence that subjects used fewer flips in the comparison process for the far trials, so the pattern-separation manipulation was ineffective in inducing subjects to use appreciably larger units in the comparison process.

GENERAL DISCUSSION

This paper has presented two experiments using a pattern-matching paradigm, which has proven to be a viable tool in the study of expertise in other domains. Both studies have demonstrated clear quantitative and qualitative differences in performance between subjects of different skill levels. The pilot study demonstrated that the paradigm was skill sensitive, that skilled readers were faster to compare temporally coherent material than temporally randomized material, and that experts and novices made different types of error. To investigate the comparison processes underlying these differences, eye-movement recordings were taken as subjects performed the comparisons in Experiment 2. This study produced results broadly comparable to those of the pilot study, Experiment 1. The eye-movement data suggested that more experienced musicians *do* use larger units in the comparison process than do nonmusicians and, moreover, that they take fewer fixations and require less viewing time to process those units. Furthermore, the average fixation duration of the most experienced group was significantly shorter than that of the nonmusicians. Finally, the more experienced groups, but not the nonmusicians, showed some evidence of using larger units to compare temporally coherent material.

We have also carried out a further pattern-matching study using pianists as subjects and piano music as stimuli. Consistent with the findings of the experiments presented here, there was a significant positive correlation between performance in a sight-reading task and speed of comparison at the pattern-matching task. Taken together, these studies suggest that the pattern-matching task is a viable tool for the study of expertise and perceptual processing of music notation.

We feel that three major strengths of the task are that (1) it is highly *adaptable*, in the sense that the task parameters (e.g., number of notes, simultaneous vs. successive presentation) can be readily manipulated to address

different issues; (2) it allows a *direct* test of the experts-process-larger-units hypothesis discussed in the introduction when eye-movement recordings are taken; and (3) it can ultimately be used to delineate the *contents* of the patterns used in the comparison process, and can thus shed light on the nature and composition of the vocabulary of pattern knowledge of skilled music readers. Nevertheless, there are a number of possible objections to the task which we wish to discuss.

First, it might be argued that the ultimate goal of this research is to understand the skills underlying fluent sight-reading (or some other ecologically relevant activity), yet the pattern-matching task is clearly very different from such activities. Therefore, one should perhaps just study sight-reading in its more "natural" state, replete with complex motoric responses and stimulus materials. However, sight-reading, as indicated in the introduction, is an extremely complex psychological activity involving a series of perceptual and motoric processes. Thus, differences in sight-reading performance between two individuals may have multiple causes (e.g., pattern recognition/chunking ability, knowledge of structure, motoric facility, etc.). The purpose of the pattern-matching paradigm is to attempt to *focus* the research to examining rather basic pattern-recognition processes of subjects with different skill levels. In doing so, the paradigm allows a direct test of the chunking hypothesis, which may only be achieved indirectly if the conventional sight-reading task is used. In any case, we would argue that the pattern-matching paradigm should *complement* (rather than substitute for) research using the full sight-reading task. For example, one may predict that patterns that are compared with few flips (i.e., in large units) should be relatively easier to sight-read than should patterns that require many flips.

A second objection to the task is that it is not obvious whether the expertise differences documented here are in the perception of the notes (encoding ability), or in storing the stimuli in a short-term store (memory ability), or in comparing two structures within short-term memory. It is, at present, unclear whether this task can really decipher much about the *stages* of processing. A third possible objection to the experiments reported here is that the manipulations of temporal and pitch structure were rather crude, and the musical stimuli used were musically impoverished. For example, there were no accidentals, rests, dynamics, or articulation markings, and there were no vertical structures present. We acknowledge these limitations, and suggest that the manipulations should preferably be based on music theoretic formulations which could more objectively characterize the degree of coherence of the stimuli.

Indeed, it is noteworthy that the effects of structure on expert performance were, though significant, generally rather weak, and the expertise by temporal (and pitch) structure interaction quite small (or nonexistent). Most expertise studies have reported dramatic reductions in performance when material is scrambled through randomization, to the extent that experts are performing at the

same level as novices. However, Saariluoma (1985) documented a complete absence of an expertise \times structure interaction for speeded perceptual classification tasks using chess positions and chess players, with the experts convincingly outperforming the novices with both structured and randomized material. It is thus possible that in speeded perceptual or search tasks, such as the one used here, effects of structure become less important in governing expert performance, with the present finding supporting Saariluoma (1985) in the musical domain. However, one can plausibly argue that the manipulations made in the present studies were simply rather weak; any sequence in music makes some kind of sense, and the terms "coherent" and "randomized" may suggest a stronger differentiation than actually exists. In this case, the small expertise \times structure interactions may be an artifact of a weak manipulation.

Finally, since the error rates in both experiments were rather high, and in Experiment 2 the novices were the most accurate group, it is reasonable to argue that it becomes more difficult to interpret reaction-time differences between groups than if there were simply no errors at all, or no differences between groups. However, it is difficult to imagine that the large differences in reaction time between groups represents a speed-accuracy tradeoff between groups with such small (and nonsignificant) differences in accuracy (in all studies we have completed). Moreover, the error data produced the interesting, if rather serendipitous, finding that experts and novices appear to make qualitatively different patterns of error, indicating different processing mechanisms for durational information.

In the light of the considerations above, we suggest that future research using this paradigm should use manipulations of structure that are more firmly grounded in music theory. Accuracy could be improved by manipulating the instructions and by providing feedback during the practice and experimental phases of the experiment. If accurate location information is attainable from eye-movement recordings, the contents of the comparison units can of course be inspected for subjects of different skill levels, and fine-grained information about saccade sizes within a unit can be determined. Delineating the content of such chunks, in terms of size and composition, is the goal state of the pattern-matching paradigm. Extended testing of subjects should reveal substantial amounts of data concerning the nature and sizes of units used in the comparison process and illuminate our understanding of the pattern knowledge of music readers.

REFERENCES

- BEAL, A. L. (1985). The skill of recognizing musical structures. *Memory & Cognition*, **5**, 405-412.
- BEAN, K. L. (1938). An experimental approach to the reading of music. *Psychological Monographs*, **50** (Whole No. 50).
- BOWER, G. H., & SPRINGSTON, F. (1970). Pauses as recoding points in letter series. *Journal of Experimental Psychology*, **83**, 421-430.
- CHARNESS, N. (1979). Components of skill in bridge. *Canadian Journal of Psychology*, **33**, 1-6.
- CHASE, W. G., & SIMON, H. A. (1973). Perception in chess. *Cognitive Psychology*, **4**, 55-81.
- CLIFTON, J. V. (1986). *Cognitive components in music reading and sight reading performance*. Unpublished doctoral dissertation, University of Waterloo.
- DE GROOT, A. (1978). *Thought and choice in chess*. The Hague: Mouton.
- DEUTSCH, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, **28**, 381-389.
- DOWLING, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, **14**, 37-40.
- ELLIS, S. H. (1973). *Structure and experience in the matching and reproduction of chess patterns*. Unpublished doctoral dissertation, Carnegie Mellon University.
- ERICSSON, K. A., & SMITH, J. (1991). *Toward a general theory of expertise: Prospects and limits*. Cambridge: Cambridge University Press.
- FASANARO, A. M., SPITALERI, D. L. A., & VALIANI, R. (1990). Dissociation in music reading: A musician affected by alexia without agraphia. *Music Perception*, **7**, 259-272.
- GENTNER, D. R. (1987). Timing of skilled motor performance: Tests of the proportional duration model. *Psychological Review*, **94**, 255-276.
- GOOLSBY, T. (1989). Computer applications to eye movement research in music reading. *Psychomusicology*, **8**, 111-126.
- GOOLSBY, T. W. (1994a). Eye-movement in music reading—Effects of reading ability, notational complexity, and encounters. *Music Perception*, **12**, 77-96.
- GOOLSBY, T. W. (1994b). Profiles of processing—Eye-movements during sightreading. *Music Perception*, **12**, 97-123.
- HALPERN, A. R., & BOWER, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology*, **95**, 31-50.
- HALVERSON, D. L. (1974). *A biometric analysis of eye movement patterns of sight-singers*. Unpublished doctoral dissertation, Ohio State University.
- HANDEL, S. (1973). Temporal segmentation of repeating auditory patterns. *Journal of Experimental Psychology*, **101**, 46-54.
- KINSLER, V., & CARPENTER, R. H. S. (1995). Saccadic eye movements while reading music. *Vision Research*, **35**, 1447-1458.
- MACKENZIE, C. L., & VAN EERD, D. L. (1990). Rhythmic precision in the performance of piano scales: Motor psychophysics and motor programming. In M. Jeannerod (Ed.), *Attention and performance XIII: Motor representation and control* (pp. 375-408). Hillsdale, NJ: Erlbaum.
- MCKEITHEN, K. B., REITMAN, J. S., RUETER, H. H., & HIRTLE, S. C. (1981). Knowledge organization and skill differences in computer programmers. *Cognitive Psychology*, **13**, 307-325.
- REITMAN, J. (1976). Skilled perception in go: Deducing memory structures from interresponse times. *Cognitive Psychology*, **8**, 336-356.
- RESTLE, F. (1972). Serial patterns: The role of phrasing. *Journal of Experimental Psychology*, **92**, 385-390.
- REULEN, J. P. H., MARCUS, J. T., KOOPS, D., DE FRIES, F. R., TIESINGER, G., BOFHUIZEN, K., & BOS, J. E. (1988). Precise recording of eye movements: The IRIS technique. *Medical & Biological Engineering & Computing*, **26**, 20-26.
- SAARILUOMA, P. (1985). Chess players' intake of task-relevant cues. *Memory & Cognition*, **13**, 385-391.
- SALIS, D. L. (1980). Laterality effects with visual perception of musical chords and dot patterns. *Perception & Psychophysics*, **28**, 284-292.
- SALTHOUSE, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, **113**, 345-371.
- SCHMIDT, F. O. (1981). *The eye movement patterns of woodwind instrument performers while sight-reading*. Unpublished doctoral dissertation, Ohio State University.
- SHAFFER, L. H. (1981). Performance of Chopin, Bach, and Bartok: Studies in motor programming. *Cognitive Psychology*, **13**, 326-376.
- SLOBODA, J. A. (1974). *Music reading and prose reading: Some comparisons of underlying perceptual processes*. Unpublished doctoral dissertation, University of London.
- SLOBODA, J. A. (1976a). The effect of item position on the likelihood of identification by inference in prose reading and music reading. *Canadian Journal of Psychology*, **30**, 228-236.
- SLOBODA, J. A. (1976b). Visual perception of musical notation: Regis-

- tering pitch symbols in memory. *Quarterly Journal of Experimental Psychology*, **28**, 1-16.
- SLOBODA, J. A. (1977). Phrase units as determinants of visual processing in music reading. *British Journal of Psychology*, **68**, 117-124.
- SLOBODA, J. A. (1978a). Perception of contour in music reading. *Perception*, **7**, 117-124.
- SLOBODA, J. A. (1978b). The psychology of music reading. *Psychology of Music*, **6**(2), 3-20.
- SLOBODA, J. A. (1981). The uses of space in music notation. *Visible Language*, **25**, 86-110.
- SLOBODA, J. A. (1982). Music performance. In D. Deutsch (Eds.), *The psychology of music* (pp. 476-496). New York: Academic Press.
- SLOBODA, J. A. (1983). The communication of musical metre in piano performance. *Quarterly Journal of Experimental Psychology*, **35A**, 377-396.
- SLOBODA, J. A. (1984). Experimental studies of music reading: A review. *Music Perception*, **2**, 222-236.
- SLOBODA, J. A. (1985a). Expressive skill in two pianists: Style and effectiveness in music performance. *Canadian Journal of Psychology*, **39**, 273-293.
- SLOBODA, J. A. (1985b). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.
- THOMPSON, W. B. (1987). Music sight-reading skill in flute players. *Journal of General Psychology*, **114**, 345-352.
- UNDERWOOD, G., DEIHIM, C., & BATT, V. (1994). Expert performance in solving word puzzles: From retrieval cues to crossword clues. *Applied Cognitive Psychology*, **8**, 531-548.
- VAN NUYS, K., & WEAVER, H. A. (1943). Memory span and visual pauses in reading rhythms and melodies. *Psychological Monographs*, **55**(1), 33-61.
- WOLF, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research*, **5**, 143-171.
- YORK, R. (1952). *An experimental study of vocal music reading using eye movement photography and voice recording*. Unpublished doctoral dissertation, Syracuse University.

NOTES

1. Practical examinations of musicianship are administered by the London-based Associated Board of the Royal Colleges of Music. There are eight levels of difficulty, ranging from Grade I (beginners) to Grade VIII (required by some U.K. institutions for entrance into studying music).
2. The materials were composed with the help of a music student, Kieron O'Rierdon, at the Department of Music, Nottingham.
3. Written by Bob Kentridge, Department of Psychology, University of Durham.

(Manuscript received April 6, 1995;
revision accepted for publication May 6, 1996.)