Cognitive Mechanisms in Temporally Controlled Rhythm Reading: Evidence From Eye Movements

Marjaana Puurtinen
University of Turku, Turku, Finland

Erkki Huovinen
Royal College of Music in Stockholm, Stockholm, Sweden

Anna-Kaisa Ylitalo
Natural Resources Institute Finland, Helsinki, Finland

Music-reading research has not yet fully grasped the variety and roles of different cognitive mechanisms that underlie visual processing of music notation; instead, studies have often explored one factor at a time. Based on prior research, we identified three possible cognitive mechanisms regarding visual processing during music reading: *symbol comprehension, visual anticipation,* and *symbol performance demands*. We also summed up the eye-movement indicators of each mechanism. We then asked which of the three cognitive mechanisms were needed to explain how note symbols are visually processed during temporally controlled rhythm reading. In our eye-tracking study, twenty-nine participants performed simple rhythm-tapping tasks, in which the relative complexity of consecutive rhythm symbols was systematically varied. Eye-time span (i.e., "looking ahead") and first-pass fixation time at target symbols were analyzed with linear mixed-effects modeling. As a result, the mechanisms *symbol comprehension* and *visual anticipation* found support in our empirical data, whereas evidence for *symbol performance demands* was more ambiguous. Future studies could continue from here by exploring the interplay of these and other possible mechanisms; in general, we argue that music-reading research should begin to emphasize the systematic creating and testing of cognitive models of eye movements in music reading.

Received: September 29, 2021, accepted December 6, 2022.

Key words: eye movements, music reading, notation, rhythm, timing

Music reading has some features that set quite distinctive cognitive requirements on the reading situation. A key characteristic is that music-reading tasks are most often temporally constrained, and the amount of time the reader can spend on processing individual items is therefore limited (e.g., Huovinen et al., 2018; Kinsler & Carpenter, 1995; Puurtinen, 2018a). So far, however, there is little to no systematic testing and developing of cognitive models of music reading: indeed, prior studies seem to have focused more on the music readers' fixation behavior, and not on the mechanisms underlying this activity (for a notable exception, see Kinsler & Carpenter, 1995). To begin to fill this gap, we summarize prior research by three possible cognitive mechanims addressing how visual processing could be regulated during temporally controlled rhythm reading, and then conduct an experiment where all three mechanisms are simultaneously explored. Briefly, we examine whether the fixation behavior during rhythm reading is regulated by the demands set by currently read symbols (*symbol comprehension*), upcoming symbols (*visual anticipation*), or symbols that are currently performed (*symbol performance demands*).

Basic Characteristics of Eye Movements During Music Reading

The eye-tracking methodology offers a non-intrusive way to study the course of visual processing of written music. According to Just and Carpenter's (1980) "eye-mind hypothesis," a reader's cognitive processing is typically closely related to the currently fixated visual target; thus recordings of eye movements can be used for making inferences about the cognitive processing that underlies the execution of visual tasks. During music reading, and especially when previously unread note symbols are performed in a given temporal framework, this link between fixation targets and the cognitive processing of corresponding information can be said to be rather tight and also corroborated by the music performance itself (Puurtinen, 2018b).

Although published studies about eye movements in music reading are slowly increasing, they are still not

Music Perception, volume 40, issue 3, pp. 237–252, issn 0730-7829, electronic issn 1533-8312. DOI: https://doi.org/10.1525/mp.2023.40.3.237

numerous. Nevertheless, some general results are already supported by relatively convincing evidence (for more extensive reviews about research findings and methods in this field, see Madell & Hébert, 2008; Puurtinen, 2018b; Sheridan et al., 2020). Much of this evidence concerns how music readers allocate their *fixations*, or short "stops" lasting some hundreds of milliseconds, to different parts of the musical score to gather and process visual information. To begin with, music readers appear to fixate most of the available note symbols (see Goolsby, 1994a; Kinsler & Carpenter, 1995; Penttinen & Huovinen, 2011; Truitt et al., 1997; Wurtz et al., 2009). In practice, they target their *fovea*, the area of accurate vision, to a symbol in question. Typically more than one notehead fits in the fovea; thus, composite note symbols, such as a set of two eighth notes beamed together, can be processed with one fixation and as meaningful chunks or patterns (e.g., Goolsby, 1994a; Kinsler & Carpenter, 1995; Sheridan, et al., 2020). The identification of note symbols, or chunks of them, seems to become faster and fixation durations shorter with increasing music-reading skill or familiarity with the musical material (e.g., Goolsby, 1994b; Kinsler & Carpenter, 1995; Maturi & Sheridan, 2020; Penttinen & Huovinen, 2011; Penttinen et al., 2015; Polanka, 1995; Waters & Underwood, 1998). Some such findings may have been simply due to more experienced readers' choice of faster performance tempi (in studies where tempo was not controlled; see Puurtinen, 2018b), but another explanation is that skilled readers may use their highly developed and automated ability to process composite note symbols holistically and are perhaps less distracted by the surrounding symbols (Wong & Gauthier, 2010, 2012).

While the fovea covers the area of a reader's accurate vision, it does not cover *all* visual information available to the reader. The *perceptual span* is the visual area accessible from a given fixation; it includes the fovea but also extends beyond it into the visual symbols the reader sees as blurred (e.g., Rayner, 2009; Sheridan et al., 2020; Sloboda, 1974). Though it is known from text reading that the information within the perceptual span may be used to plan, for instance, the location for the next fixation (Rayner, 2009), the size and use of the perceptual span has only rarely been studied in music reading with modern eye trackers and during actual musical performances. The scarce available evidence suggests that the perceptual span may encompass 3–5 composite note symbols to the right of the fixated symbol (Gilman & Underwood, 2003; Truitt et al., 1997), and that it may be affected by practice (Burman & Booth, 2009; but see Rosemann et al., 2016). Nevertheless, it is clear that during one single fixation, a music reader has access to visual information not just regarding the note symbol currently being fixated, but also regarding the symbols around it.

As mentioned above, a key feature of music reading is that it typically proceeds in a given tempo. Thus, music readers need to reserve enough time for the processing of each musical symbol in order to perform flawlessly even when working with unfamiliar material. Decoding the note symbols, storing them in working memory, and transforming them into a motor response should all fit within the time period between when the target note first is visible in the reader's perceptual span and when a corresponding performance is initiated; this interval also typically includes a fixation on target (see Kinsler & Carpenter, 1995; see also Laubrock & Kliegl, 2015). Most likely the reader simultaneously also needs to monitor the timing and quality of the performance (Penttinen et al., 2015; Rosemann et al., 2016). Given the temporal constraints of typical performance and practice situations, music reading requires a constant balancing between the different processing requirements of the available symbols. Indeed, it represents a kind of a zero-sum game of time use: if more time is spent on processing one symbol, this "additional" time is necessarily away from processing other symbols (Puurtinen 2018a; see also Chitalkina et al., 2021; Hadley et al., 2018).

To ensure that there is enough time for processing each symbol so that the performance can be flawless, music readers maintain their gaze slightly ahead of the currently performed musical symbols (for reviews about the "looking ahead," see Huovinen et al., 2018; Perra et al., 2021; Puurtinen, 2018a). The distance between the point of gaze and point of performance has been called the *eye-hand span* in instrumental performances (e.g., Madell & Hébert, 2008; about eye-hand span in typing, see, e.g., Inhoff & Wang, 1992) and *eye-voice span* in singing (Huovinen et al., 2021; about eye-voice span in oral reading, see, e.g., Laubrock & Kliegl, 2015). In music-reading studies, researchers have often compared the timing or location of a reader's fixation to the timing or location of a concurrent (or near-concurrent) performance action, such as a keypress. Alternatively, they have calculated the time between fixating a target and its later performance (about different ways to calculate the eye-hand span, see Huovinen et al., 2018). More recently, Huovinen and colleagues (2018) reconceptualized the span by relating the point of gaze to the current location in the "metrical time" of the music (instead of the time of a performance action). They called this modified measure the *eye-time span*

(see also Chitalkina et al., 2021). Despite these different conceptualizations of "looking ahead," studies have reported the music readers' gaze to remain, on average, only around 1–2 seconds ahead of the performed music (Chitalkina et al., 2021; Huovinen et al., 2018; Huovinen et al., 2021; Lim et al., 2019; Penttinen et al., 2015; Rosemann et al., 2016; these studies controlled for performance tempi in their experiments). This may well be due to limitations of working memory capacity: proceeding too far from the point of performance might require storing too much information in working memory (Kinsler & Carpenter, 1995; Rayner & Pollatsek, 1997).

One standard methodological approach in many music-reading studies has been to select symbol-sized or larger segments of interest, and then study group- or task-based differences in the visual processing of these segments (e.g., Ahken et al., 2012; Drai-Zerbib et al., 2012; Huovinen et al., 2018; Penttinen et al., 2015). In a methodological sense, however, such study designs might overlook the fact that music reading consists in continuous processing and executing of a *string* of symbols. Indeed, since several symbols at a time are visible to the performer within the fovea and within the perceptual span, the visual processing of a given target symbol may well be affected by what lies around it. Likewise, the zero-sum game of time use in music reading suggests that fixation time allocated to the surrounding symbols may affect the time available for processing a target symbol. Recently, some studies have began to take these aspects into account (e.g., Chitalkina et al., 2021; Hadley et al., 2018; Huovinen et al., 2021). We believe, however, that music-reading studies should also go beyond observing and describing eye-movement patterns, and attempt to grasp the mechanims underlying the observed patterns (e.g., Goolsby, 1994a). Thus, in the following, we summarize previous research in terms of three potential cognitive mechanims for the regulation of fixation behavior during music reading. We also consider how these three mechanisms could be manifested on the eye-movement level, and finally, put them to test in our empirical study.

Three Possible Cognitive Mechanisms for Regulation of Visual Processing During Music Reading

Most prior music-reading studies have applied musical stimuli that did not differentiate between rhythmic, melodic, and/or harmonic characteristics of the performed music; they typically also allowed participants to perform in different tempi. Thus, the confounding of factors affecting the reading processes has made it difficult to convincingly isolate the effects due to different musical parameters (Puurtinen, 2018b). Some studies have focused solely on melodic features of simple melodies, simplifying the rhythmic features in performance tasks (e.g., Huovinen et al., 2018; Penttinen & Huovinen, 2011), but we are only aware of one study by Kinsler and Carpenter (1995) that suppressed melodic information in the reading of rhythm notation. Considering that the timing of each motor response in music reading is dictated by rhythm information, focusing specifically on rhythm reading seems a promising step toward unraveling some of the basic factors that might regulate the visual processing of music notation in temporally controlled performances. Thus, in our empirical study we will focus on rhythm reading. However, since studies about that particular music-reading task are scarce, we turn to other music-reading studies—especially ones that also applied simple musical stimuli—when outlining possible cognitive mechanisms for regulation of eye movements during rhythm reading.

Below, we present these three cognitive mechanims: *symbol comprehension, visual anticipation*, and *symbol performance demands*. These mechanisms are not mutually exclusive, but approach the music-reading process from different standpoints. In addition, since the mechanisms have not been studied together in any one quantitative study before, their interplay is still unexplored. In our summary of prior research, we focus on how each of the three mechanims could be expected to be reflected by two music-reading measures. The first is the Eye-Time Span (ETS), i.e., the temporal distance between the fixated note symbol and concurrent location in the metrical time of the music (Huovinen et al., 2018). This modification of the eye-hand span measure enables the study of reading rests among other music symbols, since it does not use the performance of the symbol as a point of reference; hence, we choose this measure over other measures for "looking ahead." The second measure is First-Pass Fixation Time (FPFT), i.e., the summed duration of all fixations targeting a symbol during its first encounter (Hyönä et al., 2003). This measure is a standard in text-reading research, and is generally considered to reflect immediate stimulus processing effects (Hyönä et al., 2003). Recently, it has gradually become more often used in music-reading research, too, though that field tends to be less consistent in the applied fixation measures (see Puurtinen, 2018b). At least in simple, temporally controlled music reading tasks, reading appears to rely on first-pass reading (Penttinen & Huovinen, 2011), and this measure seems therefore useful for our purposes.

Our interest here lies in how the two measures are affected by the *relative complexity* of consecutive rhythm symbols. Kinsler and Carpenter (1995) observed that the relative complexity of a rhythm symbol is dependent on its meaning. We wish to add to this that in Western music notation also the visual complexity of (composite) rhythm symbols varies, and that it is typical for more complex visual configurations to signal more complex performance actions. In practice, in our experiment we will compare the visual processing of quarter notes, quarter rests, and sets of four beamed 16th notes. The latter composite symbol will here be considered to be relatively more complex than the two others, both in terms of the associated performance task (executing a patterned string of sounds instead of just one sound or no sound at all) as well as its visual appearance (considering the visual extent of the symbol, for instance). Including quarter rests among the symbols of interest enables us to check whether the lack of a motor response affects the reading process, and if the treatment of the quarter-rest symbol differs from the that of the quarter-note symbol.

We now explain the first cognitive mechanism. Relying on Just and Carpenter's (1980) "eye-mind hypothesis," Kinsler and Carpenter (1995) presented a cognitive model suggesting that music reading is primarily regulated by the relative complexity of individual note symbols. According to their view, more difficult note symbols simply require more time for identification and planning for a motor response. The authors came to this conclusion after recording four participants' repeated performances of simple rhythm tapping. Later, Penttinen and Huovinen (2011) reported that music-reading novices spent the most FPFT on the second note symbol of a large (and for them, more complex) interval within an otherwise stepwise melody. We will use the term *symbol comprehension* to denote the cognitive mechanism whereby the complexity of a symbol is directly reflected in readers' foveal processing—more specifically, in the time needed *at* the target symbol for decoding it and for planning appropriate motor responses (Just & Carpenter, 1980; Kinsler & Carpenter, 1995). In practice, then, a relatively more complex symbol should gain more FPFT than a less complex symbol. Considering the zero-sum game of time use, however, such effects on FPFT at a target symbol should also be moderated by the complexity of symbols both *before* and *after* the target (see Penttinen & Huovinen, 2011). In this case, simpler symbols around the target could enable the reader to spend more FPFT on the target, while complex nearby symbols might hinder this possibility. This mechanism would thus be primarily apparent in the FPFT measure.

The second suggested mechanism, *visual anticipation*, takes as its starting point music readers' need to plan and prepare for upcoming performance actions and thus their "looking ahead" behavior. Huovinen and colleagues (2018, Experiment 1) have shown that in simple sight-reading tasks, skilled readers may fixate relatively more complex and/or visually salient symbols as early on as possible after becoming aware of them—thus locally lengthening their ETS (see also Gilman & Underwood, 2003). However, in Experiment 2 by Huovinen and colleagues (2018), the upcoming salient symbol only precipitated fixations to the symbols immediately preceding the salient symbol: in this case, the authors suggested that information within the perceptual span about upcoming complexity might have prompted the music reader to proceed faster ahead in reading the notation (toward the complexity). Chitalkina and colleagues (2021) reported a similar observation in their study, where unexpected melodic changes were embedded into familiar melodies. Adapting these observations to our present context, visual anticipation would thus become manifest in a target symbol's lengthened ETS either driven by the complexity of this symbol itself (Huovinen et al., 2018) and/or by the complexity of the *following* symbol(s) that the reader is rushing toward (Chitalkina et al., 2021; Huovinen et al., 2018). Notice that such regulation by anticipation could itself serve the first mechanism, symbol comprehension, by helping to allocate as much fixation time on complex symbols as possible. However, this need not be so: early viewing of complex symbols, or the ones preceding them, also maximizes the time available from initiating the visual processing of a symbol to producing the corresponding sound on an instrument, even if not all of this time is used in actually fixating the symbol in question. Thus, as indeed reported by Chitalkina and colleagues (2021), a lenghtened ETS and a longer FPFT might not coincide on the same symbol. This further supports our choice of keeping symbol comprehension and visual anticipation at this point as separate cognitive mechanims with their own eye-movement indicators.

Finally, a third possible cognitive mechanism builds on the multitasking nature of the music-reading task (Puurtinen, 2018a; see also Penttinen et al., 2015; Salvucci & Taatgen, 2011). While looking ahead from the current point of performance, music readers are forced to fixate new note symbols while still performing previous ones. Thus, we will consider the possibility that not (just) the complexity of the viewed symbols, but the complexity of the currently performed ones might affect how the new symbols can be visually processed. Such a possibility has, indeed, been suggested by some

TABLE 1. *Summary of Three Possible Cognitive Mechanisms Governing Visual Processing During Temporally Controlled Music Reading, and Their Associated Eye-Time Span and First-Pass Fixation Time Effects*

Cognitive mechanism	Eye-Time Span (ETS)	First-Pass Fixation Time (FPFT)
Symbol comprehension	–	Complexity of a target symbol increases FPFT at the target symbol. The effect might be mediated by the relative complexity of surrounding symbols (Penttinen & Huovinen 2011; see also Kinsler & Carpenter, 1995).
Visual anticipation	Complexity of target symbol and/or complexity of symbols *after* the target symbol increases the ETS at the target symbol (Chitalkina et al., 2021; Huovinen et al., 2018).	–
Symbol performance demands	Complexity of note symbols *before* a target symbol decreases ETS at the target symbol (see Penttinen et al. 2015 for a related eye-hand span finding).	Complexity of symbols *before* a target symbol increases FPFT on the target symbol (Chitalkina et al., 2021); but note a reported contrary finding (Hadley et al., 2018).

empirical studies that measured the eye-hand span (Penttinen et al., 2015; Rosemann et al., 2016; in typing, see Inhoff & Wang, 1992). Our third suggested mechanism, *symbol performance demands*, thus reflects the cognitive challenges presented by the performing of one symbol (or group of symbols) while visually processing another. Assuming the length of "looking ahead" in music reading to be around 1–2 seconds, or a couple of beats, the performance of more complex symbols in the notation *before* a target symbol might interfere on the visual processing of the target, increasing FPFT on the target symbol. Indeed, in Chitalkina and colleagues' (2021) study, the performance of unexpected changes in a melody did increase FPFT on the note symbols that followed the complex ones; but in Hadley and colleagues' (2018) study the contrary was observed. However, as with symbol comprehension, and due to the zero-sum game of time use, we need to note that the increase of FPFT due to a complex symbol before the target might be mediated by the relative complexity of all symbols visible to the performer. This might give at least a partial explanation for the above-mentioned contradicting findings, since the experimental conditions of the two studies were not exactly alike. Notably, this third mechanism would be against Kinsler and Carpenter (1995), who separated the processing of a symbol from its execution, assuming that after a symbol has been stored in a buffer, the visual system can initiate a new fixation, and that the part of the cognitive system responsible for visual processing is not affected by simultaneous motor executions. Also, the threaded cognition framework of Salvucci and Taatgen (2011) suggests that the visual process and the motor performance would not themselves compete of the same cognitive resources. To our knowledge, however, this issue has not been systematically addressed in eye-tracking studies of music reading.

Above, we have formulated three cognitive mechanisms that could govern the visual processing in simple music-reading tasks, and suggested how the mechanisms might manifest themselves in music readers' ETS and/or FPFT (for a summary, see Table 1). We now ask: *Which of the three cognitive mechanisms are needed to explain how rhythm symbols are visually processed during temporally controlled rhythm reading?*

Method

PARTICIPANTS

Forty-three adult volunteer participants took part in the study. Thirty-seven were Finnish BA (education) students (27 female, 10 male; M = 23.4 years, SD = 3.7 years), recruited from a music course. Participation was open to all course members, irrespective of musical background, and compensated with lunch vouchers. Six trained music educators from an institute of higher education (demographics omitted for ensuring participants' anonymity) were later invited to complement the data set. As some music-reading studies have been known to suffer from heterogeneity of their data (see Puurtinen, 2018b), only temporally stable and correct performances, combined with quality eye-movement data, were included in the analyses. The final data set thus consisted of data from 29 participants; the exclusion criteria and details about the final data set are described in "Data Analysis" below.

There was variability in participants' musical background; we therefore divided the participants into two skill-based groups in order to control for possible effects of expertise in our statistical analyses. Based on their answers to a background questionnaire, "novices"

(n = 11): (A) reported not practicing music at the time of the measurement, (B) did not report training or current experience on any musical instruments, and (C) estimated their skill of sight reading simple rhythm notation with one of the four lowest choices on a six-point scale. The music educators and all other student participants were categorized as "musicians" (n = 18). These student participants were typically: (A) practicing music at the time of the measurement, (B) mentioned prior instrumental training (sometimes on several instruments), and (C) estimated their skill of sight reading simple rhythm notation with one of the four highest choices on a six-point scale. Some individual students satisfied these criteria only in part (e.g., four gave lower ratings for the sight-reading skill), but their musical background was still considered more extensive than those in the "novice" group.

STIMULI

Target Symbols and Target Patterns

We selected three types of *target symbols*—quarter notes, quarter rests, and sets of four 16th notes beamed together. The last of these was assumed to be the most complex symbol, both in terms of its visual appearance and with respect to the required motor response. Target symbols were embedded in larger *target patterns* with the duration of four quarter notes, placing the target symbol on the third beat (see Table 2). The target symbols were thus preceded by two beats of either "sparse" (quarter notes) or "dense" rhythmic material (16th notes) and likewise followed by "sparse" or "dense" material. In this way, the rhythmic contexts before and after the target were varied systematically. Following the same line of thought as with the target symbols, sparse contexts were assumed to be less complex and dense contexts more complex with respect to both their visual appearance and their motor requirements.

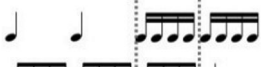
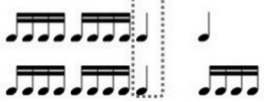
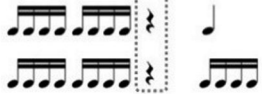
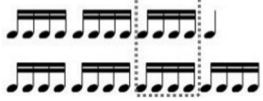
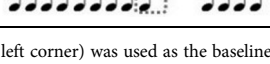
Complete Rhythm Tasks

The target patterns were hidden within longer rhythmic tapping tasks. Four tasks were composed, each consisting of four consecutive one-line staves typical of rhythmic notation, and each including six target patterns (see Figure 1). Tasks were written in 4/4 time. The second or third staff lines of the tasks included a two-bar long rest and no target patterns. Each of the 12 target patterns (see Table 2) came to be used twice. The target patterns were placed in the tapping tasks by alternately beginning on the third and ninth or the fifth and eleventh beats of the staff line, using every third area for patterns involving a given target symbol. Running this way consecutively through tasks A and B, and continuing through tasks C and D for another complete set of target patterns, each target symbol (see Table 2) thus appeared four times on the first beat of the bar and four times on the third beat. All in all, the stimulus design had the structure of 3 (target symbols) x 2 (preceding context) x 2 (following context) x 2 (metrical locations). Finally, the rhythm tasks were completed with filler rhythm symbols. In addition to the experimental tasks, we created two practice tasks consisting of the same rhythmic symbols in similar but non-systematic arrangements. The tasks were notated using Sibelius music notation software, to be shown on a computer screen with a bar width of 84.5 mm (3.33 in).

APPARATUS

Eye movements were recorded with a Tobii T60XL Eye Tracker with a 60 Hz recording frequency and screen resolution of 1,920 x 1,200. The tapping tasks were carried out using a HandSonic HPD-10 hand percussion pad, placed on a table between the participant and the eye tracker. No chin rest was used; the tapping position (with one arm in a 90-degree angle in front of the

TABLE 2. *Composition of Target Patterns: Rhythmic Contexts for Three Types of Target Symbols*

Rhythmic context		Target symbol (marked with a rectangle)		
preceding	following	Quarter note	Quarter rest	16th notes
sparse	sparse			
sparse	dense			
dense	sparse			
dense	dense			

**Note*: The four-quarter-note pattern (upper left corner) was used as the baseline in the analyses.

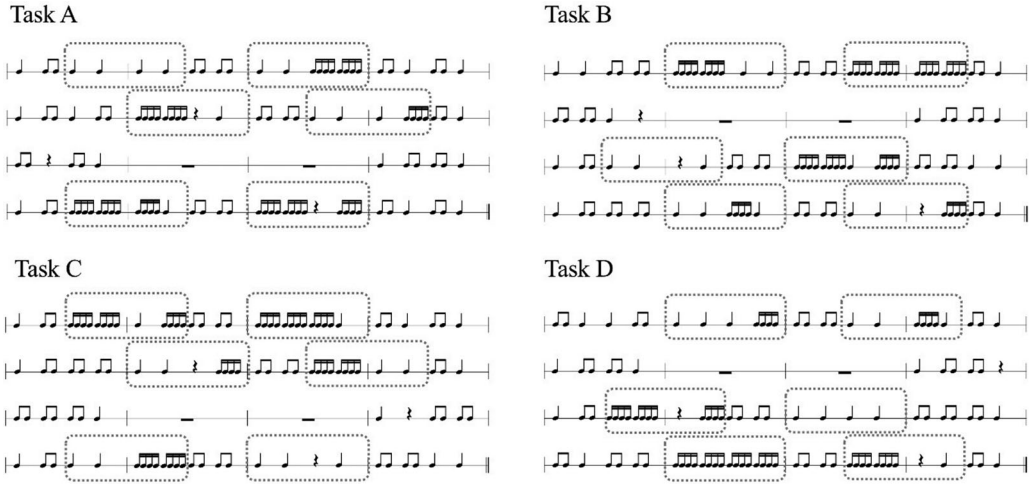


FIGURE 1. The four rhythm tasks applied in the experimental trials. The rectangles have been added here to indicate the locations of the target patterns presented in Table 2.

participant) aided the participant in remaining in position throughout the recording. The rhythm pad was connected to a separate laptop via a MIDI interface, thus recording the performances in MIDI format directly into Logic Pro software. External speakers were used for monitoring the tapping as well as the metronome click (provided by the Logic Pro software).

PROCEDURE

Student participants were invited to perform the tasks twice with a long time gap between the sessions (September and February; for data handling and control for measurement effects, see Data Analysis, below). This allowed us to include more observations per participant in our analyses. The music educators, who were later invited to complement the data set, took part in one measurement session (February) and thus performed the tasks once. Each participant was tested individually in the laboratory by the first author.

In the measurement session, a participant was first introduced to the setup and invited to try out the rhythm pad. The participant could choose whether to use the left-hand or right-hand index finger for tapping. After a 5-point calibration procedure, a practice trial began with written instructions presented on a computer screen. These informed the participant about the following: First, a cross marking the location of the beginning of a notated rhythm task (at the top left corner) would appear on the screen for two metronome beats and, on its disappearance, the task would appear. Second, on the appearance of the rhythm task, the participant should wait for two metronome beats and then start tapping in time with the metronome, continuing through the whole task irrespective of possible performance errors. (Some of the participants delayed their start by one or two beats, but it was later confirmed from the eye-movement recordings that no participant fixated on any of the target symbols during these metronome beats.) After completing the first task, a new cross and a new task appeared. A metronome was set at 50 beats per minute and it was ticking throughout each recording. The experimenter (first author, a musician) changed each slide on the screen on the metronome beats.

After the practice, the participant was allowed to ask for clarifications about the protocol: the experimenter corrected any misunderstanding concerning the instructions, but did not advise on performing the rhythms. The main experiment started with a new calibration, and a similar set of written instructions as during practice. The participant then performed one of the four experimental trials (see Figure 1). The tasks were presented to a participant in one of the following orders: A-B-C-D, B-A-D-C, C-D-A-B, or D-C-B-A, rotating the presentation orders between successive participants. In the first measurement session, the participant also signed a consent form prior to the task and filled in a questionnaire on his/her musical background after the task had been completed.

DATA ANALYSIS

Performance Data Analysis

Despite the simplicity of the musical task, some participants still committed temporal and performance

errors. Of the original 320 rhythm task performances, we first excluded 83 performances whose total duration differed more than two quarter beats from the ideal total performance duration; these performances typically contained a significant amount of performance errors. Next, based on audio recordings created from the MIDI data, any local temporal instabilities and performance errors in the rhythm tasks were identified by a research assistant (with a conservatory degree). For data from a participant's measurement session to be included in the analysis, we required the session to involve at least two successful target-pattern performances. Finally, since each staff could include two target patterns (see Figure 1), all data until the first performance error on each staff were included in the final analyses (similarly to the data handling of Truitt and colleagues, 1997). This enabled including data at the first target symbol of the staff, even though there would be an error before the second target symbol on the same staff.

Eye-Movement Data Analysis

The analyses of eye-movements proceeded in five steps. First, each participant's eye-movement recordings were examined for quality in the visualization and gaze replay tools of Tobii Pro Studio software, using the Tobii Fixation Filter and averaged data from both eyes. Two participants were excluded from further analyses due to a large amount of missing eye-movement data.

Second, eye movements were allocated spatially and temporally to the four staves. The spatial distance between two staves was 124 pixels. The lower bound of fixations to be counted for each staff was set to 70 pixels below the line in order to include also those fixation locations that had dropped a bit below the line due to temporary calibration errors. We also noticed that fixations targeting the lowest staff line were in some cases (due to calibration error on the y-axis) located below (and outside) the stimulus image, data outputs showing that no fixations had landed on the target. Thus, all target symbols missing a fixation were rechecked in the visualization tool of Tobii Pro Studio by two of the authors to ensure whether a target symbol should indeed gain 0 ms as the FPFT, or whether the information should be coded as missing.

Third, Areas of Interest (AOI) were drawn around each target symbol (quarter note, quarter rest, or a set of 16th notes). AOI borders were placed exactly between two note symbols (i.e., between the rightmost pixel of note symbol A and the leftmost pixel of the following note symbol B). However, if there was a barline within the target area (when the target symbol occurred on the first beat of a bar), the AOI border was set at the barline. The width of an AOI varied from 47 to 114 pixels. Our choice of drawing AOIs based on the stimulus features (instead of making them similar in size) was based on: (A) the earlier findings of music-reading studies that (composite) note symbols are treated as meaningful visual units and fixations therefore typically land on (or near) them (e.g., Penttinen 2018b), as well as (B) the general recommendation that "each AOI should cover an area with homogeneous semantics, and the semantics should be founded in [sic] the rationale behind your experimental design" (Holmqvist et al., 2011, p. 188). The First-Pass Fixation Time (FPFT) for the particular target symbol consisted of the sum of fixation durations from the first fixation landing on an AOI until the first fixation landing outside of it. The FPFTs in the data set consisted of 1–4 fixations, but most often (in 78.4% of the cases) of only one. Performers tended to fixate on most of the target symbols: 833 (95.3%) out of 874 correctly performed target symbols received a fixation.

Fourth, we synchronized the eye-movement recording with the metrical time (given by the external metronome) by using the keypress timestamps of the change of slides on the eye-tracker monitor. The change of slides were performed manually by the experimenter at metronome beats. Each of the four tappings tasks (see Figure 1) had a duration of 64 beats (or, given the tempo, 76.8 s). Thus, the first beat onset of one tapping task should have occurred exactly 64 beats before the experimenter switched off the slide showing the task. This manual synchronization was rather accurate: according to 308 pairs of timestamps ideally produced 2400 ms apart, the 95% confidence interval for the experimenter was (2411 ms, 2426 ms).

Fifth, we defined a suitable measure of "looking ahead" in the musical notation. As explained in the introduction, we described the three cognitive mechanisms in terms of the Eye-Time Span (ETS), but preliminary analyses of fixation landing positions urged us to adjust the actual measure used in the analysis. As shown in Figure 2, most of the fixations targeting quarter notes and quarter rests landed close to the center of the single symbol, but for the horizontally wider 16th-note patterns, the gaze tended to land 20–30 pixels to the right from the center of the first note head. This was, in our layout, approximately the location of the second note head of the composite symbol.

In order not to unduly bias the span measurements in favor of the 16th-note pattern, we decided to refrain from using the ETS measure, in which the "front end" of the span exactly tracks the location of the first

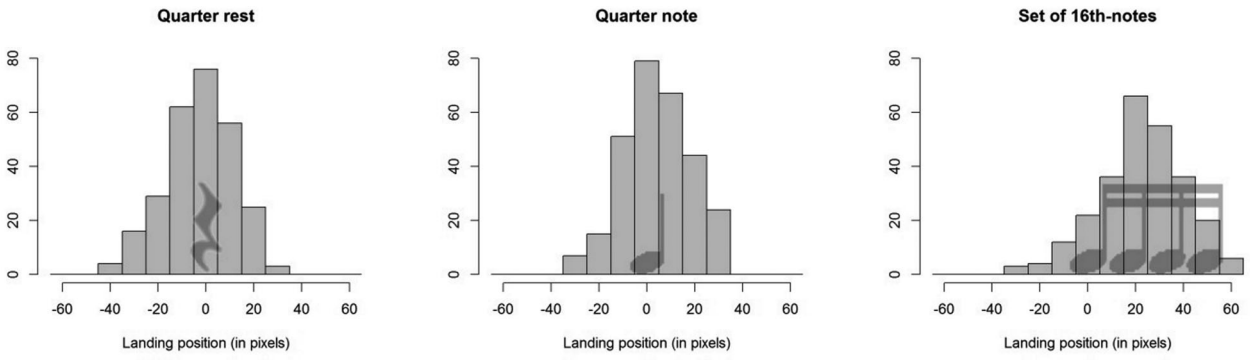


FIGURE 2. Histograms of the landing positions for first fixations (N = 818) with respect to the visual center of the first note head (or rest symbol) appearing on the beat (marked with "0" on the x-axes) for three types of target symbols. The difference in the landing positions resulted in our use of an adjusted Eye-Time Span measure, the Beat-Specific Eye-Time Span (BETS).

fixation targeting the symbol (for more information and illustrations about eye-hand and eye-time span measures, see Huovinen et al., 2018). We thus needed a simpler measure comparable with Furneaux and Land's (1999) "time index" or Rosemann and colleagues' (2016) equivalent "eye-hand span in latency"—both of which record the time lag between the first fixation on a target and its instrumental execution. In order to handle rests in the notation (for which there is no "hand" action) we could not call our measure eye-hand span; thus we defined a measure that will be called *Beat-Specific Eye-Time Span* (BETS).

At the time of first glancing at a given AOI, BETS simply measures the metrical distance between the current point of time of the music and the beat onset corresponding to the fixated AOI. Here, the time of the music is counted in beats and decimal parts thereof. We may demonstrate this by stimulus A in Figure 1 in which the first target symbol was the quarter note beginning the second bar. Glancing at the AOI of this target note, say, exactly at the time point between the previous two beats in the music would yield a BETS value of 1.5 beats. For better comparability with the above-mentioned eye-hand span measures (Furneaux & Land, 1999; Rosemann et al., 2016), we further converted such measurements from metrical time (beats) to absolute time (milliseconds), given the tempo of the performance at 50 bpm. In our example, the measurement of 1.5 beats would thus be converted to 1800 ms.

Final Data Set and Statistical Modeling

As mentioned above, data from 29 participants fulfilled the above criteria with respect to performance and eye-movement data quality (with 13–45 acceptable target pattern performances from each). Two BETS measurements longer than 4 beats (4800 ms), eleven negative spans, and one FPFT of only 33 ms long (excluding this, the minimum FPFT was 117 ms) were regarded as outliers and excluded from the data set. The final data set thus consisted of 818 target observations, derived from 69 task performances by 19 students from the 1st measurement, 88 performances by 23 students from the 2nd measurement, and 20 performances by five music educators. The four rhythm tasks (see Figure 1) were represented by approximately equal numbers of performances (45, 46, 42, and 44 performances for rhythm tasks A–D, respectively).

Linear mixed-effects modeling was chosen as an analysis method due to correlated observations. The analyses were carried out in R (R Core Team, 2021, version 4.1.2) using the package lme4 (Bates et al., 2015). For visualisation we used the emmeans (Lenth, 2022) and ggplot2 (Wickham, 2016) packages. Post hoc pairwise comparisons were carried out with the emmeans package, using the Tukey method for adjusting p values.

FPFT and BETS were considered as dependent variables. The analysis regarding FPFT was carried out both in original and in log-transformed scale due to skewness, and since there was no discrepancy between the respective results, we report the ones in a non-transformed scale. The factor *target symbol* (quarter rest, quarter note, 16th-note pattern) and the factors *preceding context* and *following context* (both with two values, sparse or dense) were considered as fixed effects. In addition, we included the three-way interaction term consisting of all these factors (and hence all two-way interactions). Control variables concerning the experimental setup, *expertise* ("novice" or "musician"), *measurement* (1 or 2 for student participants; 1 for music educators) and *location* of target symbol (on the 1st or 3rd beat of a musical bar),

TABLE 3. *Mixed-Effects Model Analyses for Beat-Specific Eye-Time Span (BETS) and First-Pass Fixation Time (FPFT): Results From Likelihood-Ratio Tests.*[a]

		BETS		FPFT	
Fixed effects	df	χ^2	*p* value	χ^2	*p* value
Expertise	1	0.046	.830	0.530	.467
Measurement[b]	1	0.006	.937	0.571	.450
Location	1	4.143	.042*	2.623	.105
Target Symbol	2	4.714	.095	102.23	<.001***
Preceding Context	1	101.730	<.001***	7.404	.007**
Following Context	1	4.420	.036*	2.584	.108
Target Symbol: Preceding Context	2	7.659	.022*	3.288	.193
Target Symbol: Following Context	2	0.241	.887	2.038	.361
Preceding Context: Following Context	1	0.318	.573	5.387	.020*
Target Symbol: Preceding Context: Following Context	2	0.099	.952	0.258	.879

* $p < .05$, ** $p < .01$, *** $p < .001$

[a] Restricted maximum likelihood (REML) was used as the estimation method.

[b] The effect of measurement session was additionally examined with only those 18 participants whose data from both measurements were included (total of 622 observations; for further details for data handling, see Data Analysis section). Even then, the measurement session did not have a significant main effect nor any significant interactions with other factors.

were also considered as fixed effects. By including *participant* as a random effect we took into account the within-subject correlation in the observations and controlled for the subject-level variation; thus, each participant could have their own average level of FPFT or BETS. These fixed and random effects and interactions formed the full model fitted to the data.

Goodness-of-fits for the final models were checked from the residuals, which were fairly well normally distributed in both cases. Statistical powers were assessed through simulations using the simr (Green & MacLeod, 2016) package and each power is based on 1,000 simulations. This power analysis utilizes the fitted model by simulating new values for the response variable and refits the model to the simulated response. A statistical test is then applied to the simulated fit (Green & MacLeod, 2016). The power calculations were done for fixed effect sizes; these were chosen to be close to the estimated ones.

Results

BEAT-SPECIFIC EYE-TIME SPAN

We used likelihood ratio tests to attain significance of the fixed effects and interactions (see Table 3). Starting from the full model, terms were excluded one by one from the highest term onwards (i.e., from the bottom of the list in Table 3), each time comparing the model with the term in question with the one without it. The final model for Beat-Specific Eye-Time Span (BETS) was formed based on the results of the likelihood ratio tests such that all significant fixed effects and interactions were included. Location was the only control variable having a significant effect on BETS (see Table 3), hence the final model did not include measurement or expertise effects (for the mean values of these variables, see Appendix A). The final model for BETS was *location* + *target symbol* + *preceding context* + *following context* + *target symbol*: *preceding context* (difference in Akaike information criterion: AIC_{null} - AIC_{final} = 174). We also checked all the smaller (nested) models for BETS, but each of these had a larger AIC value than the final model.

For BETS, there were thus significant main effects of *location*, *preceding context*, *following context*, and an interaction between *target symbol* and *preceding context*. The parameter estimates of the final model including these effects are presented in Table 4. To begin, Figure 3a visualizes the main effect of *following context* using predicted values given by the model. BETS at the target was predicted by the model to be 74 ms longer when the target was followed by a dense context than when it was

TABLE 4. *Parameter Estimates (in Milliseconds) of the Final Model Fitted for Beat-Specific Eye-Time Span*

	b	*SE*	*t*
(Intercept)	1235.93	63.18	19.561
Location (Third)	67.86	34.40	1.973
Target Symbol (Quarter Rest)	153.69	57.02	2.696
Target Symbol (Set of 16th Notes)	180.95	58.12	3.113
Preceding Context (Dense)	−233.20	58.00	−4.021
Following Context (Dense)	74.12	34.34	2.158
Preceding Context (Dense): Target Symbol (Quarter Rest)	−219.63	83.22	−2.639
Preceding Context (Dense): Target Symbol (Set of 16th Notes)	−165.77	84.13	−1.971

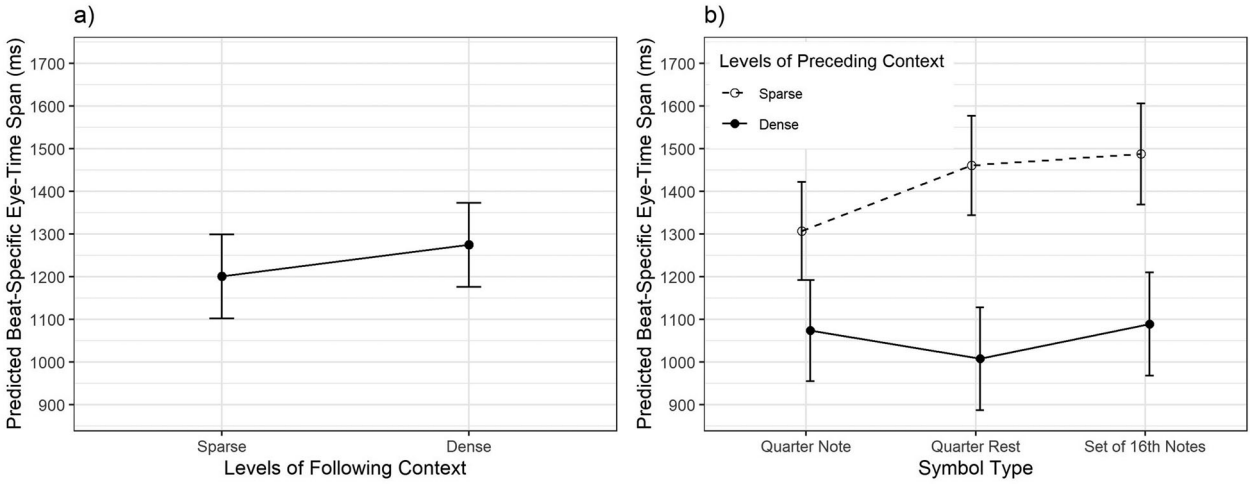


FIGURE 3. Beat-Specific Eye-Time Span (A) at different levels of following context and (B) at different levels of preceding context and for different target symbols. Error bars represent 95% confidence intervals.

followed by a sparse context (*SE* = 34 ms, *t* = 2.158). The power for an effect of size 70 ms was 55%. (This means that if there is a significant difference in BETS when the target is followed by the dense context compared to sparse one and we observe this difference to be 70 ms, in such a case the model is able to detect the effect with the probability of 0.55.) Considering the *location* effect, if a target symbol is located on a third beat of a bar, BETS is estimated to be 68 ms longer on average (*SE* = 34 ms, *t* = 1.973) compared to when it is located on the first beat. However, the power for the location effect of size 70 ms is basically 0%, and we will thus ignore this result.

The interaction between *preceding context* and *target symbol* is visualized in Figure 3b using predicted values given by the model. The powers of the interaction effects of size 170 ms were 51% (*preceding context* is dense; *target symbol* is quarter rest) and 52% (*preceding context* is dense; *target symbol* is set of 16th notes). According to Figure 3b, BETS is generally estimated to be several hundreds of milliseconds longer (up to ∼1.5 seconds; this equals 1.25 beats in the tempo of 50 bpm) when the target is preceded by a sparse context (quarter-note symbols), compared to a dense preceding context (16th notes). Post hoc tests for the interaction effect indicated that BETS was significantly larger in the sparse preceding context compared to the dense preceding context for each of the target symbols (all Tukey-adjusted *p* values < .001). The significant interaction between *preceding context* and *target symbol* seems to be provoked by the difference between quarter notes and a set of 16th notes after sparse preceding symbols, *t*(785) = -3.113, Tukey-adjusted *p* value = .024. In sum, the effect of target symbol type on BETS depended on the complexity of the preceding context: given a sparse preceding context, 16th notes were approached with a longer BETS than quarter notes.

FIRST-PASS FIXATION TIME

The final model for First-Pass Fixation Time was formed similarly as the one for BETS (see Table 3 and Appendix A). Unlike with BETS, however, location did not have a significant effect on FPFT. Thus, the final model for FPFT is *target symbol* + *preceding context* + *following context* + *preceding context*: *following context* (difference in Akaike information criterion: AIC_{null} - AIC_{final} = 153). We also checked all the smaller (nested) models for FPFT, only to observe that they had a larger AIC value than the final model.

The parameter estimates for the final model including the significant fixed effects as well as the significant interaction are presented in Table 5. As for the main

TABLE 5. *Parameter Estimates (in Milliseconds) of the Final Model Fitted for First-Pass Fixation Time.*

	b	SE	*t*
(Intercept)	1006.52	50.46	19.947
Target Symbol (Quarter Rest)	78.04	43.90	1.778
Target Symbol (Set of 16th Notes)	441.59	44.53	9.916
Preceding Context (Dense)	−176.39	51.08	−3.453
Following Context (Dense)	−131.25	50.02	−2.624
Preceding Context (Dense): Following Context (Dense)	149.81	72.70	2.061

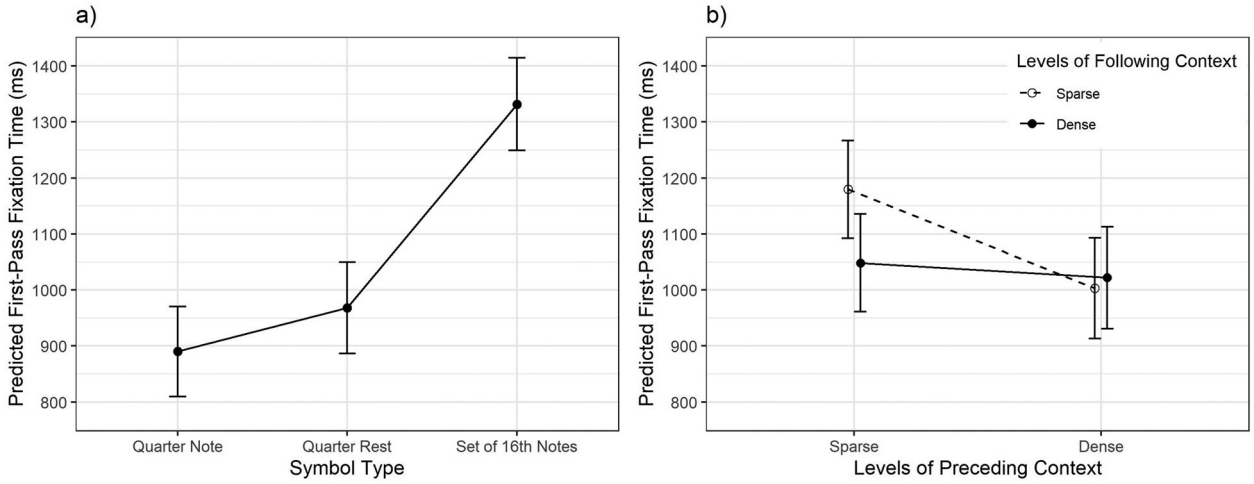


FIGURE 4. First-Pass Fixation time (A) for the three symbol types and (B) at different levels of preceding and following context. Error bars represent 95% confidence intervals.

effect of *target symbol*, the FPFT was estimated to be 442 ms longer on average at 16th notes compared to quarter notes (SE = 44.53, t = 9.916), and an average of 78 ms longer at quarter rests than at quarter notes (SE = 43.90, t = 1.778; for visualization with predicted values, see Fig. 4a). The powers of the effect size 80 ms are 44% and 45% for the main effect of quarter rest and set of 16th notes, respectively.

The interaction between *preceding context* and *following context* is visualized with predicted values in Figure 4b. The power of the effect size 150 ms is 55%. In sum, effect of the preceding context on the FPFT at a target symbol depended on the following context; only when the following context was sparse, the preceding context had an effect in such a way that the FPFT was larger for sparse preceding contexts. Post hoc tests verified the significant difference between the sparse-sparse condition and the three other conditions (all Tukey-adjusted p values < .05).

Discussion

In this study, we identified three possible cognitive mechanisms governing eye movements during temporally controlled music reading: *symbol comprehension, visual anticipation,* and *symbol performance demands*. The mechanisms were derived from prior research, where they have been studied separately from one another. We then conducted an empirical study where we asked which of the three mechanisms are needed to explain how strings of note symbols of varying relative complexity are visually processed during simple, temporally controlled rhythm reading. In each case, prior research suggested that the cognitive mechanisms underlying successful rhythm reading would be reflected in measurements of Eye-Time Span (ETS) and/or First-Pass Fixation Time (FPFT) at selected target symbols. For methodological purposes, we simplified the ETS measure to *Beat-Specific Eye-Time Span* (BETS): at the time of first glancing at a note symbol, it measures the temporal distance between the current point of time in the music and the beat onset corresponding to the fixated symbol. In our data set, all analyzed observations were temporally stable and correctly performed; thus, our pattern of findings can be taken to reflect the cognitive processes underlying successful rhythm reading.

It turned out that what we considered the most complex rhythm symbol in the stimuli (both in terms of visual layout and motor execution)—a set of 16th notes—was processed with a longer FPFT than the two other target symbols, a quarter note and a quarter rest. This points toward a *symbol comprehension* effect, and is in accordance with Kinsler and Carpenter's (1995) music-reading model (see also Penttinen & Huovinen, 2011). In addition, target symbols surrounded by simpler symbols and target symbols with more complex ones nearby resulted in different distributions of FPFT: in practice, simpler surrounding symbols allowed more first-pass fixation time to be spent on the target. Importantly, this interaction occurred irrespective of the target symbol itself. We thus see that in temporally controlled music reading, the zero-sum game of time use is constantly at play even in such simple tasks as the rhythm reading discussed in this paper.

We also found evidence of *visual anticipation*. First, relatively more complex symbols *following* a target symbol increased the BETS at the target, irrespective of the target symbol itself. Second, when a string of two quarter notes preceded a target symbol, a more complex target symbol (the 16th-note pattern) tended to attract the gaze earlier than when the target was yet another quarter note. Together with prior reports (Chitalkina et al., 2021; Huovinen et al., 2018), these observations support the notion that in temporally controlled music reading, visual attention is indeed drawn sooner toward relatively more complex note symbols than less complex ones. This occurs when the surrounding symbols are simple enough to enable it. However, we need to note that this finding might be explained either by the saliency of the visual symbol or by the greater relative complexity of the musical pattern indicated (see Huovinen et al., 2018): these two factors were confounded in our design, as they indeed tend to be in Western music notation. Nevertheless, given the above-mentioned *symbol comprehension* effect, it seems that *visual anticipation* may have served the same overall purpose of allocating fixation time to more complex (or salient) symbols: readers continuously readjust their span of looking ahead from the point of performance, allowing more complex upcoming symbols to attract their gaze earlier than less complex ones. While logically separable, the mechanisms of *visual anticipation* and *symbol comprehension* thus seem empirically intertwined.

Finally, also the cognitive mechanism we call *symbol performance demands* found some support in our findings, but the support was more ambiguous than for the two other mechanisms. In our study, the BETS at the target symbol appeared to shorten when the symbol was preceded by relatively more complex symbols, and this could suggest that the performance of complex symbols restricts the performer's simultaneous "looking ahead" to the target symbol (as in, e.g., Penttinen et al., 2015). However, in our modeling, this effect was in interaction with the target symbol type (the interaction supporting the *visual anticipation* mechanism), which cautions against emphasizing the effect of the preceding symbols alone. Also in terms of FPFT, the results gave partial support for the notion of *symbol performance demands*: the symbols *before* a target did affect the time spent at the target symbol, as might be expected based on prior studies (Chitalkina et al., 2021; Hadley et al., 2018), but only in interaction with the type of symbols *following* the target (the interaction supporting the *symbol comprehension* mechanism). Further studies with different research designs are clearly needed to elucidate the role that varying performance demands may have on the eye-movement process during music reading. However, our findings point toward the possibility that in music reading, the visual process and the motor performance might compete of (some of) the same cognitive resources, after all (cf. Kinsler & Carpenter, 1995; Salvucci & Taatgen, 2011). Although *symbol comprehension* and *visual anticipation* found stronger support in our empirical data, we suggest that *symbol performance demands* should not yet be ignored when studying cognitive factors that affect eye-movement processing during music reading.

We also wish to point out one way in which the participants' tapping performances might be reflected in our results. Notice that what we have called the most "complex" symbol in our study was always the same beamed group of four 16^{th} notes. Therefore, allocating most FPFT to the 16th-note groups might not so much have reflected any great difficulties in decoding these visual symbols, or in figuring out their proper tapping execution. However, despite the redundancy of this composite symbol in the tapping session, it still collected more fixation time than the other symbols—perhaps simply reflecting the slightly more extended action sequence that it signalled to be carried out. We propose that patterns of FPFT might thus illustrate an embodied view of rhythm reading in which the meanings of notation symbols are not consummated in their proper recognition but only in their proper execution (here, by tapping on a drum pad). Saliently "thick" visual symbols also tend to require more elaborate performance actions, and readers indeed seem to adjust their gaze as if to support their motor effort.

In summary, similarly to the studies by Huovinen and colleagues (2018) and Chitalkina and colleagues (2021), our study points towards music readers' sensitivity to the visual stimulus itself, highlighting the bottom-up effects brought about by the characteristics of musical notation. Unlike many other studies focusing on skill-based differences (see Madell & Hébert, 2008; Puurtinen, 2018b), we found no effects of expertise in this simple tapping task. Arguably, this is because of the highly simplified stimuli used in the study. More diverse and realistic musical stimuli could introduce more problems of decoding the symbols, thus perhaps bringing about skill-based differences in the visual processing as well. While the testing of basic assumptions regarding cognitive processes in music reading best begins with simple enough stimuli, future work should address these bottom-up effects with more sophisticated musical materials. Notice, too, that our handling of musical expertise was rather coarse, since participants were roughly divided into two skill-based groups. With

a larger participant pool, musical background (or, preferably, pre-tested sight-reading skill as in, e.g., Gilman & Underwood, 2003) could perhaps have been treated as a continuous variable.

This work also shows some technical limitations. Although the original number of participants was, in our view, adequate, this number was then reduced due to our strict criteria for data included in the final analyses. We considered a data set consisting of correct, temporally stable performances, with each participant providing several observations, the best way to test our hypotheses. In music-reading research, data sets often include eye-movement data from higly different kinds of (correct or more or less erroneous) performances (Puurtinen 2018b). Though such pooling of data adds to the number of observations and might therefore increase statistical power, it also necessarily adds noise to the data set, making it more difficult to relate the pattern of results to any specific factor of interest. Lastly, we also need to note the lowish recording frequency (60 Hz) of the applied eye tracker that might well have compromized the quality of our findings, although probably not their overall form.

As the present study demonstrates, our understanding of the course of the music-reading process in still at a relatively early stage. We already know quite a lot about the typical eye-movement processes observed in simple music-reading situations, but the underlying cognitive mechanisms are not yet as clearly understood. We urge music-reading research to begin to emphasize the systematic creating and testing of cognitive models of eye movements in music reading. We propose that the still missing cognitive model (or models) about the regulation of eye movements in music reading should take at least *symbol comprehension* and *visual anticipation* into account, and continue to explore the possible significance of *symbol performance demands*. Future studies could continue from here by exploring the interplay of these and other possible mechanisms, and thus enhance our understanding of this unique type of reading task.

Author Note

The project was supported by Turku Institute for Advanced Studies (for the first author) and the Academy of Finland (grant number: 275929). The authors would like to express their gratitude to research assistant Suvi Heinonen and to the participants of this study for their time and effort, and to the reviewers of the manuscript for their excellent and supportive feedback.

Correspondence concerning this article should be addressed to Marjaana Puurtinen, Department of Teacher Education, FI-20014, University of Turku, Turku, Finland. E-mail: marjaana.puurtinen@utu.fi

References

Ahken, S., Comeau, G., Hébert, S., & Balasubramaniam, R. (2012). Eye movement patterns during the processing of musical and linguistic syntactic incongruities. *Psychomusicology: Music, Mind and Brain*, *221*, 18–25.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Burman, D. D., & Booth, J. R. (2009). Music rehearsal increases the perceptual span for notation. *Music Perception*, *26*, 303–320.

Chitalkina, N., Puurtinen, M., Bednarik, R., Gruber, H., & Bednarik, R. (2021). Handling of incongruences in music notation during singing or playing. *International Journal of Music Education*, *39*, 18–38.

Drai-Zerbib, V., Baccino, T., & Bigand, E. (2012). Sight-reading expertise: Cross-modality integration investigated using eye tracking. *Psychology of Music*, *40*, 216–235.

Furneaux, S., & Land, M. F. (1999). The effects of skill on the eye-hand span during musical sight-reading. *Proceedings of the Royal Society of London, Series B*, *266*, 2435–2440.

Gilman, E., & Underwood, G. (2003). Restricting the field of view to investigate the perceptual span of pianists. *Visual Cognition*, *10*, 201–232.

Goolsby, T. W. (1994a). Profiles of processing: Eye movements during sightreading. *Music Perception*, *12*, 96–123.

Goolsby, T. W. (1994b). Eye movement in music reading: Effects of reading ability, notational complexity, and encounters. *Music Perception*, *12*, 77–96.

Green, P. & Macleod, C. J. (2016). SimR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*, 493–498.

Hadley, L. V., Sturt, P., Eerola, T., & Pickering, M. J. (2018). Incremental comprehension of pitch relationships in written music: Evidence from eye movements. *The Quarterly Journal of Experimental Psychology*, *71*, 211–219.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (Eds.). (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.

Huovinen, E., Timoshenko, M., & Nyström, M. (2021). Eye movements in sight singing: A study with experts. *Psychomusicology: Music, Mind, and Brain*, *31*, 134–148.

Huovinen, E., Ylitalo, A.-K., & Puurtinen, M. (2018). Early attraction in temporally controlled sight reading of music. *Journal of Eye Movement Research*, *11*(2), 1–30.

Hyönä, J., Lorch, R. F. Jr., & Rinck, M. (2003). Eye movement measures to study global text processing. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 313–334). Elsevier Science.

Inhoff, A. W., & Wang, J. (1992). Encoding of text, manual movement planning, and eye-hand coordination during copytyping. *Journal of Experimental Psychology: Human Perception and Performance*, *18*, 437–448.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, *87*, 329–354.

Kinsler, V., & Carpenter, R. H. S. (1995). Saccadic eye movements while reading music. *Vision Research*, *35*, 1447–1458.

Laubrock, J., & Kliegl, R. (2015) The eye-voice span during reading aloud. *Frontiers in Psychology*, *6*.

Lenth, R. (2018). *emmeans: Estimated Marginal Means, aka Least-Squares Means*. R package version 1.2.3. https://CRAN.R-project.org/package==emmens.

Lim Y., Park, J. M., Rhyu, S-Y. Chung C. K., Kim, Y., & Yi, S. W. (2019). Eye-hand span is not an indicator of but a strategy for proficient sight-reading in piano performance. *Scientific Reports*, *9*.

Madell, J., & Hébert, S. (2008). Eye movements and music reading: Where do we look next? *Music Perception*, *26*, 157–170.

Maturi, K. S., & Sheridan, H. (2020). Expertise effects on attention and eye-movement control during visual search: Evidence from the domain of music reading. *Attention, Perception and Psychophysics*, *82*, 2201–2208.

Penttinen, M., & Huovinen, E. (2011). The early development of sight-reading skills in adulthood: A study of eye movements. *Journal of Research in Music Education*, *59*, 196–220.

Penttinen, M., Huovinen, E. & Ylitalo, A. (2015). Reading ahead: Adult music students' eye movements in temporally controlled performances of a children's song. *International Journal of Music Education*, *33*, 36–50.

Perra, J., Poulin-Charronnat, B., Baccino, T., & Drai-Zerbib, V. (2021). Review on eye-hand span in sight-reading of music. *Journal of Eye Movement Research*, *14*. https://doi.org/10.16910/jemr.14.4.4

Polanka, M. (1995). Research note: Factors affecting eye movements during the reading of short melodies. *Psychology of Music*, *23*, 177–183.

Puurtinen, M. (2018a). Learning on the job: Rethinks and realizations about the use of eye tracking in music-reading studies. *Frontline Learning Research*, *6*, 148–161.

Puurtinen, M. (2018b). Eye on music reading: A methodological review of studies from 1994 to 2017. *Journal of Eye Movement Research*, *11*(2), 1–16.

R Core Team (2021). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Rayner, K. (2009). The 35th Sir Frederik Bartlett Lecture. Eye movements and attention in reading, scene perception and visual search. *The Quarterly Journal of Experimental Psychology*, *62*, 1457–1506.

Rayner, K., & Pollatsek, A. (1997). Eye movements, the eye-hand span, and the perceptual span during sight-reading of music. *Current Directions in Psychological Science*, *6*, 49–53.

Rosemann, S., Altenmüller, E., & Fahle, M. (2016). The art of sight-reading: Influence of practice, playing tempo, complexity and cognitive skills on the eye-hand span in pianists. *Psychology of Music*, *44*, 658–673.

Salvucci, D. A., & Taatgen, N. A. (2011). *The multitasking mind*. Oxford University Press.

Sheridan, H., Maturi, K., & Kleinsmith, A. L. (2020). Chapter Five - Eye movements during music reading: Toward a unified understanding of visual expertise. *Psychology of Learning and Motivation*, *73*, 119–156.

Sloboda, J. A. (1974). The eye-hand span – An approach to the study of sight reading. *Psychology of Music*, *2*, 4–10.

Truitt, F. E., Clifton, C., Pollatsek, A., & Rayner, K. (1997). The perceptual span and the eye-hand span in sight-reading music. *Visual Cognition*, *4*, 143–161.

Waters, A., & Underwood, G. (1998). Eye movements in a simple music reading task: A study of experts and novice musicians. *Psychology of Music*, *26*, 46–60.

Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag.

Wong, Y. K., & Gauthier, I. (2010). Holistic processing of musical notation: Dissociating failures of selective attention in experts and novices. *Cognitive, Affective, and Behavioral Neuroscience*, *10*, 541–551.

Wong, Y. K., & Gauthier, I. (2012). Music-reading expertise alters visual spatial resolution for musical notation. *Psychonomic Bulletin and Review*, *19*, 594–600.

Wurtz, P., Müri, R. M., & Wiesendanger, M. (2009). Sight-reading of violinists: Eye movements anticipate the musical flow. *Experimental Brain Research*, *194*, 445–450.

Appendix A

Mean BETS and FPFT From a Series of Rhythm-Tapping Tasks in the Tempo of 50 bpm: Standard Deviations in Parentheses

	Mean Beat-Specific Eye-Time Span (ms)	Mean Beat-Specific Eye-Time Span (beats in the tempo of 50 bpm)	Mean First-Pass Fixation Time (ms)	Number of participants / Number of observations
Novices	1196 (537)	1.00 (0.45)	1103 (597)	11 / 268
Musicians	1263 (589)	1.05 (0.49)	1037 (554)	18 / 550
Measurement 1	1243 (568)	1.04 (0.47)	1078 (589)	24 / 421
Measurement 2	1238 (579)	1.03 (0.48)	1039 (547)	23 / 397