Learning on the job: Rethinks and realizations about eye tracking in music-reading studies

Marjaana Puurtinen[1]

University of Turku, Finland

Abstract

The application of new methods and measures in domains with few methodological traditions of that kind often presents researchers with a challenge; they may have to take up the task of developing their understanding of the phenomenon while, at the same time, creating the practices for its study. For us, the method was eye tracking, and the topic, music reading. One key characteristic of music reading is that the music reader's gaze moves slightly ahead of the current point of performance. This gap allows the performer to prepare for the upcoming motoric responses. In this paper, we present our 10-year-long path, describing the steps we have taken while studying this "looking ahead" in music reading. We will point out how we have, after both advances as well as setbacks, come to change our views on how best to explain the various components affecting this specific act and how it is best measured. Finally, we discuss some of the lessons we have learned, hoping in this way to provide practical suggestions for others who plan to take up methods from other domains and use them in novel ones.

Keywords: Expertise; eye tracking; eye-hand span; eye-time span; music reading

---

1   Introduction

Similar to all researchers who plan to apply a method more often used for studying other topics than the one they are wishing to target, we in our team have had to learn about expertise in music reading in parallel with developing the eye-tracking methodology for its study. Our initial educated guesses about how to go about this were based on one team member's prior experience with eye tracking and on all team members' much better comprehension of Western music notation than of the cognitive side of things (due to our background in music and academic training in other disciplines than psychology). As often occurs in research, some of our hunches were better than others, and our views about both what to study and how to study it have therefore changed considerably throughout the 10 years we have conducted our work. It may be that the interest in process measures in educational sciences puts more and more project leaders, research coordinators, doctoral candidates and their supervisors in similar situations to where we were some time ago. To be sure, many will be much faster in learning their topics than we have been. Still, for anyone intending to take up eye tracking or any other similarly complicated method and apply it to a little explored topic, we suggest that it is good to prepare for similar kinds of rethinks and realizations to what we have faced and, in advance, already be thinking of both practical and emotional ways to cope with them.

We begin our history by giving some background about eye tracking in the music domain (based on our current understanding of the matter), and then we briefly describe some of our more or less successful data collections and attempts to report our findings from 2008 onward with respect to the study of "looking ahead" during music reading. For the sake of presentation, we write about our sub-studies as Steps 1, 2 and 3, although the whole process consisted of several overlapping stages. Finally, based on our history, we state some of the lessons we have learned from the perspective of taking up new research methods in little explored fields.

1.1 Why use eye tracking in the music domain?

With eye tracking, we can accurately record the durations and locations of *fixations*, that is the short moments (typically around 200–400 ms) when we target our gaze to a certain location in our visual array and process information from it. During one fixation, we only see a narrow area accurately (for instance one word of this text), and in order to collect more visual information, we must quickly move our eyes to fixate upon another spot (for instance the next word). The *perceptual span* is the area from which useful visual information is obtained. For instance, when we fixate on <u>this</u> word, a few words to the left and right of it fit within the perceptual span but are blurred. This blurred visual information helps us, however, to see where the next word begins, and we can then target our next fixation on that word. Our eye movements are so rapid that we get the impression that our eyes do not perform these skips and jumps, and we may also fixate upon targets (for instance reread a word) without noticing it ourselves. (For details, see Rayner, 2009; Holmqvist et al., 2015.) We therefore need eye trackers to record and show us how we are looking at a painting, reading a newspaper, checking the traffic while driving or otherwise examining our surroundings.

Eye tracking has a long research tradition in psychological studies about text reading. In these studies, eye-movement data is considered to give insight into the cognitive processes related to the reading and comprehending of words and longer texts (see Rayner, 1998; 2009). The method has also been used to study visual expertise in various domains (Jarodzka, Holmqvist, & Gruber, 2017), of which chess and medicine are the ones typically mentioned first. In expertise research, the goal is to make experts' automated visual practices visible and, similarly to text-reading research, to search for indicators of the underlying cognitive processes. Often this is done by comparing the visual processing of experts to that of novices or intermediates while they perform a domain-specific task.

Music reading would seem like a natural part of studies about visual expertise. It is a visual motor task rehearsed and widely used by a large group of amateur musicians and used professionally (and on very high level of expertise) by some, while there are still many almost illiterate with respect to Western music notation. Music reading also has one feature which points especially to the usefulness of the eye-tracking method; while reading, the performer produces a continuous motor response, as she is constantly "acting out" the stimulus. This gives the researcher constant verification of the quality of the visual processing as well as the possibility

of synchronizing, throughout a recording, the "intake" of visual information with a motor output. This is unlike the process with many other visual tasks. (See Puurtinen, 2018.)

At this point, we need to specify what we mean by "music reading". To us, it is a task where someone reads music notation and executes the symbols in one way or another. The motor part may consist only of tapping rhythms, singing or playing an instrument (see Video 1). We consider it important to make a distinction between performance tasks and tasks where music notation is only read and not performed in any way. To highlight the difference, we have come to call the latter *silent (music) reading* (Penttinen, Huovinen, & Ylitalo, 2013). As an example, think of a piano teacher taking a pile of sheet music in her hands in order to go through the sheets and select which piece to play with a student. The teacher is probably not reading all of the note symbols to the extent that they could be performed (cf. the performer in Video 1) but is scanning the music quickly in order to see whether one piece would seem to be of suitable difficulty. In this type of a task, there are no temporal restrictions imposed upon the reading (see section 1.2, below), nor any need to plan for motor responses. Thus, compared to a performance task, we believe that the silent-reading task presents the readers with very different goals for going through the material and, accordingly, with very different cognitive demands (see Penttinen, et al., 2013; Puurtinen, 2018). However, due to the lack of a need to plan for a motor response, silent-reading tasks could be used as a reference when comparing music reading with the processing of other types of domain-specific symbol systems. For those interested in the visual processing of note symbols with no performance requirements, we suggest that they turn to studies which have applied a non-performance design (e.g., Waters & Underwood, 1998; Burman & Booth, 2009; Drai-Zerbib & Baccino, 2013; Penttinen, et al., 2013).

Video 1. The author's eye movements during the piano performance of a familiar children's song, "Mary Had a Little Lamb". A metronome (the click heard in the background) is set at 60 beats per minute and provides the temporal framework. Recorded with a Tobii TX300 Eye Tracker and a Yamaha electric piano.

Overall, then, the use of the eye-tracking method seems well suited for music-reading studies. Still, for some reason, such studies are quite few, and the field lacks a coherent research tradition (for reviews, see Madell & Hébert, 2008; Puurtinen, 2018). We can, however, say that in general, experts in music reading are faster at encoding note symbols than are novices and intermediate performers, and therefore, they can read the music with shorter fixation durations. This expertise effect has been noticed both in silent-reading tasks (Waters & Underwood, 1998; Penttinen, et al., 2013) and during-performance tasks. Considering the latter, experts' advanced skill has appeared both when more and less experienced musicians' performances differ in terms of performance tempo (Truitt, Clifton, Pollatsek, & Rayner, 1997; Gilman & Underwood, 2003; Penttinen & Huovinen, 2011) but also when all performances are alike in a temporal sense and in performance quality (Penttinen, Huovinen, & Ylitalo, 2015). The features of the music notation, too, play a role in the reading, though this side of the process has only rarely been studied and when it has been, often only on a very general level (Madell & Hébert, 2008; Puurtinen, 2018).

1.2 Music reading is all about the use of time

The uniqueness of music reading as a visual motor task is in the fact that it is *temporally regulated*. Time can be thought to affect the reading on two levels. First, the overall tempo affects the amount of *absolute* time a performer has for inspecting and performing the visual elements on the score. For example, if one performer uses 10 seconds to perform the melody in Video 1, and another one plays the same melody in 5 seconds, it is obvious that the latter performer has less time to study the written symbols and plan for the motor responses. Secondly, the *course* of reading is also regulated, because the *relative* durations of individual notes are signalled by the symbolic system itself. In Figure 1, in measure 2, the performer has to perform more (two eighth notes) during the first beat and less (only one quarter note) during the second beat. Naturally, the performer may choose to ignore the durations of individual symbols and stop to correct errors, but that will lead to a "staggering" performance, where the flow of the music is disrupted. This is exactly what beginning musicians do when they are incapable of interpreting the symbols and performing them in a given tempo (e.g. Drake & Palmer, 2000). All of this is unlike the situation in text reading, where readers can return to difficult sections or words and where, in fact, targeted rereading may even result in better text comprehension and be considered a preferable strategy (e.g. Hyönä, Lorch, & Kaakinen, 2002).

Another complexity of Western music notation is the fact that the same symbols contain information about both the relative duration of each symbol and the pitch height of each (e.g. which key to press on a piano keyboard), and the reader has to process both features in order to perform accurately. Figure 1 demonstrates the relative durations of notes in "Mary Had a Little Lamb" (the first line) along with the whole melody, with pitch heights included (the second line). (Note that this is a highly simplified example of Western music notation; the amount of information on musical scores can vary from this type of simple information to detailed information about simultaneously performed notes, performance tempos and expressive interpretation.)
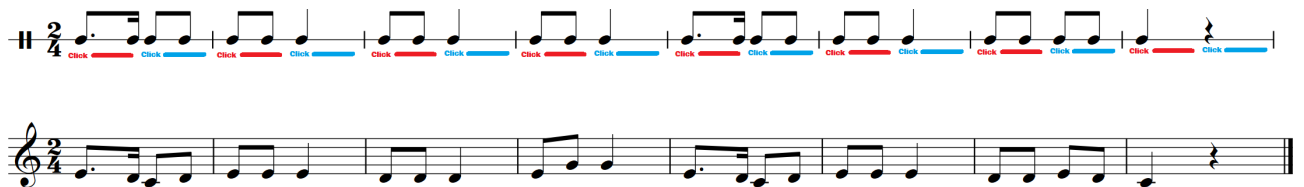

Figure 1. Above, the rhythms of "Mary Had a Little Lamb". In Western music notation, the score is typically divided into "measures" marked by vertical bar lines. In this example, in measure 2, the performer first performs two eighth notes during one quarter beat (the beat marked in red) and then one quarter note which lasts for the whole quarter beat (the beat marked in blue). The correct timing of the notes (which land either on a beat onset or between two beat onsets) is of importance, as it secures the steady flow of the music. Below is the full song, the melody represented through the different pitch heights of the note heads.

Thus, musicians have to make sense of the complex symbolic system and select what to target their gaze at, as well as when and for how long, in order to perform successfully while operating within the given temporal framework. We know that musicians manage this by maintaining their gaze slightly ahead of the current point of performance, and with the help of this buffer (typically of perhaps around 1–2 seconds [Furneaux & Land, 1999; Penttinen, et al., 2015; Huovinen, Ylitalo, & Puurtinen, 2018; Rosemann, Altenmüller, & Fahle, 2016; see Video 1]), the performer prepares for the upcoming motoric responses. The gaze may typically tend to remain very close to the performed notes (Truitt et al., 1997; Penttinen, et al., 2015), but instead of a steady "looking ahead", at least for skilled music readers, the reading may also consist (mainly or in part) of rapid back and forth eye movements (Goolsby, 1994; cf. Penttinen, et al., 2015). Overall, this "looking ahead" behaviour, often called the *eye-hand span* (Madell & Hébert, 2008; Holmqvist et al., 2015) should work as an indicator of music-reading-related cognitive processes, since it seems to vary due to performers' expertise as well as stimulus features. However, the interplay of all the factors involved is still not properly understood, and the field is methodologically scattered to the extent that only the most elementary features of the "looking ahead" can be regarded to be established ones. (For a recent review of findings thus

far, see Huovinen et al. [2018].) Before 2008, the year of our first data collection, published works on this topic were indeed very few, and we therefore had to start considerably near the beginning.

2 Our eye-tracking investigations about the "looking ahead" in music reading

2.1 Step 1: Focus only on gaze targets (Penttinen & Huovinen, 2009; 2011)

2.1.1 *Study summary*

We began our work in 2008–2009 by recording non-professional pianists' sight-reading of simple melodies prior to, during and after nine months' training. Our aim was to track down potential eye-movement indicators of early skill development, especially relating to the reading of larger melodic intervals. Our author team then consisted of a PhD candidate in education (who was also an active musician) along with a musicologist, who were supported by a group of colleagues from a department of teacher education. Forty-nine education majors took part in the experiment, of whom 15 novices and 15 amateur musicians' data was included in the final analyses. The data collection was organized alongside a compulsory music course (with piano playing as one of the course topics), and this allowed us to follow the novices' development during a real-life learning task. Our four experimental melodies consisted of quarter notes (see Figure 2) played only with the right hand, with a metronome signalling the onset of each quarter note at a relatively slow pace (60 beats per minute, where each beat thus lasted for 1 s). The melodies were stepwise apart from two larger melodic intervals (see the "skips" at two of the bar lines in Figure 2).

Figure 2. One of the four piano performance tasks applied in Penttinen and Huovinen (2009; 2011). In this melody, each of the right-hand fingers could be put on one key (the thumb on the first note), and the performer did not have to move her hand. For the most part, even for the novices, this eliminated the need to look at the fingers; with difficult pieces, pianists often need to divide their visual attention between the score and the keyboard to ease motor coordination.

We built a set-up where music was presented from the screen of a 50 Hz Tobii 1750 Eye Tracker, and an electric piano, attached to a laptop with sequencer software, was positioned in front of the tracker. After careful consideration, no chin rest was used, to allow the performers as natural a playing position as possible. We assumed that although during our simple tasks the performers did not need to look at their fingers to any great extent (apart from prior to performing when placing the hand on the keyboard), preventing that altogether with a chin rest might have had larger effects on the eye-movement data than the rare looks toward the keyboard. Previous music-reading studies have not been very precise about how participants in them were trained for the study protocols (sometimes there is no indication that any training took place), but we also took care in preparing a practice trial with exactly the same kind of protocol (down to every instruction slide appearing on the screen) than what was then applied in the actual recording. This proved a useful procedure, and we have applied that practice ever since.

We attempted to launch both the eye-tracking and sequencer recording simultaneously from a third computer, but though the system worked in pilots, inconsistent lags started to appear during the actual data collection. Thus, we could not synchronize the performance and eye-movement data after all, and our analyses had to be limited to the allocation of fixation time without information about where the performance was ongoing at each fixation. We reported indicators of novices' skill development in three respects. First, after training, the novices performed the large intervals with better timing and accuracy. Second, together with increasing performance accuracy, the novices began to allocate more fixation time to the last two notes of each measure (similar to the amateurs, who did so from the first measurement session onward). Third, we noticed

that the novices, with training, perhaps began to identify the latter notes in large intervals more quickly. This was suggested by the gradual shortening of the average first-pass fixation durations for those symbols. There were, however, some matters which suggested that some of the results should be treated with caution, and we therefore tried to corroborate them later on. After two revisions, the manuscript was accepted for publication (Penttinen & Huovinen, 2011).

2.1.2   Rethinks and Realizations after Step 1

This was the first music-reading study we designed, and it certainly had some drawbacks, but perhaps with beginners' luck, some benefits as well. To begin with the latter, due to some of our background in text-reading studies, we decided to apply in our music-reading studies the eye-tracking measures suggested by Hyönä, Lorch and Rinck (2003) as suitable common ones for text-reading research. We only later fully realized the need for music-reading research, too, to be much more consistent in the use of the measures; we continued with these measures in our later work. Another benefit was the choice of very simple melodies; we realized that the set-up should perhaps be kept simple, but the reasons were not solely due to our less sophisticated understanding of the eye-tracking method and the kind of stimuli with which it was most usefully applied. We also simply considered what kinds of melodies our novices could learn to play during the nine-month-long training. However, we did come to think (to some extent, at least) of the fact that due to the multidimensional nature of Western music notation (see section 1.1), we should be able to distinguish whether any eye-movement effects were due to the interpretation or planning of the rhythm or melodic features of a certain melody. By simplifying the rhythm in these tasks, we could narrow down our interpretations and suggest that our novices' eye-movement effects were indeed indicators of the interpreting and planning of the melodic features in the melodies, rather than being about reading and planning the execution of the rhythmic patterns. This has not been the case in many music-reading studies (see, Puurtinen, 2018).

However, we were also faced with several challenges, both while collecting the data and especially when analysing it. First, the synchronization of performance and eye-tracking data failed, as described above. For this reason, we could only speculate about what had caused shorter or longer fixation times and, importantly, *when* particular fixations took place. We had no idea whether our novices and amateurs read in a steady manner approximately one second ahead of the performed notes or whether they tended to do longer advance inspections and then return toward the point of performance, as was proposed as one possible pattern for skilled readers (Goolsby, 1994). When discussing the findings, we could only offer alternative hypotheses about the "looking ahead" in this specific task. The fixation durations suggested that not all note symbols were treated equally, but we did not have the full explanation of why. And in the analysis of note-specific fixation times, we also applied the Kruskal-Wallis analyses with dependent data in one part of the analyses, which, we know now, omitted the within-subject dependence. Furthermore, we presented the four melodies in the same order to all participants and did not counterbalance them. Thus, although the mean values for first-pass fixation times for selected note symbols decreased for the novices during training, the statistical testing should be interpreted with caution and retested with a better study design. We tried to address these matters in our follow-up studies.

2.2 Step 2: Adding the hand (Penttinen, Huovinen, & Ylitalo, 2012; 2015)

2.2.1   Study summary

In 2008–2009, when we collected the data set described above, we also invited the same participants to perform another task, to play a familiar children's song, "Mary Had a Little Lamb", a few times. Some versions of the melody contained a one-measure-long "mistake" or notes which were "wrong", and our idea was to trace the eye-movement indicators of coping with such surprising misprints in the music. Here, too, we applied a metronome, which forced the performers to solve the problem of playing something against their expectations at a given tempo. Our attempt with the 49 participants failed (for details, see section 2.2.2, below), and we ended up using only five experienced musicians' data as a pilot study (reported as study 1 in Penttinen, et al. [2012]).

We developed the setting further and collected a new data set in 2011. At this point, we had also included a student of statistics in our team. Due to many novices' difficulties in performing the melody accurately and in the given tempo in our first try, this time we only asked skilled performers (amateur and professional-level pianists) to perform the same song and two of the variations from the previous attempt. An electric piano was placed in front of a 300-Hz Tobii TX300 Eye Tracker, and the performances were recorded into a separate computer. The metronome was provided by the second computer, which recorded the performances in MIDI format. Both computers were operated by the first author, who coordinated the change of slides on the eye tracker monitor so that they occurred at metronome beats. This procedure provided both the eye tracking and MIDI data streams with simultaneous events, and we created a laborious way to synchronize the data streams based on these markers (for details, see Penttinen, et al., 2015). This enabled us to calculate what we called the eye-hand span, that is, how much the gaze was ahead of the onset of a performed note (see Figure 3).

Figure 3. How we calculated the eye-hand span in Penttinen and colleagues (2015). In this melody, two quarter beats (beat 1 in red, beat 2 in blue) fit into one measure. The performer plays all of the note symbols. When performing the first note symbol of measure three (in red), the performer's gaze is targeted near the note in the second beat area of the measure (in blue). Thus, at this point, the eye-hand span is, roughly put, one quarter beat. In the tempo of 60 beats per minute, this equals one second. We also calculated what we called gaze activity, namely how many quarter beat areas were fixated on when performing the notes of one quarter beat (i.e. notes fitting under a red or a blue line). This was typically only two beat areas.

We observed that professional performers applied longer eye-hand spans more often than amateurs did. Also, during one-second intervals, professional pianists fixated on more of the music than amateur performers did. We called this finding an increase in their *gaze activity*. However, these between-group differences disappeared in the face of melodic deviations, suggesting that performing against expectations inhibited the professional performers' reading, to some extent. We were also able to compare in detail the "looking ahead" during the performance of different rhythm symbols and obtained some information about the reading of quarter notes and the more rapidly performed eighth notes (see Figure 1). For the analyses, we used t- and Chi-square tests and were fairly content with our protocols for data collection and analyses — for a while. Our first full-paper submission was not successful, but the second one was (Penttinen, Huovinen, & Ylitalo, unpublished manuscript; Penttinen, et al., 2015).

2.2.2   Rethinks and Realizations after Step 2

As described above, our first attempt (in 2008-2009) with this particular research design failed. First of all, there was a lack of synchronization of the performance and eye-movement data. Secondly, the task was too difficult for the most novice participants, and this reduced the data set considerably. Erroneous performances were so much unlike each other (with different types of errors appearing in different parts of the melodies) that they could not be pooled. Thirdly, to present the melody in a big enough font size, we placed it on two rows (first four measures on row 1, last four on row 2). Thus, the performers' gaze moved from the end of the first row to the beginning of the second, and in such a short reading task, all additional and accidental fixations landing here and there during these sweeps from one line to the next made the eye-movement data even more complicated to handle. Finally, one of the three variations we used had the "mistake" in the second-to-last measure. However, during music reading, the endings differ from other parts of the reading, since there is nothing more to look in advance, and thus, the "extra" time gained is just spent on the final measures. Thus, this variation did not work out similarly to the other two variations.

After we addressed these deficiencies in our second data collection in 2011, we began to get a bit closer to what we were after. Skilled performers' adjustment of the eye-hand span and gaze activity, when something unexpected had to be performed, suggested that the "looking ahead" is indeed involved in the planning of motor responses and that it is affected both by the performers' expertise and by stimulus characteristics. This time, we had also been able to synchronize the performance and eye-movement data with what we thought to be sufficient accuracy, and we had found measures which had brought forth something new about the "looking ahead".

However, even though we were content with this improved data set and thought it could indeed contribute to our understanding about the "looking ahead", the first manuscript we submitted was rejected and only later accepted by another journal. This was, at the latest, the moment when we realized that it really is hard to find publication forums for this type of work (apart from music education, which we thought to be one relevant field). As we framed our work both within educational sciences, where music is by no means a mainstream topic, and within cognitive musicology, where in 2011, eye-tracking was still a relatively unknown method, and as none of us were psychologists and therefore not accustomed to write to that audience, we were caught somewhere in between these fields. (At this point, we need to thank all reviewers who took up the job of reviewing our work; some of them have stated that they were not experts in music or eye-tracking, and we do appreciate the effort they put into their reviews.)

In addition, we later came to be more critical of our analytical approaches; they were still only a series of separate (piecemeal) comparisons. We started to think that if several factors (e.g. performer characteristics, even the minutest features of the stimulus, and most likely also the performance tempo) did co-influence the "looking-ahead" in music reading, as now seemed to be the case, we should study them in a way which would separate the overriding factors from those which interacted with other factors. We had also applied only one tempo in these studies. (Some of these realizations were reached while already working on our next data sets [see 2.3, below]). Although this performance task, with its authentic melody and violation of expectations, differed from the simple performance task in Step 1 and was designed to address very different kinds of research questions, in retrospect, the studies seem to have shared more common features in a methodological sense than we had previously considered.

2.3 Step 3: Replacing the hand with metrical time (Huovinen, et al., 2018)

2.3.1 Study summary

Step 3 had its origins in the data collection in 2008–2009. We first attempted to publish the 2008–2009 data as a separate paper which focused not on novices' reading but on the reading of the two different kinds of experimental melodies (ones which had a large interval in the middle of a measure and ones where the "skip" was across a bar line) during correct performances. However, even after submitting a revised version, our work was rejected, and we realized that we did need the synchronized performance data in order to fully explain our findings (Huovinen & Penttinen, unpublished manuscript; Huovinen, Penttinen, & Ylitalo, unpublished manuscript). Thus, in 2011, we asked the same skilled participants who had performed the "Mary Had a Little Lamb" task to also perform another task. We modified the 2008–2009 set-up (see 2.1, above) by adding eight more melodies to the set, including four with no large intervals, or "skips", in them. We had realized that we needed these stepwise melodies so that we would have a baseline with which to compare the melodies with the large intervals. We also asked the performers to play half of the melodies in a slower tempo and half in a faster one. The Tobii TX300 Eye Tracker was used, and all was considered to go smoothly; our skilled performers did the tasks with very good temporal accuracy and few mistakes, and though the work was laborious, we were able to synchronize the performance and eye-movement data with a similar technique as in the task in Step 2.

Our previous measure for the eye-hand span was calculated from the point of performance onward (see Figure 3). This was the traditional approach in other prior studies, too. However, we now noticed that in fact, this measure describes more those moments in the performance when the performer *can* look ahead (Huovinen, et al., 2018), that is, moments when performing a symbol is easy enough to allow the reading of symbols further ahead. In our case, that was not the issue of interest (especially when the melodies were almost

overly simple to perform). Since we had simple melodies with "more difficult" notes in them at the large intervals, our question was whether these notes were fixated upon, or "looked ahead", as early on as possible and whether the visual processing was therefore affected by such minute features of the written music. This we did not know beforehand. Thus, we came to realize that the eye-hand span measure should be calculated exactly the other way around from how we had done it; if we wanted to examine what the performer *needed* to look at in advance, we needed to take the first-pass fixation upon a specific target as the starting point and calculate the "span" *backward* to where the metrical time was at that particular moment (see, Figure 4). In practice, this would tell us when a performer attempts to fixate a symbol as early on as possible. We therefore differentiated between two measures, the traditional eye-hand span$_{PERFORMANCE}$ and the new eye-hand span$_{FIXATION}$.

Figure 4. How we calculated the eye-hand span$_{FIXATION}$, or eye-time span, in Huovinen and colleagues (2018). (This is only for illustration; please note that those studies actually applied stimuli resembling that in Figure 2.) The red arrow marks the steady progress of metrical time. The eye-time span was calculated *from the first fixation* to a note symbol (in this example, from the last note in measure three). By regarding the metrical time as a continuous axis, we could calculate where the metrical time was at the onset of a fixation. In this example, the eye-time span is 1.5 beats or, in a tempo of 60 bpm, 1500 ms (i.e. the metrical time is ongoing 1500 ms behind the fixated note). The keypress (or "hand") information was only used to estimate the correctness of the performance as well to synchronize the metrical time with the eye-movement recording; after that, all calculations were based on the timestamps for fixation onset with respect to the thus abstracted metrical time.

We considered the adding of the eye-hand span$_{FIXATION}$ measure a significant advance in our thinking. In terms of data analyses, we also turned to generalized estimation equations and began to statistically model our data instead of listing a series of pairwise comparisons. This shift in analytical thinking later greatly affected how we designed our studies. At this point, we thought we had finally come to understand the "looking ahead" from all its angles and what kinds of questions should be answered with what kind of approach. (Later, we noticed that in domains other than music reading, the eye-hand span can also be calculated as the distance between the first fixation on a target and its subsequent performance [Holmqvist et al., 2015; Huovinen, et al., 2018].) However, our measures and findings were far from easy to explain or present, and our first manuscript, too, was not accepted for publication (Huovinen, Ylitalo, Penttinen, & Penttinen, unpublished manuscript).

In 2015, we decided to collect one more data set in order to corroborate our previous findings. This data collection was enabled by large project funding, which also gave the three authors a new support team, with representatives from the fields of education and musicology. In what was titled Experiment 2 (data from 2011 being Experiment 1), we had only professional pianists performing longer melodies which had more difficult larger intervals, again in the same two tempos as in the 2011 study. This time we used the 60-Hz Tobii T60XL Eye Tracker. The idea was to make the large intervals more salient and thus make any potential "looking ahead" they may cause more profound. From this perspective, it was only the eye-hand span$_{FIXATION}$ measure that could answer this question. Thus, we left out the comparison of the two eye-hand measures and moved on to explain the effects caused by the now ever more salient large intervals only with the eye-hand span$_{FIXATION}$ measure, and we gave it a new more appropriate name, the *eye-time span* (since the hand was not involved in the calculations). Again, we thought that we were on to something; we were focusing on one characteristic of the music notation and how it affected the "looking ahead", having also developed a measure which could be used for its study. The feedback for this was not favourable, however, and the manuscript, in which we reported Experiments 1 and 2 together, was rejected (Huovinen, Ylitalo, & Puurtinen, unpublished manuscript). After rethinking again the nature of the phenomenon under study, and the measures, we prepared yet a new manuscript, this time including a more detailed description of the eye-hand span and eye-time span,

their measurement, and use. In it we reported that besides the general effects of performance tempo and musical expertise on the eye-time span, the local melodic complexity (i.e. the larger intervals) could elicit longer advance inspections; however, these inspections might not land on the "complex" notes themselves but on the ones preceding them (Huovinen, et al., 2018).

2.3.2   Rethinks and Realizations after Step 3

All of Step 3 indeed took a while, and those years included a number of disappointments and moments when we had to admit that what we had thought to be a good idea was not. Nevertheless, despite the emotional struggle, in retrospect, we know that we needed all of that earlier work. During each stage (be it data collection or the writing of a rejected manuscript), we realized and came to think of something which then helped us with the next study design and analyses. We also improved the formulation of research questions and are now better able to estimate what kinds of evidence eye tracking can provide us with in this context. As a result, we can now argue for slightly more straightforward explanations of our findings and can see more clearly which way we or some other team should turn next.

However, the results we reported by Huovinen and colleagues (2018) are still not the full answer to all of our questions. If, in suitable circumstances, as the data suggests, the pianist's gaze is drawn toward salient features, but fixations do not necessarily land exactly on them, this suggests that the perceptual span (see 1.1, above) has come more clearly into play. In some cases, then, the salient features observed (albeit in a blurry form) within the perceptual span may attract a visual response so early that the performer can not target a fixation upon that exact note; if she did, the gaze might move too far away from where the music is ongoing. This would perhaps make it difficult to keep all the notes in between in short-term memory and perform them correctly. Whether this undershooting is involuntary due to temporal restrictions, or deliberate because the procedure is enough for fitting the problematic symbols into the area of accurate vision, we can not be sure. However, perhaps the visual information which is visible from that fixation point is still enough to allow the performer to get ready for a correct motor response. In order to search for the full answer to this question, we would have to conduct studies that apply the gaze-contingent methodology, that is, where a limited part of the notation is shown to the reader at a time (Rayner, 1998). The method has apparently been applied only in two studies about the size of the perceptual and eye-hand spans during music reading (Gilman & Underwood, 2003; Truitt et al., 1997), but not with the most controlled music stimuli—and we do claim, based on all our research, that the stimulus features should be included as part of the analyses and should not be thought of only as part of the study protocol (Puurtinen, 2018). We have, to this day, opted for a "natural" layout of the music, showing all of the music during our recordings, but after Step 3 it now seems that even more experimental set-ups may be needed in order to further explain the interplay between the eye-time span and the perceptual span during music reading.

Another matter we have come to recognize is that both the eye-hand span and the eye-time span we have calculated have one drawback: we have only calculated them at certain points during the performances (although in Step 2, the gaze activity measure described to some extent what happened between two quarter beats). The eye-time span is actually a continuously changing parameter, similar, for instance, to pupil size. Thus, to truly understand the "looking ahead", the study of this aspect should not be limited in future only to within-subject or between-subject comparisons at certain notes. Instead, we should also develop methods which allow the examining of the "looking ahead" as a process, a characteristic of a full performance, varying according to the factors which appear to affect it.

3   Lessons learned

As the examples above perhaps demonstrate, our study of the "looking ahead" in music reading has been a long and, on occasion, also tiring endeavour. We have came to change a number of things in our methodologies, such as the skill level of our study participants (starting with novices and amateurs and finally recording performances of professionals only), our measures (developing one, the eye-time span, that we thought will be the answer to our questions but which actually lead to even more open ones), as well as our analytical approach (from piece-meal comparisons to statistical modelling). Already these examples should demonstrate the methodological twists and turns of our work. To be sure, each step we have taken, from

planning an experiment to the final published report of the findings, has made us less the novices we initially were. Still, we are aware that the path toward expertise in this methodology and domain will require the taking of many more similar steps.

Hopefully, this history of our choices and our reasons for moving away from our earlier choices are of benefit to others planning to use new methods and measures and perhaps even use them in domains where there is little methodological support available. We will now attempt to summarize some of the more general lessons we have learned for the benefit of others who may find themselves in a similar situation to where we have been (and still are), using our own experiences as an example.

First, many of us who are trained in something other than the computer sciences might not be in our comfort zone when we first start to apply research methods which require the use of technology we are not used to. For us, for instance, the synchronizing of the two recording onsets in Step 1 was done with remote support from a manufacturer. We were able to build the system ourselves, but when it failed, there was nothing we could do about it. When learning the use of new equipment, and with limited skills in solving problems related to it, we recommend that researchers create a plan B in their research design; they should ensure that their research questions can be answered, at least in part, even when something goes wrong on the technical side. In our case, the comparison of novices and amateurs helped us to make use of the data, and we also included a task of silent music reading where the synchronization was not needed (Penttinen, et al., 2013). However, more insight into the participants' reading would have been gained had the technical side been more successful.

Second, when starting to work with a new method and/or new types of stimuli, an exploratory pilot with a complex and perhaps realistic task is a possible starting point for creating research hypotheses (see, Puurtinen, 2018), but to get beyond descriptive data, it may be useful for researchers to control their stimuli (or task) as much as possible. This might not answer the most interesting of research questions, but enables one to test the possibilities of the method. For instance, in prior work about eye movements in music reading, the role of the applied music was thought of more as a part of the study protocol, something "read" or "played", although it now seems that even the minutest details of the music (whether there is one quarter note or two eighth notes to be played, or a slightly larger interval among an otherwise stepwise melody) are already reflected in the eye-movement process. Thus, following this line of thought, before knowing how a complex task is actually represented in the measure of the researchers' choice, they should start with simple set-ups. The reporting of the findings might be challenging (we have heard a number of times how our tasks are not "musical" at all, since they are so simple), but in the end, one can argue with more certainty what one's findings might mean and what their cause might be.

Third, for others beginning their work with new methods, we suggest the use of measures from other domains (or previous studies, if available). Importantly, pay attention to the fact that measures which different research teams consider the same may, in fact, be calculated quite differently. With unique measures and their definitions, one's results might be very hard to align with others' findings. In music-reading studies, there is great variability in the measures and in how they are calculated (Puurtinen, 2018); the eye-hand span, with its many ways of operationalization, is a very good (or bad) example (Huovinen, et al., 2018). In our work, partly because they were familiar to us, we began from the very beginning with some measures applied in text-reading research, and we have been content with that early choice we made. In developing areas for research, we should aim at methodological consistency to the extent that it is possible and should provide data and reports which can later be used in meta-analyses. To be sure, measures also need to be developed further and new ones tested; still, existing measures provide a good starting point.

Fourth, after conducting some studies with similar protocols, test the protocols. This is something we should also do. For instance, our team members' own background in music made us think of using a metronome in order to secure the temporal similarity of the performances. In practice, we had an external "click" marking the onset of certain beats. We only later noticed that this was not a standard protocol, up until that point, at least. The external temporal control means that the performer herself does not have to use any of her cognitive capacity to maintain the tempo (something that is challenging, especially for novice and intermediate performers), but still, the possible effects caused by the external click should also be tested. It may be that in our data, there are main effects caused either by the strain of following the external click or by the lack of strain due to the relaxation of the need to maintain the tempo without any help. Similarly, with

other methods, there may be protocols which first seem sensible and which are ones we become accustomed to, but those could also raise new research questions when further thought through.

Fifth, if possible, create or join a team, and make it something more than just a group of colleagues. In research areas which reach out to several background theories or methodologies in particular, no one can be the expert on all of the relevant topics, and it is difficult for one person to reach all the audiences which might be interested in and whose input would be beneficial for the work. Also, potential disappointments and difficulties are much easier to handle with people whom the researchers trust and who support them — and the celebrations are also much more fun like that. Thinking of our Step 3, one may stop to think whether, alone, they would have done the three data collections, written and submitted the several manuscripts and revisions which were rejected, or kept one subproject going for the ten years it took to get one piece of the work published. We now think it is a good thing that we made it (though we immediately saw the next steps someone should definitely take). Of course, in long-term projects and among people who work with something as personal as their intellectual skills, small or large conflicts necessarily arise at one point or another — but with good enough emotional bonding and similar goals, these do not break the team and just become a part of life.

To conclude, we have here presented how we have tried to develop our methodologies for the study of "looking ahead" in music reading, openly listing our unpublished work and what we now know were mistakes or simply bad ideas. Overall, as will probably also be the case for many others, we do not think our path has simply been either a success or a failure. Instead, it has been a back-and-forth movement (cf. Goolsby, 1994), and we are glad to have the chance to report also the backward steps. Hopefully, other researchers, too, can share those parts of their projects which are faded out of polished publications and presentations — those are, after all, the ones we could really learn from.

Keypoints

- The application of a method in a research domain where there is little methodological tradition of that kind presents the researcher with a challenge; likely, there is a need to develop the understanding of the phenomenon and the methodology in parallel.
- By using our own 10-year-long work as an example, we demonstrate the advances and setbacks we have faced in our eye-tracking studies about music reading.
- We discuss some of the lessons we have learned in order to provide practical suggestions to others planning to take up a new method or apply one in a novel domain.

Acknowledgments

References

Burman, D. D., & Booth, J. R. (2009). Music rehearsal increases the perceptual span for notation. *Music Perception: An Interdisciplinary Journal, 26*(4), 303–320. DOI: 10.1525/mp.2009.26.4.303

Drai-Zerbib, V. & Baccino, T. (2013). The effect of expertise in music reading: Cross-modal competence. *Journal of Eye Movement Research, 6*(5), 1–30. DOI: 10.16910/jemr.6.5.5

Drake, C., & Palmer, C. (2000). Skill acquisition in music performance: Relations between planning and temporal control. *Cognition*, *74*(1), 1–32. DOI: 10.1016/S0010-0277(99)00061-X

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*(1436), 2435–2440. DOI: 10.1098/rspb.1999.0943

Gilman, E., & Underwood, G. (2003). Restricting the field of view to investigate the perceptual span of pianists. *Visual Cognition*, *10*(2), 201–232. DOI: 10.1080/713756679

Goolsby, T. W. (1994). Eye movement in music reading: Effects of reading ability, notational complexity, and encounters. *Music Perception*, *12*(1), 77–96. DOI: 10.2307/40285756

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2015). *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press

Huovinen, E., & Penttinen, M. (unpublished manuscript). The allocation of fixation time in simple sight-reading tasks.

Huovinen, E., Penttinen, M., & Ylitalo, A.-K. (unpublished manuscript). The visual processing of melodic group boundaries: An eye-movement study.

Huovinen, E., Ylitalo, A.-K., Penttinen, M., & Penttinen, A. (unpublished manuscript). The where and when of sight reading: Effects of performance tempo and musical structure on eye movements.

Huovinen, E., Ylitalo, A.-K., & Puurtinen, M. (unpublished manuscript). The eye-time span: Structural salience and looking ahead in simple sight-reading tasks.

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. DOI: 10.16910/jemr.11.2.3

Hyönä, J., Lorch, R. F., Jr., & Kaakinen, J. K. (2002). Individual differences in reading to summarize expository text: Evidence from eye fixation patterns. *Journal of Educational Psychology, 94*(1), 44–55. DOI:10.1037//0022-0663.94.1.44.

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). Amsterdam: Elsevier Science.

Jarodzka, H., Holmqvist, K., & Gruber, H. (2017). Eye tracking in educational science: Theoretical frameworks and research agendas. *Journal of Eye Movement Research, 10*(1):3, 1–18. DOI: 10.16910/jemr.10.1.3

Madell, J., & Hébert, S. (2008). Eye movements and music reading: Where do we look next? *Music Perception*, *26*(2), 157–170. DOI: 10.1525/mp.2008.26.2.157

Penttinen, M., & Huovinen, E. (2009). The effects of melodic grouping and meter on eye movements during simple sight-reading tasks. In J. Louhivuori, T. Eerola, S. Saarikallio, T. Himberg, & P. Eerola (Eds.), *Proceedings of the 7th Triennial Conference of European Society for the Cognitive Sciences of Music*, Jyväskylä, Finland. (pp. 416-424). Available at https://jyx.jyu.fi/dspace/handle/123456789/20910

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*(2), 196-220. DOI: 10.1177/0022429411405339

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (unpublished manuscript). Eye-movement effects of melodic deviations in temporally controlled sight reading.

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2012). Unexpected melodic events during music reading: Exploring the eye-movement approach. In E. Cambouropoulos, C. Tsougras, P. Mavromatis, & K. Pastiadis (Eds.), *Proceedings of the 12th International Conference on Music Perception and Cognition and the 8th Conference of the European Society for the Cognitive Sciences of Music,* Thessaloniki, Greece. (pp. 792-798). Available at http://icmpc-escom2012.web.auth.gr/sites/default/files/papers/792_Proc.pdf

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2013). Silent music reading: Amateur musicians' visual processing and descriptive skill. *Musicæ Scientiæ, 17*(2), 198-216. DOI: 10.1177/1029864912474288

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2015). Reading ahead: Adult music students' eye movements in temporally controlled performances of a children's song. *International Journal of Music Education: Research, 33*(1)*,* 36-50. DOI: 10.1177/0255761413515813

Puurtinen, M. (2018). Eye on music reading: A methodological review of studies from 1994 to 2017. *Journal of Eye Movement Research, 11*(2), 1-16. DOI: 10.16910/jemr.11.2.2

Rayner, K. (1998). Eye movement in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422. DOI: 10.1037/0033

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62*(8), 1457–1506. DOI: 10.1080/17470210902816461

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*(4), 658–673. DOI: 10.1177/0305735615585398

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*(2), 143–161. DOI: 10.1080/713756756

Waters, A., & Underwood, G. (1998). Eye movements in a simple music reading task: A study of experts and novice musicians. *Psychology of Music, 26*(1), 46–60. DOI: 10.1177/0305735698261005