Eye movements reveal the visual component of music expertise: evidence from a visual search task

Kinnera Savitri Maturi

University at Albany, State University of New York, maturi.kinnera@gmail.com

The University at Albany community has made this article openly available. Please share how this access benefits you.

Follow this and additional works at: https://scholarsarchive.library.albany.edu/legacy-etd

Part of the Psychology Commons

Recommended Citation
https://scholarsarchive.library.albany.edu/legacy-etd/2330

This Master's Thesis is brought to you for free and open access by the The Graduate School at Scholars Archive. It has been accepted for inclusion in Legacy Theses & Dissertations (2009 - 2024) by an authorized administrator of Scholars Archive. Please see Terms of Use. For more information, please contact scholarsarchive@albany.edu.
EYE MOVEMENTS REVEAL THE VISUAL COMPONENT OF MUSIC EXPERTISE:
EVIDENCE FROM A VISUAL SEARCH TASK

by

Kinnera S. Maturi

A Thesis
Submitted to the University at Albany, State University of New York
In Partial Fulfillment of
the Requirements for the Degree of
Master of Arts

College of Arts & Sciences
Department of Psychology
2019
Abstract

An important component of expertise is the ability to rapidly recognize domain-related perceptual patterns. To explore this ability in the domain of music reading, a unique visual search paradigm was used to compare the eye movements of 30 expert musicians (with at least 10 years of music reading skill) and 30 novices (who could not read music). Participants had to match a section of a piece of sheet music (search template) to its identical counterpart within a larger music score (search array). Both the search template and array were presented on the screen simultaneously, which allowed for visual comparisons between the target and the array. By employing a visual search task to isolate the visual component of music, it was possible to study music expertise across a wide range of expertise. The task elicited strong expert/novice differences. Relative to the novices, experts had higher accuracy, and spent more time looking at the relevant regions and less time looking at irrelevant regions. Also, as evidence that the experts and novices adopted qualitatively different search strategies, the experts spent more time than novices looking at the search template at the beginning of the trial, and the experts returned to this region less often than novices. Taken together, these results indicate that experts can acquire accurate representations of highly complex search template, which allows them to focus their attention on relevant rather than irrelevant regions during their search. These findings are consistent with the predictions of chunking and template theories of expertise.

Keywords: attention, chunking, eye tracking, expertise, music, visual search
Eye Movements Reveal the Visual Component of Music Expertise

Eye-tracking research can unveil many important qualities of expertise differences. While visual expertise has been studied in other domains such as chess (Reingold et al., 2001), medicine (Sheridan & Reingold, 2017), sports (Mann, Williams, Ward, & Janelle, 2007), and airport security (McCarley, Kramer, Wickens, Vidoni, & Boot, 2004), the perceptual component of visual expertise in music has been less extensively studied. Music is a multi-modal domain that has the potential to yield significant insights into expertise, given the fact that visual perception can be isolated within a domain that contains its own syntax and structure. The aim of this study was to investigate this visual component of music expertise by using eye tracking to contrast the search strategies of expert musicians and novices while they performed a novel music-related visual search task. The purely visual nature of the task allows for performance comparisons between expert musicians and novices to provide support for the idea that experts use chunks or templates in order to solve problems within their domain. I will begin by discussing prior investigations of visual expertise, followed by a discussion of the chunking and template theories of expertise that will provide the motivation for our predictions, and I will then introduce the current study’s visual search paradigm.

An expert can be characterized by the amount of domain-specific knowledge they have, skill level, hours practiced, or even their decision-making ability. Generally, experts can perceive patterns in a more meaningful way, are flexible in their problem-solving approach, can understand complex problems, and have an efficient filter of relevancy when faced with a new problem (Shanteu, 1992; Chi, Glaser, & Farr 2014). Dreyfus (2004) identifies five stages of skill acquisition to become an expert: novice, in which the student learns the rules, facts, and domain-specific features of a skill; advanced beginner, in which the novice learns to consolidate their
semantic knowledge with real-world applications and experience; competence, where the learner not only deals with the overload of information gained from acquisition, but is able to identify and correct mistakes; proficiency, where the individual learns to draw away from rigidity and analyze situations independently based on previous experience; and finally the expert, when one is fully engulfed in their domain and can skillfully understand how to complete the task at hand. Experts do not have higher overall intelligence, but simply have more domain-specific knowledge gained from years of practice, as DeGroot (1978) showed within the domain of chess.

Within the domain of music, experts show superior advantages within their domain compared to novices. The visual aspect of music is unique in nature, as there is a prerequisite amount of knowledge required to be able to understand the syntax and grammar structure of sheet music. Previous studies have shown various perceptual advantages that music experts utilize in their performance. Expert musicians are able to understand larger portions of the music through its overall themes and familiar patterns (Wong & Gauthier, 2010a, 2010b; Penttinen, Huovinen & Ylitalo, 2013), especially when there is contextual similarity within the music score (Kinsler & Carpenter, 1995). In comparison, novices tend to focus on individual features, such as the music notes, more than experts do (Gilman & Underwood, 2013). Experts also tend to have a larger eye-hand span, which indicates how far their eyes are ahead of the hand during instrument playing (Wristen, 2005; Penttinen, Huovinen, & Ylitalo, 2015; Truitt, Clifton, Pollatsek, & Rayner 1997).

These specific behaviors exhibited by experts lead to better performance by experts than novices. Looking across domains, experts are more quickly able to solve problems with better accuracy than novices are (DeGroot, 1946; Larkin, McDermott, Simon, & Simon, 1980). One possible explanation for this is provided by Chase and Simon’s (1973) chunking theory, which
tested the problem-solving abilities of chess masters and novices. In seminal work by Chase and Simon (1973) and de Groot (1946), chess experts were able to reproduce patterns of chessboards they viewed for a few seconds to a higher degree of accuracy than novices. However, Chase and Simon (1973) also showed that experts and novices demonstrated similar accuracy when shown randomized chess patterns, which ruled out the idea that experts simply have better memory and instead supported the viewpoint that the expert’s knowledge of domain-specific visual patterns supported their memory performance. Chase and Simon theorized that as skill level increases, experts are able to recognize familiar patterns, and by recognizing a pattern, they can create and access meaningful chunks stored within the pattern. These chunks, in turn, store information about the individual chess pieces, which is why experts are rapidly and accurately able to make decisions about the best move on the board during a chess game. Moreover, due to the experts’ meaningful schemas or patterns in their long-term memory, experts can process visual information in larger sections rather than individual features. This global or holistic manner of processing is a hallmark trait of expertise that also extends to domains beyond chess, including medical image perception (see Sheridan & Reingold, 2017, for a review). Experts’ ability to process information in a more global manner allows them to solve problems quicker and more efficiently than novices can.

Building on chunking theory, Gobet and Simon (1998) introduced template theory to further explain the role of domain-specific patterns in chess. According to template theory, the chunks that experts in a field acquire become larger structures of information called templates when they are frequently recalled. Gobet and Simon posit that a template not only contains information about patterns, but also contains open “slots” that can store new information. This alternate theory of expertise and memory accounts for the way experts are able to retrieve
information from their memory within a matter of seconds, but also provides the elasticity to adapt their knowledge as they are exposed to more domain-specific information (Sala & Gobet, 2016) making it less rigid than chunking theory.

To understand chunking and template theories, as well as to investigate eye movement control, eye tracking can be used to provide rich information about search strategies, attention allocation, and visual and cognitive processing. Eye tracking measures such as the number of fixations, saccade amplitudes, and fixation locations can indicate the way someone parses the information in front of them in order to solve a problem (see Rayner, 1998 for a review). Regarding music, eye movement analyses provide insight as to how music is processed at a perceptual level (Madell & Hebert, 2008). As such, eye movement measures can be a very useful way to evaluate expertise. If experts are able to view situations in a global manner compared to novices, such that they are taking in more information (Cellier et al., 1997), then this would be shown through their pattern of fixations, dwell times, and saccades. Based on their findings within the chess domain, Reingold, Charness, Pomplun, and Stampe (2001) posit that experts are able to extract visual information from a larger region than novices during a fixation, which means that experts are better able to interpret feature locations and the relationships between those features using their parafoveal (or peripheral) vision. In other words, experts have a larger visual span than novices do, since they are able to take in and interpret a large amount of complex information from a single glance. This visual span (also known as the perceptual span) can increase in size as expertise level increases (Reingold & Sheridan, 2011).

A visual search paradigm can be an effective way to investigate the differences between music experts and novices using eye movements (see Eckstein, 2011, for a review). Eye movements have previously shown to be different across tasks, such as reading and searching in
a scene (Rayner, 2009; Rayner & Pollatsek, 1992), so we cannot assume that a visual search task involving music scores will elicit the same pattern of eye movements as a music reading task. During a visual search task, eye movements can be influenced by the quality of an expert or novices’ representation of the search template in visual working memory to help guide their search (Wolfe & Horowitz, 2004). As evidence for the importance of the search template in guiding the search, eye movements during visual search tasks are impacted by the familiarity of search template. Previous studies have found that the more similar the search template is to the target, the easier and more efficient a search becomes (Yang & Zelinsky, 2009; Vickery, King, & Jiang, 2005). Search templates are sensitive to discrepancies between the visual input and the target, and the more precise a search template is, the more effective it is in facilitating a successful search (Hout & Goldinger, 2015). Although visual search has often been examined in abstract, low-level contexts, it is important to consider real world situations as well, and when we observe objects in our visual field in organic situations, we process them as objects, and not a composition of features (Neider & Zelinsky, 2006). As an example of a real-world visual search task with complex targets, Bravo and Farid (2009) set up a visual search design in which participants had to find a specific species of fish amongst a coral reef visual. The target fish served as extremely specific and complex search templates in a real-world context. Bravo and Farid found that when people form search templates that fit a category but allow flexibility for changes in specific detail, they performed better in the task as compared to participants who held rigid and specific search templates. While familiarity does assist in visual search tasks, the knowledge of the context and ability to adapt has also been shown to be an advantage. Furthermore, Hout and Goldinger (2015) found that when the visual input is rich in detail but the similarity of the cue to the target is reduced, performance decreases. Applying this to expertise,
experts naturally have more familiarity with visual patterns in their domain, which could allow them to form chunks and templates. Therefore, chunking and template theory predicts that in the present study, experts, who are familiar with the visual patterns that frequently occur in music scores, will perform better on the task since they are better able to create precise and detailed search templates to assist them in a music-related visual search task.

The present study therefore integrates expertise research and visual search research. In the present study, I sought to explain the relationship between expertise in music reading through template and chunking theories using a visual search paradigm and eye tracking methodology. Because music itself is a multimodal skill (visual, auditory, motor), the present study chose to solely focus on the visual component of perceptual expertise. The advantage of using a music-related visual search task to study extreme expertise is threefold: 1) sheet music is used as the stimulus, which allows for complex search templates given the richness of detail contained within the syntax of music reading; 2) by isolating music to the visual component of expertise, novices as well as experts should be able to successfully complete the task which will allow us to compare performance across a wide span of expertise; 3) by including the search template simultaneously with the search array throughout the trial, it will be possible to investigate if experts are using the more precise search template than novices, resulting in differences in their search patterns as revealed by eye movements. Due to chunks and templates, experts might have a better representation of the search template in this task, and eye tracking will allow us to test this prediction by revealing how attention is spatially allocated during the task.

The present study investigated how experts and novices differ in performance in a visual search task using sheet music bars as search templates that are visually complex and rich in detail. Additionally, the search template was presented simultaneously with a search array (see
Figure 1). This aspect of our paradigm is unique compared to other visual search tasks (in which the search template is typically presented before the array) and this innovation will allow us to investigate how experts and novices allocate their attention to the search template versus the search array throughout the trial. Due to their knowledge of domain-specific patterns (e.g., chunks and templates), it is likely that experts would have a more precise search template during the task. As a result, the experts may show higher accuracy, and they may not need to move their eyes back to the search template as often as novices, to the degree that they form a better memory representation of the search template. Moreover, I predict that experts will be able to locate relevant regions of the visual field more efficiently than novices will, as indicated by shorter overall reaction times, shorter dwell times in distractor regions, shorter dwell counts, and shorter times to first fixation within the target region. Thus, pre-existing chunks and templates should support experts’ performance in the task, since they already have a memory for the visual pattern, which could lead to experts and novices having qualitatively different search strategies when completing the task. Finally, I plan to test for possible interactions between complexity and expertise, to explore if increasing the visual complexity of the search template and music score might differentially impact experts and novices.

Method

Participants

Participants were categorized into two groups: music experts and music novices. The experts (N = 30) were recruited from the surrounding campus and community and were compensated $10 per hour. Expert musicians were defined as those who had studied music through any instrument in a formal or informal setting for a minimum of ten years. Expert musicians were also fluent in reading and interpreting sheet music and had a strong grasp of
music theory. Novices (N = 30) were undergraduate students at the University at Albany, SUNY and were compensated with course credit. Novice musicians were defined as those who have little to no experience in music expertise of any kind. All participants had self-reported normal or corrected-to-normal vision.

**Materials and Design**

The materials for this experiment consisted of 108 music scores. The music scores were three staves of piano sheet music from the Baroque and Classical music eras. The composers were lesser known composers from these eras to eliminate familiarity bias among experts. Fifty-four of these images were classified as “visually simple” and the other fifty-four were classified as “visually complex”. Complexity was defined as the amount of ink on the page (determined by number and types of notes, the use of accidentals, etc.). Because previous expertise studies involving music have had varied definitions of complexity manipulations, ratings were collected from five participants prior to the trial (none of whom participated in the main experiment) to validate the present definition of “amount of ink on the page”. Participants were instructed to rate the perceived visual complexity of music scores on a Likert scale ranging from one to seven, one being not complex and seven being very complex, and to ignore any technical details about the music. The images that were validated as being simple all had a collective rating that was under three and the images that were validated as complex all had a collective rating that was five or above. Only images that fit these criteria were used in the main study to ensure strong differences between the visually simple and visually complex images. Therefore, this study employs a visual search paradigm with a 2 (Complexity: Visually Complex or Visually Simple) x 2 (Expertise group: Experts or Novices) design. Complexity was within-subjects and expertise was between-subjects.
On each trial in the experiment, a target bar was presented above the music score, and this target bar was identical to one of the bars within the score. To select the target bars, we randomly selected three bars from each music score (one bar was selected from each of the three lines in the score) to serve as possible target bars. Across participants, we counterbalanced which of the three possible target bars was designated the search target, and the remaining two bars were designated as distractors. Therefore, a bar of the music was either a target region or distractor region based on the context of the presentation. In this way, a bar served as a control for itself since the bar could be analyzed as either a target or a distractor based its presentation context. Each participant saw a given music score only once, and there was an equal chance of the target being located in each of the three lines.

**Apparatus**

Participants’ eye movements were monitored using an SR Research Eye Link 1000 Plus system with high spatial resolution and a sampling rate of 1000 Hz. Participants were fitted on a head and chin rest to stabilize the head, and then calibrated to track the right eye, though viewing was binocular. The gaze-position calibration was less than 0.5° for all participants. Visual stimuli were displayed on a 24-inch Asus CG248QE computer screen with a resolution of 1920 x 1080 pixels. The screen was seventy centimeters away from the participant. Images were presented on a white background. The target image was located above the music score on the screen. While there were slight variations in the sizes of the music score and target images, the distance between the center of the music score and the center of the target image remained at a fixed distance. The music score ranged in size from a minimum of 345 pixels to a maximum of 650 pixels in height. A gamepad was used to collect responses.

**Procedure**
At the beginning of the trial, participants looked at a fixation cross, which was centered at the location where the target image subsequently appeared and pressed a button to initiate each trial. Once the participant looked at the cross and pressed the button on the gamepad, the trial began. During each trial, the target bar appeared above the music score, and remained there throughout the trial, and the participants had to locate the target bar within the music score. When they located the target bar, they had to fixate on the target within the music score and press the button on the gamepad. Once the button was pressed, their fixation location was recorded, and the study immediately moved on to the next trial. There was a time-out limit of two minutes for each trial, but all participants completed all the trials within this time limit. Because the search template was presented at the same time as the music score, the participants could look back and forth between the two images as much as they wanted. This sequence repeated for five practice trials, which were followed by one hundred and eight experimental trials.

**Results**

The main purpose of this study was to investigate the differences in expert musicians’ and novices’ ability to detect relevant regions of an intricate music score using a novel visual search paradigm. In this section, global eye tracking measures and local dwell-based eye movement behavior will be reported.

Regarding reaction time and all eye movement measures reported below, all inaccurate trials were removed from the analysis. Overall, experts were more accurate than the novices, though differences in reaction time between experts and novices were not statistically significant for accurate trials. The accuracy and RT measures did not show any interactions between complexity and expertise, though participants were more accurate but slower in the complex condition, compared to the simple condition (see Table 1 for full statistics).
Global Eye Movement Measures

Experts had longer fixation durations and shorter saccade amplitudes than novices. Though there were no expertise x complexity interactions for either of these measures, Table 1 shows that the complex condition had higher accuracy and slower reaction times than the simple condition across both participant groups, indicating a speed-accuracy tradeoff for the complexity manipulation. Additionally, the complex condition had shorter fixation durations, and differences in saccade amplitudes between the complex and simple conditions were not statistically significant. There were no significant complexity x expertise interactions for these measures.

The expertise effects for the global measures were as predicted by our hypotheses for the accuracy, but the experts’ pattern of longer fixation durations and shorter saccade amplitudes warranted further analysis. A possible explanation for this pattern of results is that experts employ qualitatively different search strategies, such that they prioritize their attention to the search template region at the start of their trial relative to novices. This would account for the longer fixation durations, and since they would not need to return back to the search template as often, would also account for the shorter saccade amplitudes. To examine this hypothesis, dwell-based analyses on local eye movement measures were performed for the search template region, as well as for comparisons between the target and distractor region.

Dwell-Based Analyses

In order to examine region effects, several local measures of eye movements were analyzed across the three different regions: the search template region (i.e., the bar presented above the music score), the target region (i.e., the location of the target within the score), and the distractor region (see Figure 1). For each of these three regions (i.e., the search template, target,
and distractor regions), the following local eye movement measures were examined: 1) first dwell duration (i.e., the sum of the durations of all the fixations during the first run in the interest area), 2) total dwell duration (i.e., the sum of the durations of all the fixations during the trial within the area of interest), and 3) number of dwells (i.e., the number of times a participant’s eyes entered the area of interest and then left). Dwell count was used instead of probability of a re-fixation in the interest area; Since the probability of a re-fixation was always extremely likely, dwell count was a more informative measure of eye movement behavior.

**Search Template.** To examine search behavior for the search template region, a 2 (complexity) x 2 (expertise) ANOVA was conducted for first dwell, total dwell, and dwell count. Table 2 shows that there are expertise effects for all three measures such that experts spend more time in the search template than novices in the first dwell and total dwell but have smaller dwell counts. Overall, participants had longer dwell times in the complex condition than the simple condition for both first dwell and total dwell, but no significant differences in dwell count. The first run measure indicated that experts prioritize time spent in the search template at the start of the trial much more than novices do, as well as for the rest of the trial, as indicated by the total run analysis, though the difference is not as drastic. While experts spend more time in the search template than novices do, their dwell count indicates that they do not return to the search template as often as the novices did. This supports the hypothesis that experts use qualitatively different search strategies that prioritize the search template.

There were also expertise x complexity interactions for first dwell, total dwell, and dwell count. Specifically, the difference between the complex and simple trials in the search template was greater for experts than novices for first dwell and total dwell. There was a crossover interaction for dwell count such that experts had a higher dwell count for the complex condition,
whereas novices had a higher dwell count for the simple condition. Overall, experts had shorter first dwell and total dwell durations for the simple condition than the complex condition, and while novices do as well, this difference is not as stark. Post-hoc t-tests also indicate that both experts and novices spend less time in the simple condition for first dwell (Experts: $t(29) = 3.57, p < 0.01$; Novices: $t(29) = 2.16, p < 0.05$) and total dwell (Experts: $t(58) = 7.21, p < 0.001$; Novices: $t(29) = 4.90, p < 0.001$). Post-hoc t-tests for dwell count show that experts have significantly fewer dwells in the simple condition ($t(29) = 5.32, p < 0.001$), but novices have more dwells in the simple condition ($t(29) = 4.33, p < 0.01$). The interactions between complexity and expertise in the dwell-based measures provide further evidence that experts approach this task using qualitatively different strategies from novices.

**Target Region vs. Distractor Region.** I contrasted the target region and the distractor region within the music score, to test the hypothesis that experts would be better than novices at focusing on relevant rather than irrelevant regions of the music score.

Because the target bar had an equal chance of being in one of the three lines of the music score, the same bar served as either a target or a distractor depending on its context. Therefore, there were theoretically two possible distractors for each trial (i.e., the bars from the other two lines that were not used as the target) that could be analyzed. Out of the two possible distractors, the distractor with the quickest time to first fixation was used for the region analysis. We used this approach because there was more data available for the first-fixated distractor bar than the second-fixated bar. Participants were less likely to fixate the second distractor bar because they typically ended the trial as soon as they found the target.

A 2 (region: target, distractor) x 2 (expertise: expert, novice) x 2 (complexity: complex, simple) ANOVA was conducted, with complexity and region as within-subjects variables, and
expertise as a between-subjects variable. Overall, there were no main effects of expertise, no complexity x expertise interactions, and no three-way interactions.

However, there was a significant region x expertise interaction for first dwell, such that the difference between dwell times for the target region as compared to the distractor region was larger for experts than for novices. This finding supports our hypothesis that experts would be better than novices at rapidly detecting the relevant region. Post hoc t-tests also show that both experts and novices have significantly longer dwells in the target region (Expert: \( t(29) = 13.66, p < 0.001 \); Novice: \( t(29) = 15.19, p < 0.001 \)) relative to the distractor region. Also, as indicated by Tables 3 and 4, and Figure 2, there was also a main effect of region such that participants spent more time in the target region for first dwell, and total dwell, and have a higher dwell count in the target region, indicating a priority of the target region compared to the distractor\(^1\).

Regarding complexity, there were significantly shorter and fewer dwells in the simple condition than the complex condition, as indicated by all three dwell-based measures. Overall, it seems that all participants spend less time processing the simple music scores, as indicated by dwell time and dwell count. Participants also show an interaction between region and complexity for total dwell and dwell count. Post hoc t-tests show that participants spend significantly more time in the target region (Complex: \( t(59) = 13.45, p < 0.001 \); Simple: \( t(59) = 14.33, p < 0.001 \)) and significantly higher dwells in the target region than the distractor region for both complexity

\(^1\)To address concerns that the first fixated distractor bar would always be the first-line distractor, the same set of target vs. dwell analyses were performed for the first line only. The pattern of the results for first line targets and distractors follows the exact same pattern of behavior as the overall analysis (significant main effects of region for first dwell, total dwell, and dwell count; significant region x expertise interactions for first dwell, all \( p \)'s < 0.01). Experts always prioritize the target region than the distractor region to a greater magnitude than novices do, and the first-fixated distractor method of analysis does not present a confound.
conditions (Complex: $t(59) = 14.16, p < 0.001$; Simple: $t (59) = 15.25, p < 0.001$). The fact that there were complexity effects in the local dwell-based measures but not the global measures is an avenue for future research.

**Discussion**

In this study, the visual component of expertise within the domain of music was examined. A novel visual search task was developed in which a search template was presented simultaneously with the search array beneath it, and participants had to locate the search template within the array. Strong expertise effects were shown in the study, including experts’ superior ability to discriminate between relevant regions of the search array compared to novices, as well as evidence suggesting that experts were able to develop precise representations of complex search templates in order to guide their search.

The overall results are consistent with the idea that experts and novices use qualitatively different search strategies when completing their task. This idea is supported by evidence that suggests that experts allocated their time and attention to the different regions in a different manner from novices. For example, experts spent significantly longer time looking at the search template at the start of the trial, as compared to novices, while also having significantly shorter saccade amplitudes during the trial. First dwell duration analyses suggest that experts were allocating more attention to the search template at the start of the trial, which is potentially because they were taking time to encode chunks or templates to assist them in learning the search template. This explanation would also account for why experts needed less time to identify the target region as a relevant area and is also supported by the fact that experts have smaller dwell counts and shorter saccade amplitudes overall and in the target region, since experts likely did not need to move their eyes back to the search template as often as novices. Thus, experts did not
need to compare the search template and target region as often, as indicated by various eye movement measures, but they still solved the task more accurately as compared to novices.

In addition to the qualitatively different search strategies, another difference experts exhibited was their superior ability to detect relevant regions of the music score compared to novices. Experts spent more time in the search template (which is a relevant region), at the beginning of the trial, and experts were also better able to ignore irrelevant areas, which is similar to findings in the chess domain (Sheridan & Reingold, 2014; Bilalić et al., 2010a, 2012; DeGroot & Gobet, 1996) and the medicine domain (Kundel et al., 2008; Donovan & Litchfield, 2013; Wood et. al, 2013). This early discrimination of relevant areas (as revealed by the first dwell duration and dwell count measures) by experts across different domains is in strong support of the idea that experts have a strong perceptual advantage within their domain as compared to novices.

These perceptual advantages that experts can exhibit could be explained by chunking and template theories (Chase & Simon, 1973; Gobet & Simon, 1998). Experts are able to form and use chunks or templates based on extensive practice and building their knowledge base within a domain. In the present study, it appears that the experts’ behavior was indicative of their use of chunks or templates, given their qualitatively different search strategies and allocation of attention. To the extent that experts are able to store the search template using chunks or templates, they will not be as dependent on referring back to it during their subsequent search of the music score. The experts in our study relied on this strategy in order to complete this visual search task, resulting in superior performance and enhanced ability to detect relevancy, which is similar to expertise findings in other domains, including chess (Chassy, 2013; for a review, see Reingold & Sheridan, 2011). Also, previous studies regarding expertise in music have found that
experts have larger eye-hand spans in reading sheet music (Wristen, 2005; Penttinen, Huovinen, & Ylitalo, 2015; Truitt, Clifton, Pollatsek, & Rayner, 1997), process sheet music in a more holistic manner than novices do, (Drai-Zerbib, Baccino, & Bigand, 2011; Wong & Gauthier, 2010a, 2010b; Penttinen, Huovinen, & Ylitalo, 2013), and do not focus on the individual notes (Kinsler & Carpenter, 1995; Gilman & Underwood, 2003). The present study extends these findings by providing evidence that experts use chunks or templates in order to store the search template into their working memory.

In addition to the above theoretical contributions, the present study also has several methodological contributions. The nature of the visual search task itself is novel, because the search template was presented for the entirety of the trial, to allow the participants to make comparisons as they performed their search of the array. Related to this paradigm, there is another task in the literature called a comparative visual search paradigm. It is similar to the paradigm used in this study in that participants must compare two images to detect similarities or differences between the two (Humphrey & Lupker, 1993). Comparative visual search tasks have also been combined with eye tracking methodology to examine search behavior, and often use complex stimuli (Pomplun et al., 2001). Thus, many properties of the comparative visual search paradigm and the paradigm used in this study are similar. The present study’s task allowed me to investigate properties of search behavior while using search templates that were complex, rich in domain-specific detail, and more ecologically valid than many other tasks within the visual search literature (for a review, see Wolfe & Horowitz, 2017). Furthermore, since the domain of music was isolated to simply its visual component, the task was accessible to both experts and novices, which allowed me to compare performance across groups, since confounds such as tempo and technical detail were made irrelevant. Finally, using the domain of music to
investigate expertise was advantageous due to the fields’ large and diverse population. Using music allowed for an unprecedented large sample size of experts without compromising on the standards required to qualify as an expert.

Based on the results of this study, a possible future experiment could manipulate similarity of the features of sheet music to explore how this variable interacts with visual complexity. In the present study, the simple condition unexpectedly yielded worse performance than the complex condition in both participant groups. One possible explanation for these results could be a similarity effect, since high levels of visual similarity in the simple condition could have made it more difficult to distinguish between the relevant and irrelevant regions. Similarity effects have been discovered in other domains, including visual search tasks and reading in Chinese (e.g., Yu et al., 2018). It is known that visual span decreases as both task difficulty increases and feature similarity increases (Duncan & Humphreys, 1989; Pomplun, Reingold, & Shen, 2001; Carmody, Nodine, & Kundel, 1981). Evidence from saccade analyses show that eye movements are not random but instead are directed toward regions that share characteristics with the target (Motter & Belky, 1998). Additionally, Rayner and Fisher (1987a, 1987b) found that fixation durations in visual search tasks increase as similarity between distractors and targets increases. Thus, further research with similarity manipulations could help to pinpoint why the simple condition was more challenging than the complex condition in the present study.

Building upon the above findings, other avenues for future research could further illuminate the differences in performance and search strategies across experts and novices in the music domain. A planned follow up experiment is to evaluate how performance in both experts and novices if the search template were to be presented before the search array. This would force participants to memorize as much detail about the search template as possible before performing
the task, therefore eliminating the ability to compare the two images, which would allow us to investigate the role of memory, specifically chunking, in experts and novices. Additionally, while a stark contrast in expertise level was desired for the current study, an interesting idea would be to include an intermediate group, such that expertise could be studied on a continuum, as opposed to absolutes.

The present study contributes several important findings, both methodological and theoretical. In this study, a unique visual search paradigm was developed that allowed us to compare expert and novice performance within a domain that is rich in complexity, detail and syntax. Experts are also superior in discriminating between relevant and irrelevant areas of a visual field within their domain. Evidence showed that experts likely use chunks to commit the search template to working memory before examining the search array, which replicates expertise findings from other domains and also accounts for the relevancy effect. Overall, this study suggests that experts use chunking to aid their superior performance in domain-specific tasks.
References


Table 1

Means for the Global Eye Movement Measures. Standard Error of the Mean shown in parentheses.

<table>
<thead>
<tr>
<th></th>
<th>Expert</th>
<th>Novice</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complex</td>
<td>Simple</td>
<td>Complex</td>
</tr>
<tr>
<td>Accuracy (proportion of correct trials)</td>
<td>0.85 (0.02)</td>
<td>0.83 (0.02)</td>
<td>0.78 (0.02)</td>
</tr>
<tr>
<td>Reaction Time (ms)</td>
<td>11217 (652)</td>
<td>10900 (653)</td>
<td>10972 (504)</td>
</tr>
<tr>
<td>Saccade Amplitude (°)</td>
<td>2.34 (0.05)</td>
<td>2.37 (0.05)</td>
<td>2.65 (0.05)</td>
</tr>
<tr>
<td>Fixation Duration (ms)</td>
<td>275 (7)</td>
<td>281 (8)</td>
<td>233 (5)</td>
</tr>
</tbody>
</table>

For all $F$ tests, $df$ for $F = (1, 58)$
Table 2

Local Dwell-Based Eye Movement Measures for the Search Template only. Standard Error of the Mean shown in parentheses.

<table>
<thead>
<tr>
<th>Complexity</th>
<th>Expert</th>
<th>Novice</th>
<th>Significance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex</td>
<td>Complex</td>
<td>Simple</td>
<td>Simple</td>
</tr>
<tr>
<td>First Dwell</td>
<td>2870</td>
<td>2512</td>
<td>855</td>
</tr>
<tr>
<td></td>
<td>(261)</td>
<td>(199)</td>
<td>(81)</td>
</tr>
<tr>
<td>Total Dwell</td>
<td>4811</td>
<td>4188</td>
<td>2711</td>
</tr>
<tr>
<td></td>
<td>(329)</td>
<td>(294)</td>
<td>(182)</td>
</tr>
<tr>
<td>Dwell</td>
<td>3.66</td>
<td>3.27</td>
<td>3.92</td>
</tr>
<tr>
<td>Count</td>
<td>(0.19)</td>
<td>(0.17)</td>
<td>(0.24)</td>
</tr>
</tbody>
</table>

For all $F$ tests, $df$ for $F = (1, 58)$.
Table 3

*Local Dwell-Based Eye Movement Means Comparing the Target Region and the Distractor Region. Standard Error of the Mean in Parentheses.*

<table>
<thead>
<tr>
<th></th>
<th>Expert Target</th>
<th>Distractor Target</th>
<th>Novice Target</th>
<th>Distractor Target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complex</td>
<td>Simple</td>
<td>Complex</td>
<td>Simple</td>
</tr>
<tr>
<td><strong>First Dwell</strong></td>
<td>565 (25)</td>
<td>531 (19)</td>
<td>282 (8)</td>
<td>276 (11)</td>
</tr>
<tr>
<td><strong>Total Dwell</strong></td>
<td>2053 (180)</td>
<td>1745 (147)</td>
<td>496 (32)</td>
<td>458 (27)</td>
</tr>
<tr>
<td><strong>Dwell Count</strong></td>
<td>3.20 (0.16)</td>
<td>2.89 (0.14)</td>
<td>1.58 (0.05)</td>
<td>1.51 (0.04)</td>
</tr>
</tbody>
</table>
Table 4

Significance results for the three-way ANOVA comparing the target and distractor regions. Significant results are shown in bold.

<table>
<thead>
<tr>
<th></th>
<th>Region</th>
<th>First Dwell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>$F = 1.24$, $p = 0.27$</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>$F = 8.06$, $p &lt; 0.05$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise</td>
<td>$F = 15.20$, $p &lt; 0.01$</td>
<td></td>
</tr>
<tr>
<td>Region x Complexity</td>
<td>$F = 0.04$, $p = 0.84$</td>
<td></td>
</tr>
<tr>
<td>Expertise x Complexity</td>
<td>$F = 0.01$, $p = 0.92$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise x Complexity</td>
<td>$F = 3.08$, $p = 0.08$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Region</th>
<th>Total Dwell</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>$F = 0.53$, $p = 0.47$</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>$F = 62.94$, $p &lt; 0.01$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise</td>
<td>$F = 2.25$, $p = 0.15$</td>
<td></td>
</tr>
<tr>
<td>Region x Complexity</td>
<td>$F = 30.63$, $p &lt; 0.01$</td>
<td></td>
</tr>
<tr>
<td>Expertise x Complexity</td>
<td>$F = 0.11$, $p = 0.74$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise x Complexity</td>
<td>$F = 1.31$, $p = 0.26$</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Region</th>
<th>Dwell Count</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Expertise</td>
<td>$F = 0.36$, $p = 0.55$</td>
<td></td>
</tr>
<tr>
<td>Complexity</td>
<td>$F = 41.78$, $p &lt; 0.01$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise</td>
<td>$F = 0.002$, $p = 0.97$</td>
<td></td>
</tr>
<tr>
<td>Region x Complexity</td>
<td>$F = 19.67$, $p &lt; 0.01$</td>
<td></td>
</tr>
<tr>
<td>Expertise x Complexity</td>
<td>$F = 0.952$, $p = 0.34$</td>
<td></td>
</tr>
<tr>
<td>Region x Expertise x Complexity</td>
<td>$F = 0.02$, $p = 0.89$</td>
<td></td>
</tr>
</tbody>
</table>

For all $F$ tests, $df$ for $F = (1, 58)$
Figure 1. A visual example of the three regions from the dwell-based analyses. The target region, or correct answer, was the matching bar of the search template, which is the section of music located above the music score. The example of a distractor region was the control for the target as it served as a possible correct answer in a different context (see Method section for a more detailed explanation).
Figure 2. Target region versus first-fixated distractor region (see Results section for a more detailed explanation) comparisons for a) first dwell (ms), b) total dwell (ms), and c) dwell count measures, collapsed across complexity. Error bars represent standard error of the mean.