Expertise in Processing Musical Notation

An Investigation of chunking, working memory, and eye movements

Lucas Lörch

Inaugural Dissertation

Submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences at the University of Mannheim

Thesis Submission:

December 17th, 2021

Thesis Defence:

April 1st, 2022

Supervisors:

Prof. Dr. Stefan Münzer

Prof. Dr. Erkki Huovinen

Dean of the Faculty of Social Sciences:

Prof. Dr. Michael Diehl

Academic Director of the CDSS:

Prof. Dr. Thomas Gschwend

Thesis Reviewers:

Prof. Dr. Stefan Münzer

Prof. Dr. Erkki Huovinen

Prof. Dr. Beatrice Kuhlmann

Defense Committee:

Prof. Dr. Stefan Münzer

Prof. Dr. Erkki Huovinen

Prof. Dr. Beatrice Kuhlmann

Contents

Abstract

Expertise develops as a result of deliberate practice, that is, as a result of activities that are designed with the primary purpose of attaining and improving skills. As a by-product of deliberate practice, experts acquire superior memory for information from their domain. This superior memory is enabled by chunking and by the rapid access to long-term memory. Chunking denotes a process of searching for known sequences in encoded stimuli and, if found, recoding them as single units. The present work investigated expertise in the cognitive processing of musical notation. More specifically, the three studies of the present work respectively examined (1) chunking in short-term memory for musical note symbols, (2) working memory for musical note symbols in the context of the Time-Based Resource Sharing model, and (3) eye movements during sight-reading, i.e., during the instrumental performance of unfamiliar notated melodies.

The first study was based on the finding that chunking in short-term memory for musical notation is supported by systematic compared to random tonal structures. To extend this finding, I analyzed which specific features of systematic tonal structures support chunking. To this end, I designed a serial recall task with single quarter note symbols as memoranda. The number of memoranda and the tonal structure of the note sequences were varied within-participants. Chunking-supportive tonal structures resided in a clearly recognizable tonal context, i.e., contained only notes from the scale defined by the first note of the sequence, and contained meaningful melodic cells with clear labels. In the first simple span experiment, these melodic cells were major triads, while in the second, they were based on authentic cadences. Note sequences that had a chunking-obstructive tonal structure did not reside in a clearly recognizable tonal context and did not contain melodic cells with clear labels. Participants' musical expertise was indicated by the Gold-MSI questionnaire. I expected that participants on a higher expertise level would recall notes more accurately, that chunking-supportive sequences would be recalled more accurately, and that the advantage in chunking-supportive sequences would be larger for participants on a higher expertise level. Analyses provided evidence for both expected main effects. Probably due to a ceiling effect, the expected interaction was only found in the long sequences containing authentic cadences. I conclude that, given that a note sequence has a systematic tonal structure, the combination of a clear tonal context with meaningful melodic cells might support chunking.

The second study investigated expert working memory in the context of the Time-Based Resource Sharing (TBRS) model. This theoretical model assumes that working memory involves a rapid

switching between maintaining already encoded information and processing new stimuli. It was translated into a computational model that simulates recall in complex span tasks. In a recent version of this computational mode (TBRS*C), a chunking mechanism was added. Although various theories, such as template theory or long-term working memory theory, conceptualize working memory as being inherently influenced by expertise, TBRS does not account for expertise differences. Thus, using a newly developed complex span task for musical notation, I investigated how expertise might be conceptualized in the context of the TBRS model.

In the task, participants had to memorize single quarter notes for serial recall of pitch. In between the visual presentation of each note symbol, they had to perform a distractor task, i.e., they had to sight-read a short melody. To manipulate the potential for chunking, the tonal structure of the sequences of to-be-remembered notes was varied. The sequences contained melodic cells that were either more meaningful (major triads) or less meaningful (arbitrary trichords). The complex span task was completed by a music student and a hobby musician sub-sample. Using the Gold-MSI questionnaire, a higher-expertise and a lower-expertise group was created in both sub-samples. Recall was simulated using TBRS*C. Estimates for certain parameters, namely the strength of encoding (*R*), the chunk search duration (*cSD*), the probability of chunk retrieval (*PCR*), and the time used for distractor processing (*Ta*), were compared between expertise groups. I expected to find evidence for experts' more rapid LTM access (i.e., a larger *R* and a smaller *cSD*) and more reliable chunking (i.e., a larger *PCR*). This expectation was supported in the hobby musician sub-sample. In addition, it was found that lower-expertise hobby musicians might have spent less time on the processing of distractors. They might have compensated inefficient memory processes by increasing refresh times during the distractor task. Parameter differences in the music student sub-sample were marginal, which probably was due to a ceiling effect. I conclude that expert working memory in the TBRS model might be conceptualized by rapid LTM access which increases refresh times, and by reliable chunking which increases the efficiency of refreshing.

In the domain of musical sight-reading, there is some initial evidence that sight-reading accuracy, i.e., how accurately notes are performed on an instrument, might be associated with eye movements. This was found for measures such as pupil size, the size of the eye-hand span, or overall gaze duration. However, a systematic investigation of the association between eye movements and performance accuracy during sight-reading is still missing. I developed a software tool, the *MidiAnalyzer*, to assess the accuracy of experimental *MIDI* (Musical Instrument Digital Interface) data. Using eye movement and MIDI data from the sight-reading task that was embedded in the complex span task, I investigated how the number and duration of fixations during sight-reading was associated with the accuracy of pitch and note onset. In the statistical analyses, I controlled for the effect of certain covariates, namely musical expertise, practice, and features of notes. Musical expertise was indicated by the Gold-MSI questionnaire. Practice was represented by the number of trials participants had already completed. Rhythmical features of notes were accounted for by the type of note pair (eighth-eighth, eighth-quarter, quarter-quarter, quarter-eighth) that participants were currently reading. Results showed that the number of fixations was negatively associated with the accuracy of note onset in both the reading of the whole melody and the reading of note pairs. I conclude that reading with fewer fixations might require less cognitive resources as eye movement planning and information integration is less demanding. The saved cognitive resources might be used to increase the accuracy of the instrumental performance.

In summary, the present work provided evidence for experts' superior processing of musical notation in short-term memory, working memory, and sight-reading. It highlighted the central role of chunking in expert memory, and showed that chunking is a process with both universal and domain-specific aspects. In addition, the present work demonstrated that unraveling the role of eye movements in accurate sight-reading is substantial for the comprehensive understanding of this skill.

1. Articles

The present thesis is based on three manuscripts of which two are published and one is submitted for publication:

Lörch, L. (2021). Chunking in tonal contexts: Information compression during serial recall of visually presented musical notes. *Psychology of Music*. https://doi.org/10.1177/03057356211013396

Lörch, L., Lemaire, B., Portrat, S. (2021). A computational simulation of expertise in the Time-Based Resource Sharing model. Manuscript submitted for publication. *Cognitive Computation.*

Lörch, L. (2021). The association of eye movements and performance accuracy in a novel sight-reading task. *Journal of Eye Movement Research*, 14(4). https://doi.org/10.16910/jemr.14.4.5

In the following, I will explain the overarching theoretical foundation for these three manuscripts and provide a general discussion. For the sake of logical structure, the manuscripts are located between the general introduction and the general discussion, not in the Appendix.

2. Introduction

Since the pioneering work of Binet (1894), the exceptional performances of experts have stirred the interest of psychologists. As described by Ericsson and Charness (1994), prior to the scientific investigation of the phenomenon, exceptional performances have classically been considered as resulting from an innate talent, from a gift provided by some divine source. This view was overthrown, though, by the recent decades of scientific research, which revealed that individual differences in basic abilities are poor predictors of performance (Ericsson et al., 1993). Rather, the amount of individualized training activities, and environmental factors that foster them are associated with individual differences in performances (Ericsson & Lehmann, 1996).

With this insight, research on expertise gained a much broader relevance. First, if expertise is not merely based on innate talent, it is not restricted to some gifted individuals. Anyone is a potential expert. This is especially true, as modern society offers broad access to training and education. For example, most citizens in industrialized nations are highly trained in their respective language, having practiced speaking, reading, and writing for hundreds of hours. Second, if training and experience play a role in the development of expertise, this gives rise to new research questions, such as "How do cognitive processes change with training and experience?". This moves the focus of interest from the specific characteristics of experts to the general functioning of the human mind. Gobet (2019) stated that expertise research is "aimed to address general questions of cognition, using expertise as a means" (Gobet, 2019, p. 49). Lastly, the understanding of expertise might tell us something on how the environment can be adapted to individuals and thereby, foster them with maximal efficiency. Researchers nowadays not only want to understand experts but want to use their discoveries to derive suggestions how to structure learning environments, teaching methods, and everyday technological aids. Gobet (2012) claimed that "studying experts can illuminate which training methods are efficient and which ones are not, again with applications for nonexperts and even for education in general" (Gobet, 2012, p. 951).

The idea that practice and training are essential for expertise was mainly advanced by the studies on chess expertise by de Groot (1946) and Chase and Simon (1972). De Groot (1946) found that the accuracy of the recall of briefly presented chess boards was a function of chess skill. Chase and Simon (1972) replicated this finding and extended it by showing that the effect of skill vanished when using randomly rearranged chess boards. They concluded that experts' vast knowledge enabled them to perceive meaningful patterns of chess pieces. By encoding these meaningful patterns instead of individual pieces, they increased their short-term memory capacity, a process which they termed *chunking*.

The theory of Chase and Simon (1972) has sparked a lot of interest and has led to the development of further theories of skilled memory, such as *template theory* (Gobet & Simon, 1996), *skilled memory theory* (Chase & Ericsson, 1982; Ericsson & Staszewski, 1989), and *long-term working memory theory* (LT-WM; Ericsson & Kintsch, 1995). These theories have in common that they describe another advantage of experts besides their use of known meaningful patterns, namely the rapid access to information in long-term memory (LTM).

In addition, chunking has been incorporated into various computational models such as CHREST (Gobet et al., 2001), ACT-R (Anderson et al., 2004), and TBRS*C (Portrat et al., 2016). In these models, verbal theories are translated into computer programs. These programs simulate the behavior of participants in certain cognitive tasks. To validate the models, their simulations are compared to empirical data. The advantage of computational models is that they provide a new perspective on participants' behavior, as they describe it with theoretically meaningful parameters. Moreover, they allow to derive new theoretical assumptions. Parameters in the model, for example the speed of processing a stimulus, can be easily varied and simulations then reveal how these parameters might affect participants' behavior.

The research introduced in the recent paragraphs refers to cognitive tasks in which information is encoded, maintained, and recalled. Musical notation, though, is often not processed to memorize it, but to perform it on a musical instrument. If the person is unfamiliar with the performed melody, this activity is called *sight-reading* (Ericsson & Lehmann, 1996; Kopiez & Lee, 2006; Wolf, 1976). Waters et al. (1998) stated that "sight-reading is a complex transcription task involving a series of overlapping perceptual, cognitive, and motoric processes" (Waters et al., 1998, p. 123). The understanding of the psychological processes involved in sight-reading is both of practical interest to understand the acquisition of the skill and of theoretical importance for the more general understanding of expertise (Waters et al., 1998).

During sight-reading, the eyes move across a printed score and, through movements on a musical instrument, the note symbols are translated into specific sounds. The movement of the eyes is based on an action schema (Land & Furneaux, 1997). During their musical training, musicians learn how to move their eyes to provide the required continuous stream of information. Thus, moving the eyes during sight-reading denotes a highly relevant skill by itself. This notion has led to numerous studies on the topic in the recent decades (reviewed by Puurtinen, 2018). However, potentially due to methodological inconsistencies, "the collective data in the field are inconclusive and rather confusing" (Madell & Hébert, 2008, p. 166).

Based on these considerations, the main goal of the present work was to investigate expertise in the cognitive processing of musical notation. I focused on two cases in which musical notation is cognitively processed. These were the memorization of a single note symbol in a recall task, and the translation of a notated melody into movements on an instrument during sight-reading. With respect to these cases, the more specific goals of the present work were to investigate (1) expertise in the short-term and working memory for musical notation, and (2) the role of eye movements during musical sight-reading. Table 2.1 depicts how the three manuscripts of the present work relate to these research goals. The theoretical and methodological foundation of the manuscripts as shown in Table 2.1 will be introduced in the following. To increase readability, I will henceforth use short titles to refer to the respective manuscripts, namely *Chunking in tonal contexts*, *Expertise in the TBRS model*[1], and *The association of eye movements and sight-reading accuracy*.

2.1 A psychological understanding of expertise

2.1.1 Development of expertise

In their seminal work, Ericsson et al. (1993) studied how the reported amount of practice related to the skill level of violinists. There were four groups of participants: (1) music students who were nominated by their professors as excellent violinists, (2) music students who were nominated by their professors as good violinists, (3) music education students, and (4) professional violinists engaged at renowned symphony orchestras. Participants indicated relevant biographical information, estimated the hours they spent practicing the violin up to the age of 18, and rated a list of everyday and musical activities with respect to the time invested in it in the previous week, its relevance for violin skills, the effort it required, and how enjoyable it was. In addition, participants completed a seven-day diary to record their activities in this time.

[1] This manuscript was co-authored. For the sake of consistency, I nevertheless use the singular (e.g., "I investigated...") when talking about this manuscript in the general introduction and discussion.

Table 2.1

Overview of methods and theoretical foundations of the present manuscripts

Research goal	Manuscript	Research question	Paradigm	Design	Hypotheses
Investigate expert memory for musical note symbols	Chunking in tonal contexts: information compression during serial recall of visually presented musical notes	Which tonal features of melodic sequences support chunking in STM?	Simple span task	Within-participant factors *tonal structure* (meaningful vs. arbitrary) and *list length* (short vs. long)	Interaction of musical expertise and tonal structure: additional advantage for participants on a higher expertise level in chunking-supportive sequences
	A computational simulation of expertise in the Time-Based Resource Sharing model	How can expert memory be conceptualized in the Time-Based Resource Sharing model?	Complex span task	Within-participant factor *tonal structure* (major triads vs. arbitrary trichords)	Faster encoding (larger *R);* faster chunk search (smaller *cSD*); more reliable chunking (larger *PCR*) of experts
Investigate eye movements during sight-reading	The association of eye movements and performance accuracy in a novel sight-reading task	How are number and duration of fixations during musical sight-reading associated with accuracy of pitch and note onset?		Within-participant factor *type of note pair* (eighth-eighth vs. eighth-quarter vs. quarter-eighth vs. quarter-quarter)	Exploratory analyses on the level of melodies and on the level of different note pairs

Results revealed that solitary practice was judged as most relevant for violin skills. The week-long diary showed that the excellent and good violinists practiced almost three times longer than the music education students. The accumulated hours of practice up to the age of 18 increased from music education students to good violinists to excellent violinists with the latter group indicating a similar amount of practice than the professional musicians (Ericsson et al., 1993).

The results obtained by Ericsson et al. (1993) promoted the theoretical development in expertise research as it led to the concept of *deliberate practice*. The concept assumes that expert performance cannot merely be explained by the amount of experience as there are many activities, such as work or play, that provide few opportunities for the systematic improvement of certain kinds of skills. So, rather than mere experience, the amount of deliberate practice, i.e., "activities invented with the primary purpose of attaining and improving skills" (Ericsson et al., 1993, p. 367) is central for expertise development. For this practice to be most effective, the person should be motivated to attend to the task, the task should be oriented at the person's current state of knowledge, and the person should receive immediate informative feedback on their performance. A crucial theoretical assumption associated with the concept of deliberate practice is the *monotonic benefits assumption*, which claims that there is a monotonic positive relationship between the amount of deliberate practice and an individual's acquired level of performance. However, there are several hindrances or constraints an individual must overcome to engage in deliberate practice. Deliberate practice requires time, energy, and money for teachers and training material (resource constraint), it is not inherently motivating (motivational constraint), and can be executed only for a limited time as it leads to exhaustion (effort constraint) (Ericsson et al., 1993).

In the domain of music, the deliberate practice of a musical piece might comprise activities such as the variation of performance tempo, the targeting of specific errors, the isolation of musical dimensions such as pitch or rhythm, the focus on segments of a piece or on specific movements, and even non-playing strategies such as listening to recordings or silently reading a printed score (Mishra, 2019). As the practice is usually directly related to the demands of the respective task, the specifics of practicing might vary with instruments, genres, and musical activities (accompanying, sight-reading, conducting, etc.) (Lehmann et al., 2018). While deliberate practice in many domains is typically associated with solitary practice, practice in some musical genres might involve activities such as taking lessons, group performances, or practicing together with others (Mishra, 2019). Moreover, the quality of practice is at least as important for skill development as the duration (Lehmann et al., 2018; Mishra, 2019).

With regard to the time it takes to develop expertise, Ericsson et al. (1993) developed the so-called *ten year rule*. It evolved from the claim by Chase and Simon (1972) that it takes about 10,000 hours or ten years of practice to acquire the knowledge that is necessary to reach expert level in chess (see also Gobet, 2019). This initial statement together with further empirical data led Ericsson to derive the general statement that irrespective of the domain of expertise, “not even the most ‘talented’ individuals can attain international performance without approximately 10 years of preparation” (Ericsson & Lehmann, 1996, p. 278). The improvement of skill is generally assumed to be continuous without any evidence for abrupt leaps (Ericsson, 2018a). In the domain of music, analysis of musical works has shown that the ten year rule even applies to outstanding performers such as Mozart (Hayes, 1981) and The Beatles (Weisberg, 1999).

The progression through the thousands of hours of practice described in the ten year rule can be segmented into different phases, described by the three-phase model (Bloom, 1985). These phases comprise: (1) the first contact with the domain through playful activities often in childhood age, (2) the start of instruction and limited amounts of deliberate practice, and (3) the transition into full-time commitment. Ericsson et al. (1993) claimed the necessity to add a fourth phase to account for eminent performance, which involves unique and innovative contributions to the respective domain.

In the domain of music, Mishra (2019) developed a musical performance hierarchy to categorize the different stages on the path to musical expertise. The lowest stage is the *non-performer*, a person who has not made any efforts to systematically improve instrumental skills. The second stage, the *student musician*, addresses persons who have studied or are currently studying an instrument. This stage is followed by the *developing expert:* musicians who have demonstrated a certain level of skill but continuously strive to improve under the supervision of a teacher. A music major at a university would be a typical example for a person at this stage. The last two stages, the *expert musicians* and the *influential expert* respectively denote persons who publicly perform concerts which they prepare without supervision, and persons who make significant and influential contributions to the field (Mishra, 2019).

For the study *Chunking in tonal contexts* participants from a broad range of expertise levels were addressed. The resulting sample contained student musicians, developing experts, and expert musicians. For the other two studies, I specifically targeted developing experts and student musicians by recruiting music students and musically literate students of other subjects than music, respectively. As a label for the latter group, I used the term *hobby musicians*. In the literature, the term *amateur musicians* might be more common. However, in a strict sense, music students are amateurs as well, as they do not yet work as professional musicians. Thus, I deemed the term *hobby musicians* to be most appropriate.

2.1.2 Components of expertise

When a person engages in deliberate practice and progresses through the stages of expertise development, this results in the development of domain-specific *skills* and *knowledge*. Fadde and Jalaeian (2019) claimed that “deliberate practice includes the thousands of hours that performers typically invest in acquiring a body of declarative knowledge and mastering requisite technical skills” (Fadde & Jalaeian, 2019, p. 928). Thus, skills and knowledge can be considered the main components of expertise. While cognitively oriented expertise researchers often understand knowledge simply as the availability of information in LTM, skill denotes the practice-related superiority of goal-directed action. A rationally and intentionally controlled action can be seen as the manifestation of a specific skill (Stanley & Krakauer, 2013).

Typically, knowledge is classified in declarative and procedural knowledge with skill being independent from the former and identical with the latter. However, according to Stanley and Krakauer (2013), this is a misconception. Skill involves both declarative (or factual) knowledge and components that are not knowledge-based. Concerning the former, the authors claimed that “knowing what to do to initiate an action and knowing how to do something are both kinds of factual knowledge, factual knowledge required by skill possession” (Stanley & Krakauer, 2013, p. 5). In contrast, one component of skill that is not knowledge-based is *motor acuity*, i.e., the practice-related decrease of movement variability and increase of movement smoothness (Stanley & Krakauer, 2013).

For the present work, the relevant components of expertise were the skill to read and perform musical notation on an instrument and declarative knowledge on music theory. According to Lehmann (2005), music reading is one of the basic skills of musicians (see also Mishra, 2019). In many areas in which music is performed, musical notation is used to communicate complex musical information. The skill to perform notated melodies comprises action schemata for eye movements (Land & Furneaux, 1997) and for movements on an instrument as well as the knowledge on the association between note symbols and musical sound. Sight-reading is a special form of music reading in which the musician is not familiar with the performed melody.

In addition, the part of music theory that was relevant for the present work was how certain combinations of notes form chords and scales. As will be described in more detail below, declarative knowledge on this part of music theory might affect memory processes. For example, stimuli that entail expected musical patterns were found to be recalled more accurately (Lehmann et al., 2018; Mishra, 2019; Snyder, 2016). In summary, an expert in the present work was defined as someone who has deliberately practiced a musical instrument and while doing so, acquired declarative knowledge on music theory and the skill to perform unfamiliar notated melodies on an instrument. The former feature, declarative knowledge on music theory, was central for the manuscripts *Chunking in tonal contexts* and *Expertise in the TBRS model*, while the latter feature, sight-reading skill, was rather relevant for the manuscript *The association of eye movements and sight-reading accuracy*.

2.1.3 Assessment of expertise

Musical expertise has typically been assessed on the basis of self-reported amount of formal instruction, membership to certain groups such as professional musicians, or performance evaluations (Mishra, 2019). However, the operationalization of musical expertise often varies considerably across studies which limits their comparability (Mishra, 2019; Puurtinen, 2018). Thus, I chose a different approach in the present work by using a standardized questionnaire, the *Goldsmith Musical Sophistication Index* (Gold-MSI, Schaal et al., 2014). This questionnaire is based on the concept of *musical sophistication*, which refers not only to instrumental performance but to a broader range of musical activities. The questionnaire consists of 31 statements for which participants indicate their agreement on a 7-point Likert scale, and seven items with seven answering options that ask for the frequency and intensity of different musical activities. From these 38 items, five sub-dimensions of musical sophistication can be calculated, namely *active contact with music*, *musical perceptual abilities*, *musical training*, *vocal skills*, and *emotions* (author's translations of the German labels). The global dimension termed *general musical sophistication* is calculated from 18 items across the different sub-dimensions. In the context of the present work, I considered a high relative level of musical sophistication to indicate musical expertise.

Accordingly, whenever possible, musical expertise in the present work was not treated as a dichotomous feature but as a continuum. If two participants differed in their Gold-MSI musical sophistication score, the one with the larger score was considered to be on a higher expertise level. This is in line with Mishra (2019) who claimed that "musical performance expertise is generally considered to develop along a continuum" (Mishra, 2019, p. 575). In addition, using a continuous variable has statistical advantages over a dichotomous measures such as group membership (Royston et al., 2006).

The English version of the Gold-MSI has been found to be positively associated with performance in two listening tasks (Müllensiefen et al., 2014) and with performance of rehearsed music, playing by ear, playing from memory, and improvising (Zhang et al., 2020). While further evidence for the validity of the Gold-MSI is needed, its usage as an indicator of musical expertise has several advantages. Its questions do not solely refer to musical performance. Thus, it can be used with musicians outside the performance-oriented tradition of classical western art music. Moreover, it possesses good psychometric qualities (Müllensiefen et al., 2014) and allows a direct comparison across studies.

2.2 Expert memory

2.2.1 The emergence of the concept of chunking

The classical work by Chase and Simon (1972) on chess expertise was considerably influenced by the work of Miller (1956), who reviewed findings on the limited capacity of humans to process information. In his talk, Miller (1956) referred to two cognitive tasks: absolute judgment tasks and immediate memory tasks. In absolute judgment tasks, a newly presented stimulus has to be compared to a set of stimuli represented in memory (Wever & Zener, 1928). Typically, the stimuli are of physical nature, like colors, tones, or tastes, and the dimensions on which they should be compared are physical properties, like brightness, loudness, or saltiness. An example of a typical experimental procedure would be to present five different pitches and then present one probe pitch. Participants have to identify this probe pitch by comparing it to the set of pitches represented in memory. In immediate memory tasks, a set of sequentially presented stimuli has to be reproduced immediately after the presentation. The employed stimuli typically are numbers, letters, or words, which means they are rather abstract than physical. The most common experimental procedure would be to present ten words at a rate of one word per second and ask the participants to recall the words directly afterwards.

The main statement of Miller was that both the number of stimuli that we can hold in memory and compare new stimuli to, and the number of stimuli than we can hold in memory and recall immediately afterwards is limited. But he suggested that certain stimuli can be organized or grouped into familiar units and that thereby, in immediate memory tasks, the memory limit refers to the *number of items* and not to the amount of included information. He introduced the idea of a *chunk* of information consisting of multiple bits of information and states that "the number of bits of information is constant for absolute judgment and the number of chunks of information is constant for immediate memory" (Miller, 1956, p.349).

To explain how chess experts expanded their memory capacity for chess boards, Chase and Simon (1972) picked up on Miller's concept of a chunk. They assumed a cognitive chunking process that involves sorting features of visual stimuli through a discrimination network and thereby recognizing known patterns. They assumed that this allows to represent a stimulus by using the information in semantic memory associated with it (Chase & Simon, 1972; de Groot & Gobet, 1996; Gobet et al., 2001). The concept of a chunking mechanism in the human mind has become highly influential in cognitive psychology. However, despite its popularity, "our understanding of chunking remains incomplete" (Gilchrist, 2015, p. 1) and "most attempts to define chunks are somewhat vague, ad hoc, or severely limited in scope" (Mathy & Feldman, 2012, p. 347). Accordingly, I will formulate clear definitions of the terms *chunk* and *chunking* in the present work. As chunking takes place in short-term and working memory, I will now introduce these concepts.

2.2.2 Working memory and its measurement

While short-term memory can be conceived as a storage unit that allows the temporary maintenance of a limited amount of information (Cowan, 2008), working memory is generally defined as a set of processes that maintain information during the concurrent performance of other cognitive tasks (Shah & Miyake, 1999). Various explanations have been developed how working memory maintains information during concurrent processing. In the *Time-Based Resource Sharing* model of working memory (TBRS; Barrouillet et al., 2004), this is achieved by "a rapid and frequent switching between processing and maintenance that occurs during the completion of the task" (Barrouillet & Camos, 2007, p. 61). The model is based on four main proposals, namely (1) that limited attentional resources have to be shared between processing and maintenance, (2) that any information that is not in the focus of attention suffers from time-based decay but might be refreshed by retrieving it from memory, (3) that attention can be directed to only one central process at a time, and (4) that consequently, the sharing of attention between processing and maintenance needs to be time-based.

According to Barrouillet and Camos (2007), the Time-Based Resource Sharing model is most suitable to explain performance in complex span tasks. In contrast to simple span tasks, in which a list of memoranda is presented one at a time directly followed by serial recall, complex span tasks involve the maintenance of memoranda during the performance of a secondary task (also called distractor task) (Macnamara et al., 2011). The most classical example of a complex span task is the reading span task developed by Daneman and Carpenter (1980). In this task, participants have to read a set of sentences and are asked to remember the final word of each. Upon encountering a recall prompt, participants are asked to recall all remembered words. With each trial, the number of sentences presented before the recall prompt increases, typically from two to six. A participant's reading span is quantified as the largest number of words reliably recalled in this task.

As was initially supposed by Daneman and Carpenter (1980), complex span tasks have been repeatedly found to be predictive for complex cognitive activities such as reading comprehension, language learning, or mathematical abilities (Ackerman et al., 2005; Daneman & Merikle, 1996). Following the classical reading span task, several types of complex span tasks have been developed, such as the operation span (Turner & Engle, 1989), the counting span (Case et al., 1982), and the picture span task (Tanabe & Osaka, 2009). Although complex span tasks mainly measure domain general executive abilities, performance in these tasks is also influenced by domain-specific abilities, such as chunking (Conway et al., 2005).

The time-based sharing of attentional resources within complex span tasks as described by the TBRS theory has been implemented in a computational model called TBRS* (Oberauer & Lewandowsky, 2011). Then, Portrat et al. (2016) added a chunking mechanism to this computational model, creating TBRS*C. In TBRS*C chunking comprises the searching for sequences of stimuli in long-term memory and, if this search is successful, the representation of these sequences by single units of information. This makes refreshing more efficient as all pieces of information that are part of the single unit can be refreshed at once. Consequently, recall becomes more accurate. In line with this notion, a chunk in TBRS*C denotes a known sequence of information in LTM. Portrat et al. (2016) claimed that chunking involves several sub-processes like the reactivation of previous memoranda, comparison with LTM, and recoding. These processes take place in the focus of attention and therefore, prevent other cognitive processes to occur simultaneously. Moreover, chunking requires that the focus of attention simultaneously holds multiple items. Otherwise, it would not be possible to compare a whole sequence of previously encoded stimuli with the information in LTM.

Due to their superior declarative knowledge on music theory, I generally assumed that participants on a higher expertise level would be more likely to know certain musically meaningful note sequences. For example, they might know the sequence of tones C-E-G as it denotes a C major triad, or they might recognize the sequence of tones D-G-A-D as denoting the chord roots in a common cadential progression I–IV–V–I[2] in the key of D. Sequences of this latter type constitute a common elaboration of the authentic cadence V–I. For the sake of simplicity, I will in the following use the term *authentic cadence* for sequences of notes with this structure. Due to their knowledge of these sequences, I assumed participants on a higher expertise level to be more likely to recognize them as chunks in LTM, recode them as a single unit and, because of the more efficient refreshing, to be more accurate in recalling them.

I tested this assumption in a both a simple span and a complex span task. In the simple span task, sequences of single quarter note symbols were presented for immediate serial recall; in the complex span task, single quarter note symbols had to be memorized while short, notated melodies had to be sight-read. The tonal structure of sequences of to-be-remembered notes was varied within-participants such that more musically meaningful sequences were contrasted with rather arbitrary sequences. In both the complex span task and one of the simple span tasks, major triads were compared to arbitrary trichords. In the other simple span task, authentic cadences were compared to arbitrary cadences. Table 2.1 provides an overview of these paradigms, designs, and hypotheses.

2.2.3 Limits in working memory capacity

In his initial work, Miller (1956) suggested that immediate memory[3] is limited to 7±2 chunks of information. As this limit consistently and repeatedly appeared in the data of various experiments, Miller termed it a “magical number”. Cowan (2001) summarized the huge body of research on mental storage capacity that followed Miller’s work. This research provided comprehensive empirical evidence that the capacity of immediate memory rather amounts to 4±1 chunks.

Mathy and Feldman (2012) provided an explanation why Miller might have repeatedly found the limit of seven elements. They suggested that meaningful sequences of information appear by chance in random stimuli. They showed that the typical rate of occurrence of these meaningful sequences enabled the increase in memory capacity from four to seven elements by means of chunking. If participants are presented, for example, with a string of twelve random letters, it might contain two two-letter acronyms by chance. This might support recall in such a way that these two acronyms and three single letters are recalled successfully. This notion is supported by Sala and

[2] Roman numerals in this context refer to scale steps in diatonic scales
[3] Immediate memory in this context is an umbrella term that includes both short-term and working memory

Gobet (2017), who found superior memory of experts even when stimuli from their domain were arranged randomly.

2.2.4 Theories of expert memory

In the studies that followed Chase and Simon (1972)`s work, several findings emerged that were inconclusive with their chunking theory. For example, it was found that delay intervals and interfering tasks did not decrease the recall accuracy of shortly presented chess boards (Charness, 1976), that supplementary information on the presented chess positions, which was expected to affect the level of processing, increased recall accuracy (Frey & Adesman, 1976), and that chess players were able to play blindfold simultaneous games (Saariluoma, 1989). Based on these findings, Gobet and Simon (1996) concluded that "skilled players sometimes retain more chunks than could be held simultaneously in short-term memory" and that "the Chase-Simon model may need modification" (Gobet & Simon, 1996, p. 10).

They developed the *template theory* by introducing a distinction between different types of chunks stored in LTM: clusters of pieces, position templates, and retrieval structures. Clusters of pieces are congruent with the concept of a chunk proposed by Chase and Simon (1972). They denote known patterns of pieces that frequently occur in chess games. Templates have a comprehensive core and contain slots that can be filled with variable information. For example, a situation from a chess game with a familiar opening might denote a template. Slots would account for the fact that each game with such an opening progresses individually. Lastly, retrieval structures are identifiers for individual positions. For example, the name of a world champion who played a certain position might be used as an identifier to access this position in LTM. By these theoretical developments, template theory provided a more coherent connection of low-level to high-level knowledge (Gobet, 1998). Although template theory was originally developed especially for the domain of chess, it has been the basis for the CHREST cognitive architecture (Gobet et al., 2001) which was also applied to various other domains.

Another line of theoretical developments on expert memory besides template theory was *skilled memory theory* (Chase & Ericsson, 1982) which was later advanced into *long-term working memory theory* (Ericsson & Kintsch, 1995). Skilled memory theory is based on three principles, namely (1) the meaningful encoding principle, (2) the retrieval structure principle, and (3) the speed-up principle (Ericsson & Staszewski, 1989). The meaningful encoding principle refers to the concept of chunking. It assumes that experts have a well-organized body of concepts and relations stored in LTM, and that they use them to encode information. Thereby, experts "form more elaborate and accessible memory representations than novices" (Ericsson & Staszewski, 1989, p. 239). The retrieval structure principle assumes that experts develop cognitive structures that facilitate retrieval from LTM. In these retrieval structures, retrieval cues are used to encode information in LTM in an organized and elaborate way, enabling efficient access. The speed up principle refers to the speed of LTM encoding and retrieval operations. The principle states that, with expertise, the speed of these processes approaches the speed and accuracy of short-term memory processes.

Although the concept of skilled memory has been repeatedly used to describe exceptional memory (e.g. Anderson, 1990; Baddeley, 1990), critics have doubted its generalizability to working memory. The concept of skilled memory implies that the information encoded in LTM would be distorted by interference when other information is processed in a concurrent task (Ericsson & Kintsch, 1995). Hence, the theory needed to be adapted to account for expert working memory performance.

This reasoning has led to *long-term working memory* theory, which is based on the principles of elaborative encoding and recency (Ericsson & Kintsch, 1995). The former is analogous to the retrieval structure principle of skilled memory theory. Experts build complex, hierarchical retrieval structures, which denote organized and stable systems of retrieval cues. On a low level, the encoded information is associated with a retrieval cue. On a higher level, additional information can be used to build supergroups of multiple retrieval cues. The second principle, recency, denotes the usage of temporal separation as a retrieval cue in its own right. New information that is associated with a retrieval cue is extended by the time point and sequence of its encoding. This allows to distinguish it from other information previously encoded with the same cue and thereby protects the information from interference. In summary, skilled usage of working memory in LT-WM theory is explained by including "cue-based retrieval without additional encodings and cue-based retrieval with an elaborated structure associating items from a given trial or context" (Ericsson & Kintsch, 1995, p. 220).

The aspect of template theory and LT-WM theory that is relevant for the present work is that they both assume that experts' superior memory is not only based on chunking, but also on a more efficient, i.e., faster access to information in LTM. This assumption was the basis for the manuscript *Expertise in the TBRS model*. The TBRS*C computational model is especially suited to model the time course of cognitive processes during complex span performance. Accordingly, I used this model to separately simulate the recall performance of higher-expertise and lower-expertise musicians (this part of the work required a dichotomization into groups). The resulting differences in the model parameters enabled me to explore the expertise differences in chunking and in the time-course of cognitive processing. To clarify this analytical approach, I will now describe the functioning of the TBRS*C simulations as well as the parameters of interest.

2.2.5 Exploring expertise differences in cognitive processing with TBRS*C

TBRS*C is a computer program that mimics four basic cognitive processes, namely encoding, refreshing, distractor processing, and recall, to simulate performance in complex span tasks. Due to the central attentional bottleneck (Pashler, 1999), the program performs only one of these processes at a time. The data structure that the program uses to represent information in memory is a fully interconnected two-layer network. One layer holds the memoranda; another layer holds sets of markers that represent item positions. Each item is associated with all sets of position markers and one parameter indicates the strength of the association. If a memorandum is encoded, the strength of its association with the set of markers representing its position is increased. During encoding, the association strength is bound to a certain maximum. The *R* parameter defines the rate of increase of the association strength. Thus, a larger *R* leads to more rapid encoding, as the maximum association strength is reached faster.

Refreshing takes place in any period in the task that is not assigned to other processes. It works in a similar manner than encoding and hence is also affected by the *R* parameter. Cycling through all memoranda beginning with the first, the association between a memorandum its respective set of position markers is strengthened. While during encoding, the association strength is always increased to the maximum, during refreshing, it is only increased for a brief period of fixed duration.

The processing of distractors is not modeled per se in TBRS*C. It is rather implemented as a period in which no other central process (encoding, refreshing, chunking) takes place and thus, in which the strength of the association between memoranda and item positions decreases due to time-based decay. The length of this period is represented by the parameter *Ta*. In the present experiment, the distractor task was the performance of a short, notated melody on a piano. In this task, I assumed that each performed note would catch the attention of the participant for a short amount of time and in between the performed notes, attention would switch back to the refreshing of memoranda. This assumption might seem strange for anyone who is a musician. Attention during instrumental performance is completely captured by the task. It does not seem possible to “do something in parallel”. However, in the Time-Based Resource Sharing model, the rapid and time-based switching of attention is not assumed to be a conscious process. Thus, it is reasonable to assume that it occurred during sight-reading.

Chunking is implemented in TBRS*C as a process of searching for strings of memoranda in a pre-defined LTM. This search is performed right after the encoding of a memorandum. It is defined by two parameters, namely the chunk search duration (*cSD*) and the probability of chunk retrieval (*PCR*). The former defines how much time is invested in the search for chunks; the latter defines how likely it is that the simulation recognizes a stimulus as a chunk in LTM. Once a chunk is successfully recognized, the single memoranda are chained, and the chunk is associated with the set of markers representing the position of the first memorandum of the chunk. This makes subsequent refreshing more efficient, as all three notes can be refreshed at once. In the present study, the pre-defined LTM contained the major triads used in the experiment. Starting at the third note, after the encoding of each new note, the program searched for the last three notes in LTM. If the program would for example recognize the note sequence C-E-G as a known triad, it would associate the group of three notes with the markers representing the position of the note C.

To provide insights into experts' chunking and working memory processes, I explored expertise differences in the parameters *R*, *cSD*, *PCR*, and *Ta.* Theories of expert memory assume that experts can access information in LTM rapidly. Thus, I assumed that encoding and searching for chunks in LTM would be faster for experts. In the simulations, this would result in a larger *R* and a smaller *cSD* parameter. Moreover, due to their superior knowledge, experts might be more likely to recognize chunks in LTM, i.e., might have a larger *PCR* parameter. Lastly, I was interested if these changes in recall processes would lead to changes in distractor processing. Therefore, I explored expertise differences in the time used for the attentional capture of distractors (*Ta*).

Table 2.2 depicts an overview of the theoretical foundation for the analyzed TBRS*C parameters. In the analysis, the simulation was performed with various values for the parameters *R*, *PCR*, *cSD* and *Ta*. Each simulation provided a serial recall accuracy. Using model fit statistics and Chi-square tests, these simulated recall accuracies were compared with the empirical recall accuracies of the participants. Thereby, it was possible to investigate which parameter combination provided the best fit to the data. By using a computational model to study expertise and by exploring fundamental cognitive processes with this model, I addressed two demands by Gobet (2019) to "develop more rigorous theories, implemented as computer programs" and to "go back to fundamental questions, including setting the value of the parameters of cognition" (Gobet, 2019, pp. 48–49).

2.3 Sight-reading

2.3.1 Theoretical models of sight-reading skill

One of the first systematic attempts to describe the cognitive processes involved in sight-reading was made by Wolf (1976). He conducted two sets of semi-structured interviews, one with pianists who were excellent sight-readers and one with pianists who were professionals but lacked excellence in sight-reading. From the first set of interviews, he concluded that (1) recognizing familiar musical configurations is central for sight-reading, (2) some parts of the melody are guessed or predicted based on what is expected, (3) sight-reading is analogous to reading text with single notes being the analogue of letters, and (4) sight-reading skill is rather independent from other instrumental skills.

Wolf considered the first of these statements to be most crucial: “The musician does not sight-read note-by-note but instead looks for familiar musical patterns” (Wolf, 1976, p. 146). Referring to Miller (1956) and to the study on chess memory by Simon and Barenfeld (1969), Wolf introduced the concept of a chunk to sight-reading research. He claimed that sight-readers “will recognize familiar constellations of notes and process them as single units, or chunks, of information” (Wolf, 1976, p. 156).

Wolf then combined this idea with the multi-component model by Atkinson and Shiffrin (1968) and the link between memory and muscles proposed by Broadbent (1958) to develop a cognitive model of musical sight-reading. In the model, external input is stored in a multi-component sensory store, is compared with LTM, and then passed through a filter into short-term memory. In short-term memory, there are seven slots that each can hold one chunk of information. Information from these slots is passed through an effector system to the muscles. Wolf (1976) used the second set of interviews to “validate” this cognitive model.

Table 2.2

*Overview of TBRS*C parameters that were investigated for expertise differences*

Theoretical foundation	Assumption	TBRS*C parameter	Parameter meaning
Chunks (chunking theory); clusters of pieces (template theory); meaningful encoding principle (skilled memory theory)	Experts have superior knowledge on meaningful patterns in their domain and use this knowledge to compress information during encoding	Probability of chunk retrieval (*PCR*)	Likelihood of recognizing that a stimulus is a chunk in LTM
Retrieval structures (template theory); speed up principle (skilled memory theory); elaborative encoding (LT-WM)	Experts can access information in LTM rapidly	Encoding strength (*R*)	Rate of increase of the association strength during encoding
		Chunk search duration (*cSD*)	Time invested in the search for chunks
No assumption – exploratory analysis		Distractor processing (*Ta*)	Time used to process a distractor

The next step in the development of a theoretical understanding of sight-reading was made by Waters et al. (1998). They measured the accuracy of sight-reading performance of thirty pianists and designed several experimental tasks that they though to reflect the skill to recognize musical pattern, to predict the progression of music, and to generate auditory representations from printed notes. Using ANOVAs, correlations, and hierarchical regressions, Waters and colleagues analyzed how the performance in these experimental tasks was associated with sight-reading performance.

They concluded that "auditory skills and prediction skills are important factors underlying skilled sight reading over and above basic pattern-recognition skills" (Waters et al., 1998, p. 147).

More recently, Kopiez and Lee developed two models of sight-reading skill, namely the dynamic model of skills involved in sight-reading (Kopiez & Lee, 2006) and the general model of skills involved in sight-reading (Kopiez & Lee, 2008). 52 piano students and graduates performed five musical pieces of increasing difficulty at first sight. In addition, data on general cognitive skills (such as working memory), elementary cognitive skills (such as processing speed), and expertise-related skills (such as sight-reading experience) was collected. The same data set was used for both models. The dynamic model described how the significance of specific predictors for sight-reading skill changed with the difficulty of the sight-reading task. In easy tasks, the predictor piano expertise was sufficient to explain sight-reading performance. In tasks with intermediate difficulty, the predictors psychomotor speed, processing speed, inner hearing, and sight-reading experience explained significant amounts of variance in sight-reading accuracy. At the highest level of difficulty, psychomotor speed became the strongest predictor. The authors concluded that sight-reading skill is determined by a combination of innate and practice-related skills.

To complete this dynamic model, Kopiez and Lee developed the general model of skills involved in sight-reading (Kopiez & Lee, 2008). For this model, the association between the different component skills and sight-reading accuracy was analyzed irrespective of task difficulty. Using correlations, hierarchical regression, and factor analysis, the authors found that the combination of psychomotor speed, sight-reading experience up to the age of fifteen, processing speed, and inner hearing was optimal and explained about 60 % of the variance in sight-reading accuracy.

2.3.2 Eye movements as a distinct skill

While all these theoretical considerations might provide an idea of the skills that contribute to accurate sight-reading performance, they neglect one issue: the eye movements. The eyes play a central role during sight-reading as they provide the information that is required for an accurate performance. To this end, the oculomotor system needs to know which information is relevant and where this information can be found, i.e., it needs to have its own knowledge base (Land & Furneaux, 1997). This knowledge base might take the form of a prototypical sequence of actions, a so-called *action schema*. This logic suggests that eye movements during sight-reading can be considered a skill in their own right.

Fadde and Jalaeian (2019) defined a skill as the *practice-related* superiority of goal-directed action. Accordingly, if eye movements denote a skill, they should vary with expertise, as expertise results from practice. And indeed, various studies found expertise differences in eye movements during sight-reading (for a summary see Sheridan et al., 2020). Goolsby (1994) analyzed how eye movements during the singing of an unknown song changed with expertise of the reader, notational complexity, and rehearsal of that song. Two groups of differing expertise were confronted with four melodies of different notational complexity that they had to sing three times. While the first two times that they performed the melodies were without prior rehearsal, there was a practice period before the third trial. Participants were asked to perform in a fixed and steady tempo, but without metronome. Goolsby (1994) found that the duration of fixations differed between musicians of varying expertise levels while the number of fixations did not.

Gilman and Underwood (2003) used a moving window paradigm in their study, in which only a pre-defined area around the point of fixation was visible. Pianists had to perform three tasks of varying cognitive demand: an error detection task in which music had to be read silently and parts that did not fit into the given key had to be identified (low cognitive load), a sight-reading task in which a song had to be played at first sight (medium cognitive load), and a transposition task in which a song had to be transposed during its performance (high cognitive load). The performance tempo was freely chosen by participants in all tasks. With the help of a pretest, the music reading expertise was deduced a priori. The size of the moving window was varied across repeated measurements (one-beat, two-beat, four-beat window, no window). The authors found that more experienced participants read with fewer fixations of similar duration.

Penttinen et al. (2015) measured the eye movements of education majors minoring in music education and of music performance majors during the performance of a children's song. The performance was temporally controlled, meaning that participants played the song on a piano in time with a metronome. The participants performed the song four times. In two of the trials the song contained unexpected melodic alterations. The performance majors read with shorter average fixations than the education majors, i.e., expertise was negatively associated with fixation duration.

Arthur et al. (2016) asked participants to sight-read individually composed four-bar melodies and their visually disrupted counterparts. These disrupted counterparts consisted of the same melodies than the originals, but bar-lines were removed, and stem directions and inter-note distances were varied. In the employed experiment, four seconds after a signal sound, the melody appeared. Participants were asked to start immediately with the performance and to play as quickly and accurately as possible without stopping. Results showed that more experienced participants performed faster but with equal number and duration of fixations than less experienced participants.

It becomes clear that, across these four studies, the association between expertise and eye movements was inconsistent. This might be due to the methodological inconsistencies of these studies. Puurtinen (2018) claimed that studies investigating eye movements during sight-reading showed considerable variation concerning the selection of musical stimuli, the control of performance tempo, the measurement of participants' musical expertise, and the methods of statistical analysis. This methodological variation is especially problematic as "the effect of expertise on fixation duration seems to be largely task- and stimulus-dependent" (Holmqvist & Andersson, 2018, p. 531, see also Bertram et al., 2013). However, the fact that eye movements during sight-reading have repeatedly been found to differ with expertise supports the idea that eye movements denote a distinct skill.

2.3.3 Eye tracking technology

Eye movements during sight-reading are investigated by means of eye tracking. While there are many different types of eye trackers (reviewed in Holmqvist & Andersson, 2018), the one used in the present study (a Tobii TX 300) was a video-based eye tracker. These machines consist of an infrared light transmitter and a camera, often connected with a screen that displays the stimuli. The transmitter produces patterns of reflections on the eyes of the participant with infrared light, commonly of 870 nm wave length (Holmqvist & Andersson, 2018). The camera records the eyes and complex algorithms use the infrared reflections to deduce the 3D position of each eyeball.

From the raw data produced by these eye trackers, oculomotor events, such as fixations and saccades, need to be derived, again by using algorithms. To calculate oculomotor events in the present study, a velocity-based algorithm was used. In this algorithm a velocity threshold was defined and all movements of the eyes with a velocity above this threshold were defined as saccades. Events between two consecutive saccades were defined as a single fixation. Both the raw data as well as the derived oculomotor events provide a vast number of eye movement measures. Holmqvist and Andersson (2018) distinguished four classes of measures, namely movement measures (e.g., direction, amplitude, duration, velocity, shape), position measures (e.g., dispersion, similarity, duration), count measures (e.g., saccade rate, number of fixations, blink rate), and distance measures (e.g., saccade distance). The psychological meaning of these measures is largely dependent on the paradigm and the stimulus material. In the present work, due to their widespread use in sight-reading research, the number and duration of fixations were the central eye movement measures.

2.3.4 A missing piece in the puzzle

In addition to expertise, various aspects of the eye movements during sight-reading have been researched (for a review see Puurtinen, 2018). One aspect that remains largely unclear, however, is how eye movements are related with performance accuracy. Chitalkina et al. (2021) found that performance errors were associated with a decrease in pupil size. In the study by Lim et al. (2019), complexity of the melody moderated the association between performance accuracy and the eye-hand span. The eye-hand span denotes the distance between the point of fixation and the point of performance. In simple melodies, performance accuracy and the size of the eye-hand span were positively associated, while in complex melodies, they were negatively associated. Finally, Drai-Zerbib et al. (2012) discovered that for non-experts, the overall gaze duration during sight-reading was positively associated with the number of errors. While these studies provided first evidence that eye movements and sight-reading accuracy might indeed be associated, this association was not their major focus; the findings reported here were merely secondary findings.

It is surprising, however, that the association between eye movements and sight-reading accuracy was largely neglected by previous research. If eye movements and sight-reading accuracy are related, eye movements could potentially mediate the effect of other predictors. For example, sight-reading experience, which was established as a predictor of sight-reading accuracy by Kopiez and Lee (2008), might actually influence eye movements which, in turn, affect sight-reading accuracy. Thus, I argue that knowledge on the association between eye movements and sight-reading accuracy is crucial for a comprehensive understanding of the cognitive processing of musical notation during sight-reading. The goal of the present work was to explore this association.

During the complex span task, participants performed a short, unfamiliar, notated melody on an electric piano. Performance tempo was controlled at 70 beats per minute (bpm) by using a metronome. During this performance, participants' eye movements and the MIDI data of their performance were tracked. I developed an algorithm that assessed the accuracy of the musical performance from the MIDI data (see Appendix). Using these measures, I investigated if the number and duration of fixations were associated with the accuracy of pitch and note onset.

2.4 Overview of the present work

In the previous paragraphs, I introduced chunking as a general cognitive mechanism that needs to be conceptualized within the context of a given domain. In the processing of musical notation, tonal features were found to be highly relevant for chunking (Deutsch, 1980; Halpern & Bower, 1982; Kalakoski, 2007). Thus, the manuscript *Chunking in Tonal Contexts* investigated the effect of specific tonal features on chunking processes. The manuscript presents two simple span experiments that were administered online. Participants were presented with sequences of musical note symbols for immediate serial recall. The note sequences were varied in a 2 x 2 within-participants design with the factors *tonal structure* (chunking-supportive vs. chunking-obstructive) and *list length* (short vs. long). Chunking-supportive melodic sequences resided in a clearly recognizable tonal context and contained meaningful melodic cells with clear labels. In chunking-obstructive melodic sequences, the tonal context was unclear and melodic cells were rather arbitrary. Different types of melodic cells were used in the two experiments. The first experiment involved major triads and arbitrary trichords; the second experiment involved authentic cadences and arbitrary cadences. Arbitrary melodic cells involved intervals of eight and nine semitones to the root note, as these intervals are not part of diatonic major scales.

Based on the results of these simple span experiments, I developed a musical complex span task. In this task, participants were presented with sequences of single quarter note symbols for serial recall. In between the presentation of these to-be-remembered notes, participants had to perform an unfamiliar, notated melody on an electric piano. In other words, participants were presented with a single quarter note which they had to memorize, then they had to sight-read a short melody, then they saw another note they had to memorize, sight-read another melody, and so on, until recall was prompted. As the evidence for chunking processes was clearer in the simple span experiment involving triads than in the one involving cadences, I used this manipulation of tonal structure in the complex span task. That is, in the present complex span task, it was varied within-participants if to-be-remembered notes formed major triads or arbitrary trichords. Moreover, as I found a ceiling effect for sequences consisting of up to nine notes in the simple span experiment, I increased the list length to twelve notes in the complex span task.

The manuscript *Expertise in the TBRS model* used the recall data of this musical complex span task. By comparing the recall of major triads with the recall of arbitrary trichords in another paradigm, this experiment extended the findings from the simple span task. However, music theory and tonal considerations did not play a central role in this manuscript. Rather, music was considered to be an exemplary domain in which domain-general cognitive phenomena can be investigated. Chunking was not only investigated by analyzing the association of recall accuracy with stimulus structure and participant expertise, but also by means of computational modeling using TBRS*C (Portrat et al., 2016). Based on theories of expert memory, such as skilled memory theory (Chase & Ericsson, 1982) and template theory (Gobet & Simon, 1996) it was assumed that expert memory would be characterized by chunking and a faster access to information in LTM. Hence, expertise differences were investigated in those TBRS*C parameters that represent the speed of encoding (*R*), the chunk search duration (*cSD*), and the probability of chunk retrieval (*PCR*). In addition, expertise differences in the parameter representing the time used for distractor processing (*Ta*) were explored. By considering rapid LTM access in addition to chunking, this manuscript provided a broad perspective on expert memory.

The manuscript *The association of eye movements and sight-reading accuracy* used the eye movement and musical performance data of the musical complex span task. I analyzed if the number and duration of fixations were associated with the accuracy of pitch and note onset. In doing so, I controlled for three other variables that can be assumed to affect both eye movements as well as performance accuracy, namely musical expertise (Arthur et al., 2016; Cara, 2018; Penttinen et al., 2015), practice (Burman & Booth, 2009; Goolsby, 1994; Rosemann et al., 2015), and features of notes (Ahken et al., 2012; Lim et al., 2019). Thereby, the present work provided the first systematic and comprehensive insights into the association of eye movements and performance accuracy during sight-reading. In the following, you find the three manuscripts of the present work exactly as they were published or submitted. Thereafter follows a general discussion.

3. Chunking in Tonal Contexts

This article can be found under the following link:

https://doi.org/10.1177/03057356211013396

4. Expertise in the Time-Based Resource Sharing Model

This article can be found as a revised preprint under the following link:

https://doi.org/10.21203/rs.3.rs-2533819/v1

5. The association of eye movements and sight-reading accuracy

This article can be found under the following link:

https://doi.org/10.16910/jemr.14.4.5

6. General discussion

The present work investigated three aspects of expertise in the processing of musical notation. First, it provided insights how specific tonal features enable chunking (Chase & Simon, 1972; Gobet et al., 2016; Mathy & Feldman, 2012; Miller, 1956) in short-term memory for musical notation. Second, it explored chunking and rapid LTM access in working memory for musical notation with reference to the Time-Based Resource Sharing model (Barrouillet & Camos, 2007). Third, it analyzed the association between number and duration of fixation with the accuracy of pitch and note onset during sight-reading of notated melodies. Table 6.1 provides an overview of the main findings of the present work, the interpretation of these findings, and unresolved questions.

The manuscript *Chunking in tonal contexts* reports two experiments which tested the interacting impact of tonal structure and musical expertise on the recall accuracy in a simple span task with musical note symbols. I created two types of tonal structures, namely *chunking-supportive* and *chunking-obstructive* tonal structures. Chunking-supportive tonal structures were more meaningful in the context of traditional tonal music theory. They resided in a clearly recognizable tonal context and consisted of melodic cells with clear labels. In the first of the reported experiments, melodic cells were three-note arpeggiated major triads. In the second experiment, they were four-note sequences corresponding to a common tonal progression of chord roots representing the scale degrees I–IV–V–I. These sequences were labeled *authentic cadences*, for short. In addition, in both experiments, list length was varied as a within-participants factor with two levels (short vs. long). Based on chunking theory (Kalakoski, 2007; Miller, 1956), I hypothesized that there would be two main effects and a positive interaction. I expected notes to be recalled more accurately by more experienced participants and when the sequences have a meaningful tonal structure; I expected that the advantage for meaningful tonal structures would be larger for more experienced participants.

I found that recall was more accurate in stimuli with a meaningful tonal structure. The expected interaction was found in short sequences containing authentic cadences. Transitional errors implied that notes were grouped into melodic cells when stimuli had a meaningful tonal structure.

Table 6.1
Overview of the findings of the present work.

Manuscript	Main finding	Interpretation	Unresolved questions
Chunking in tonal contexts	More accurate recall for more experienced participants and in chunking-supportive sequences. Variation of the interactional pattern (expertise by tonal structure) across experiments and conditions.	Chunking in short-term memory for musical note symbols is supported by the combination of a clear tonal context and of melodic cells with clear labels.	Was the unexpected interactional pattern due to a ceiling effect? What are the individual effects of a clear tonal context and of melodic cells with clear labels?
Expertise in the TBRS model	Larger *R*, smaller *cSD*, larger *PCR*, and larger *Ta* for higher-expertise hobby musicians than for lower-expertise hobby musicians. Marginal expertise differences in the music student sub-sample.	In the context of the TBRS model, expert memory can be conceptualized by rapid LTM access, which increases opportunities for refreshing, and by reliable chunk recognition, which increases the efficiency of refreshing.	Were the marginal expertise differences in the music student sub-sample due to a ceiling effect? Are valid estimates of *Ta* provided by fitting simulated to empirical data?
The association of eye movements and sight-reading accuracy	Negative association of number of fixations and accuracy of note onset while controlling for expertise, practice, and type of note pair.	Sight-reading with few fixations is advantageous as less cognitive resources are needed for the planning of saccades the integration of information.	Do eye movements affect sight-reading performance or vice-versa? How is the association between other eye movement measures and sight-reading accuracy?

However, there were also unexpected effects. There was no interaction in the recall of major triads and there was a reversed pattern of interaction in the recall of short cadential sequences. These unexpected data pattern potentially resulted from a ceiling effect, as many participants performed the task with perfect or near perfect accuracy. As meaningful tonal structures in these studies were created by combining clearly recognizable tonal contexts with the presence of melodic cells with clear labels, the individual effects of these two factors remain unclear.

The manuscript *Expertise in the TBRS model* aimed to provide a first account how expert memory might be conceptualized in the context of the Time-Based Resource Sharing model (Barrouillet & Camos, 2007). I conducted an experiment involving a complex span task with musical notation. Participants were shown sequences of twelve single quarter note symbols and had to recall the pitch of these notes at the correct serial position. In between the presentation of each of these to-be-remembered notes, participants had to perform an unfamiliar, notated melody on an electric piano. This task was completed by a hobby musician and a music student sub-sample. Both sub-samples were split in a higher-expertise and a lower-expertise group by performing a median split on the general musical sophistication score of the Gold-MSI questionnaire (Schaal et al., 2014). The serial recall accuracy of each group was then simulated with the TBRS*C computational model (Portrat et al., 2016). I searched for differences between the expertise groups on parameters that represent the speed of accessing LTM and the reliability of chunking processes. The former comprised the speed of encoding to-be-remembered notes (*R*) and of searching for chunks in LTM (*cSD*). The latter refers to the probability of chunk retrieval (*PCR*). In addition, I explored if changes in these parameters would be associated with changes in the sharing of attentional resources between the task components. To this end, I analyzed expertise differences in the time used for the processing of distractor notes (*Ta*).

In the simulations, higher-expertise hobby musicians encoded memoranda faster, invested less time in chunk search, found chunks with a larger probability, and used more time to process distractor notes than lower-expertise hobby musicians. In the music student sub-sample, the only parameter difference concerned the time invested in chunk search (*Ta*). Higher-expertise music students in the simulations invested more time in chunk search than lower-expertise music students. However, auxiliary regression analysis revealed that there were in fact no expertise differences in recall accuracy in the music student sub-sample. Thus, I deem the difference in *Ta* between higher-expertise and lower-expertise music students to reflect random variation in the data. As other studies used reaction times to estimate *Ta*, it is unclear if it is a valid approach to estimate *Ta* by fitting simulated to empirical data.

The paper *The association of* e*ye movements and sight-reading accuracy* investigated how the number and duration of fixations were associated with the accuracy of pitch and note onset during sight-reading. To this end, I used eye movement and MIDI performance data collected from the distractor task of the musical complex span task. Eye movements on specific note pairs within the melodies were analyzed using AOIs. To ensure that the association between eye movements and performance accuracy was not due to a third variable, I controlled for expertise, practice, and rhythmical features of notes. Expertise was measured by the Gold-MSI global scale score. Practice was operationalized by the number of trials participants had already completed. Rhythmical features of notes were represented by the type of note pair (quarter-quarter, quarter-eighth, eighth-quarter, eighth-eighth) that participants were reading.

Number of fixations during reading of the melody was negatively associated with accuracy of note onset while controlling for the effect of expertise and practice. Number of fixations during reading of the note pairs was negatively associated with accuracy of note pair onset while controlling for the effect of expertise, practice, and type of note pair. I conclude that reading with fewer fixations might be beneficial for sight-reading as eye movement planning and information integration becomes less demanding. This saves cognitive resources which can be used to increase performance accuracy. However, the present experiment does not allow to establish causality. Thus, it remains unclear if eye movements affect sight-reading accuracy or vice-versa. Besides, fixational measures only represent one specific aspect of eye movements. It remains an unresolved question if and how other eye movement measures are associated with sight-reading accuracy.

6.1 A comprehensive notion of expertise in processing musical notation

The three experiments of the present work provide a comprehensive view on expertise in the processing of musical notation. First, when musical notation is processed in the context of a recall task, experts seem to be characterized by superior memory. I found more experienced participants to recall sequences of note symbols more accurately in both simple span and complex span task. Under specific circumstances, more experienced participants' recall was especially accurate in sequences with a meaningful tonal structure. Second, when musical notation is processed in the context of sight-reading, experts seem to be characterized by efficient perception and accurate performance of notes. I found more experienced participants to perform the notated melodies more accurately, and to read the note pairs with fewer fixations and a shorter gaze duration. So, in summary, expertise in the processing of musical notation might comprise chunking and rapid encoding of note symbols in recall tasks, and efficient perception and accurate performance of notated melodies in sight-reading tasks.

The samples in the present studies were highly heterogenous, comprising musicians of different instruments, genre, and backgrounds. It is highly likely that the way of practicing and the musical activities varied considerably among participants. It is also likely that participants were not specifically trained in memorizing and recalling musical notes. Nevertheless, evidence of expert memory could be found in all samples. This supports the notion that expert memory develops as a by-product of deliberate practice in domain-specific activities. Ericsson (2018b) claimed that experts do "not practice to improve their memory directly, but rather the superior working memory is a unintentional consequence of efforts to improve the selection and execution of superior actions" (Ericsson, 2018b, p. 700). In addition, though, the superior working memory of experts can reasonably be assumed to facilitate task-specific activities and thus to support deliberate practice.

Figure 6.1 depicts this logic. Deliberate practice is deployed to cultivate domain skills. Expert memory develops as a by-product of deliberate practice but supports it by enabling superior processing of domain-specific information. For example, a musician might practice performing a notated melody to develop instrumental skills. As a by-product, she might learn that the sequence C-E-G is a common sequence which is called "C major triad". Eventually, she might be able to access this information rapidly and to treat the three notes as a single mental unit when processing them. These changes in information processing then might come into effect when she practices a new melody.

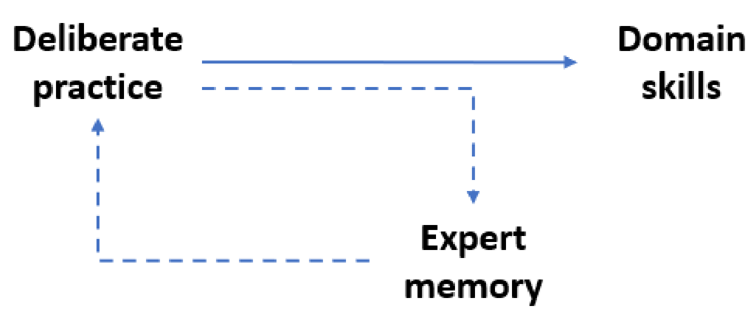

Figure 6.1. The role of expert memory in the development of domain-specific skills.

It might be asked, though, why there is no arrow from domain skills to expert memory in Figure 6.1. For example, sight-reading skill might comprise the ability to fluently perform a certain major triad on an instrument. Just like deliberate practice, this action schema might contribute to the memorization and recall of such a note sequence, i.e., to expert memory. However, I find it more intuitive to assume that skills (expert memory and domain skills) develop as a result of an activity (deliberate practice). To assume that expert memory is enabled by other domain-specific skills would rather suggest that it is a component of these domain-specific skills and not a distinct concept.

It becomes apparent that the relationship of expert memory and domain-specific skills should be investigated to develop a comprehensive idea of the cognition of expertise. “Simply showing that experts display superior performance on memory tasks [...] does not necessarily help us understand how the superior memory performance is related to the superior expert performance” (Ericsson, 2018b, p. 708). Tasks such as the musical complex span task which allow to investigate expert memory and domain skills simultaneously might proof useful in this pursuit.

Concerning the development of expertise, the present study is in in line with the claim by Ericsson (2018a) that experts’ advantages develop gradually. Although the sample was rather heterogenous, I found certain measures like the recall accuracy in the simple span task and the gaze duration on the note pairs to change gradually with the Gold-MSI score. But if expertise develops gradually, is it justified to think about stages of expertise development as proposed by Mishra (2019)? Gobet (2012) claims that it is an open question “whether there are stages in the development of expertise [...]. The presence of power laws seem to suggest that expertise development is continuous, but stages keep appearing in theories of expertise” (Gobet, 2012, pp. 952–953). Concerning this question, the manuscript *Expertise in the TBRS model* provides interesting insights. Higher-expertise hobby musicians had a similar Gold-MSI score (M = 77.84; SD = 5.66) than lower-expertise music students (M = 79.32; SD = 5.52). Nevertheless, there was a marked difference in recall accuracy (higher-expertise hobby musicians: M = 0.39; SD = 0.22; lower-expertise music students: M = 0.77; SD = 0.16) and sight-reading accuracy (higher-expertise hobby musicians: M = 0.62; SD = 0.26; lower-expertise music students: M = 0.94; SD = 0.13). While this is only a descriptive comparison, the data still suggests that there was a difference between musicians on different stages of expertise. In addition, defining and labeling different stages of expertise development might be helpful to think and speak about expertise.

With reference to the measurement of expertise in the domain of music, the Gold-MSI (Schaal et al., 2014) has proven as useful tool. It is freely accessible, easy to administer, standardized, and it was predictive for recall of musical notation, eye movements, and sight-reading accuracy in the present experiments. As the Gold-MSI measures general musical sophistication, this finding is in line with the meta-analysis by Mishra (2014), which found sight-reading to be a skill that improves with the general musicality of a person. However, the present finding that the Gold-MSI was predictive for sight-reading accuracy contradicts the study by Zhang et al. (2020). They found that the Gold-MSI was predictive for performing rehearsed music, playing by ear, playing from memory, and improvising, but *not* for sight-reading. There are a number of methodological differences between the present work and the study by Zhang et al. (2020) that provide a potential explanation for the contradictory findings. Zhang et al. (2020) measured sight-reading skill by means of self-report on one item (which was “I am very bad at ‘sight-reading’ – that is, I am not good at reading a new piece of music and performing it straight away”). The present study, in contrast, used the actually obtained accuracy in a sight-reading task. Furthermore, Zhang and colleagues did not use the global scale of the Gold-MSI but used a score that was the sum of the answers across all items. One drawback of the Gold-MSI score in the present study, however, was that it might not have differentiated well between participants on a high expertise level. This calls for additional diagnostic instruments that allow to measure expertise reliably even on high levels.

6.2 Implications for expert memory

The present work provided various evidence of experts’ superior memory. First, more experienced participants showed more accurate recall from short-term memory. This is in line with findings by Meinz and Salthouse (1998) and Kalakoski (2007) who found that experts recall musical notes more accurately. Second, I found some evidence of chunking in both short-term as well as working memory. In short cadential sequences of the simple span task and in the hobby musician sub-sample of the complex span task, experts recall was especially accurate in more meaningful note sequences. This supports the notion of Croonen (1991) and Deutsch (1980) that tonal structure is a crucial aspect in the recall of musical information. Third, simulations with TBRS*C suggested that experts accessed LTM rapidly. This is in line with template theory (Gobet & Simon, 1996) and skilled memory theory (Chase & Ericsson, 1982) which also assume that experts are characterized by rapid LTM access.

The superior short-term memory of experts in the present study can be explained by the larger knowledge network of experts that supports *redintegration*. Redintegration denotes the reconstruction of partly degraded memory traces (Ritchie et al., 2015). Previous studies have found that long-term knowledge in a task domain supported redintegration (Botvinick, 2005; Botvinick

& Bylsma, 2005; Jones & Farrell, 2018). Experts' knowledge has been described as a cueing structure for the retention of information in memory (Bellezza & Buck, 1988). In the present study, it can be assumed that the information that was activated when a note symbol was encoded was more extensive for more experienced participants. This associated information might have helped experts to restore the required pitch information during recall.

In working memory, however, not redintegration, but refreshing might be the key to experts' superiority. This was implied by the TBRS*C simulations. The rapid access to LTM might increase the time that is available for refreshing; chunking might increase the efficiency of refreshing. Thus, the interpretation of the present results calls for a detailed understanding of the process of refreshing. Camos et al. (2018) defined refreshing as a domain-general and attention-demanding process with the purpose of keeping mental representations active. Refreshing processes "increase the activation of recently presented, encoded, or retrieved information to keep it in an accessible state from moment-to-moment, thereby enabling real-time thinking" (Camos et al., 2018, p. 19). This increase in activation is accomplished by focusing attention on inner representations or thoughts. Unlike articulatory rehearsal that only concerns verbal information, refreshing might target information in all kinds of formats.

In the context of the Time-Based Resource Sharing model, refreshing is assumed to be a rapid process outside the focus of explicit awareness. Consequently, refreshing can be assumed to occur during demanding and complex tasks, such as sight-reading. Moreover, if refreshing happens unconsciously and rapidly, even small free periods during a cognitive task might be utilized for refreshing. Camos et al. (2018) claimed that prior knowledge and expertise might influence the efficiency of refreshing. However, they also stated that "the actual process of refreshing should be independent of LTM", as "refreshing interacts directly only with representations that are active in WM and no critical role assigned to LTM" (Camos et al., 2018, p. 27). The present work suggests a slightly different notion. If certain sequences of items are present in LTM and if these sequences are recognized in a stimulus, they might be refreshed by a single act of cognition. Thus, the role of LTM might indeed be described as critical, as LTM information represents the necessary condition for chunking.

6.2.1 Implications for chunking

One central theoretical notion in the present work was chunking (Miller, 1956). The hypothesis associated with this notion was an interaction between musical expertise and tonal structure in such a way that more experienced participants would have an additional recall advantage in more meaningful note sequences. This hypothesis only received partial support as it only was confirmed in long cadential sequences of the simple span task and for hobby musicians in the complex span task. However, the absence of the expected interaction can be explained with a ceiling effect. Although I increased the number of memoranda from up to nine in the simple span task to twelve in the complex span task, the data suggests that such a ceiling effect occurred in both experiments.

However, there were also various aspects of the present results that supported the idea of a chunking process in the processing of musical notation. There was some evidence for the expected interaction in both the simple and the complex span task. The proportion of transitional errors (see Figure 3.2) showed that note sequences might have been grouped into melodic cells in memory. TBRS*C simulations suggested that hobby musicians' probability of chunk recognition changed with expertise and that chunking was advantageous for recall even if this probability was low. With respect to the specifics of chunking in the processing of musical notation, the present work showed that the presence of melodic cells with clear labels might support chunking both by itself and in combination with a clearly recognizable tonal context.

Concerning the measurement of chunking, the presented studies supported the usefulness of transitional errors as claimed by Gilchrist (2015). This statistical measure allows to visualize groups in memory which is a valuable information in the context of chunking. However, transitional errors can only be calculated for the transition following a correctly recalled item. If an item is recalled incorrectly, the following transition produces a missing. Thus, I claim that the other two types of transitions – from incorrect to correct and from incorrect to incorrect – should also be considered. This would avoid missing data and would provide a comprehensive information on the transitions between memoranda.

One topic that is closely associated with chunking is limits of human memory capacity. When applying the proportion of correctly recalled notes in the complex span task (see Figures 4.3 and 4.7) to the total number of memoranda, it becomes apparent that the present work found evidence for both "magic numbers" 7±2 (Miller, 1956) and 4±1 (Cowan, 2001). Hobby musicians recalled 4±1 notes (i.e., 5.04 notes in the major triads condition and 3.24 notes in the arbitrary trichords condition); music students recalled 4±1 melodic cells in sequences of major triads (i.e., 3.2 melodic cells) and of 7±2 notes in sequences of arbitrary trichords (i.e., 7.32 notes). The fact that a limit of seven notes was found *only for experts* under conditions with *limited opportunity for chunking* supports the notion by Mathy and Feldman (2012) that the "magical number 7" results from the powerful structure detection of experts. Music students apparently managed to compress information even in sequences of arbitrary trichords, enabling the recall of seven notes. Hobby musicians lacked this powerful structure detection and hence were not able to compress information in these sequences, leading to the recall of only four notes.

To explain this phenomenon, Mathy and Feldman claimed that "Miller's magical number 7 is essentially an artifact of statistically typical compression ratios. That is, *random sequences* [...] contain some number of accidental patterns which result in chunking and thus compressed representations" (Mathy & Feldman, 2012, p. 359, italics added). The present stimuli, however, were not random but systematically ill-structured. The sequences of arbitrary trichords were created based on rules that were deliberately at odds with the regularities of the tonal system. So, while my findings provide evidence for the general notion that the "magical number 7" results from the powerful information compression of experts, the explanation of Mathy and Feldman does not fully account for them. As my stimuli were not random, the recall of seven elements in my data cannot result from the typical compression ratio in random stimuli.

My findings show that one of the most important pitfalls in chunking research is the hierarchical nature of chunking. In his article, Miller (1956) stated that the term *chunk* might denote a single sound in Morse code, the letter associated with a pattern of several such sounds, or the word multiple letters form. So, memory capacity limits can be approached from multiple perspectives, referring to these different hierarchical levels of a stimulus. In this work, the two "magical numbers" only became apparent as I considered recall in terms of notes (lower hierarchical level) *and* three-note melodic cells (higher hierarchical level). Moreover, my considerations show how important the comparison of different expertise levels is for the understanding of chunking. If the data presented by Miller (1956) would have included participants that were unfamiliar with the respective stimuli, the limit of seven elements might have never developed into an omnipresent "magical number".

Another pitfall of chunking is that the term is "used with a variety of meanings, which are often conflated, leading to considerable confusion" (Gobet et al., 2016, p. 1). In the literature that this work refers to, there are at least three different ways of conceptualizing chunking. First, there is the classical understanding coined by Miller (1956) that chunking is a process of data compression that can be found in memory tasks. Second, there is the notion proclaimed by Wolf (1976) that chunking denotes an increase in the efficiency in transcription tasks such as sight-reading by processing familiar constellations of symbols as single visual units. Third, there is the idea by Snyder (2016) that chunking denotes a deliberate segmentation of the musical surface when listening to a musical piece.

Of course, there are certain common aspects in all these conceptualizations of chunking. All of them are based on the idea that multiple single elements are treated as a single compound unit in cognitive processing – as a meaningful sequence of memoranda, a familiar constellation of note symbols, or a phrase in a musical piece. Moreover, chunking is beneficial in all three accounts, enabling more efficient refreshing, more efficient perception of note symbols, or memory representations with a clearer structure. Lastly, all accounts are based on knowledge or experience. This is most obvious in the former two, but even the segmentation of a piece of music into parts or phrases requires some knowledge on musical pieces and their structure.

Besides these similarities, however, there are various profound differences between these three accounts. Information compression is commonly assumed to be a distinct cognitive process that competes with other cognitive processes for attentional resources (Portrat et al., 2016). It utilizes information that has already been encoded. Thus, the different elements of a meaningful unit do not have to be presented simultaneously (Kalakoski, 2007). In sight-reading, chunking does not involve an additional cognitive process. The cognitive processes still comprise the retrieval of motor actions as a reaction to visual symbols. It is rather the associations between symbols and movements that change. In chunked sight-reading, instead of a single movement being associated with a single symbol, a whole sequence of movements is associated with a constellation of symbols. Lastly, the segmentation of heard music denotes a specific way of encoding in which a hierarchically structured internal representation is created in a conscious and deliberate act of cognition.

Because of the different cognitive processes that are tapped by the three accounts, the tasks in which evidence of them might be found also differ. Information compression might become apparent by recall accuracies and transitional error probabilities in serial recall tasks. Processing of constellations of symbols might be shown using performance accuracies and eye movements in sight-reading tasks. The segmentation of heard music into phrases might become apparent when a musical piece is played to participants, and they have to recall it by writing down musical notes.

So how can future studies avoid these conceptual problems I just described? First and foremost, a clear and precise definition of the terms *chunking* and *chunk* is needed whenever they are referred to. This is also demanded by Gobet et al. (2016): “progress in our understanding of chunking will be difficult until researchers recognize these different meanings and are more precise in the way they refer to them" (Gobet et al., 2016, p. 102). In doing so, researchers might refer to the relevant features of the stimulus, the relevant knowledge, the involved cognitive processes, and the ways in which they are beneficial for the performance in the task. For example, in the present complex span task, the relevant feature of the stimulus is its tonal structure, the relevant knowledge is the knowledge on major triads, the involved cognitive process is recoding, and the way in which it is beneficial is that it supports subsequent refreshing.

Another solution would be to start introducing more specific terms. However, if these terms are not chosen sensibly and are themselves not defined thoroughly, this proliferates rather than diminishes the problems. For example, Gobet et al. (2001) introduced the terms *goal-oriented chunking* and *perceptual chunking*. Other than one might intuitively assume, though, the term *perceptual chunking* does neither refer to the perception of symbol groups during reading nor to the segmentation of a perceived stimulus but denotes an automatic compression of information during perception. Apparently, Gobet realized that the terms were confusing, as he later changed them to *deliberate* and *automatic chunking* (Gobet et al., 2016).

6.3 Implications for sight-reading

In line with Chitalkina et al. (2021), Lim et al. (2019), and Drai-Zerbib et al. (2012), the present work found an association between eye movements and performance accuracy during sight-reading. Analysis revealed that reading with fewer fixations was associated with more accurate performance. I conclude that reading with fewer fixations might save cognitive resources as fewer eye movements have to be planned and less information needs to be integrated. The saved cognitive resources might be invested in the translation of visual information into motor actions which might lead to an increase in the accuracy of these actions. This finding shows that, in order to provide a comprehensive perspective on sight-reading, component skill models such as the one developed by Waters et al. (1998) should incorporate the role of eye movements.

In addition, the present results question the role of certain predictors used in the model of Kopiez and Lee (2008). Eye movements represent both a potential mediator and a potential covariate for certain of their predictors. In their study, working memory capacity was a significant predictor for sight-reading accuracy. Based on the present findings, however, it might be assumed that this effect was mediated by eye movements. A larger working memory capacity might support eye movement planning, which, in turn, might enable accurate sight-reading performance. Working memory capacity has been found to play a role in the generation of volitional eye movements

(Unsworth et al., 2004). In a study on text reading by Kennison and Clifton (1995), the number of progressive fixations was significantly smaller in participants with a large working memory capacity. Moreover, the short-term musical memory test and the inner hearing test used by Kopiez and Lee (2008) involved the reading of a melody in limited time. Thus, eye movement schemata that are beneficial for music reading might have facilitated performance in these tests. This would mean that eye movements are a covariate that needs to be statistically controlled when the association of these tests with sight-reading performance is investigated.

Besides these implications for theoretical models on sight-reading, the present study might also contribute to the understanding of the inconsistency of previous findings on the effect of musical expertise on eye movements during sight-reading. Goolsby (1994) and Gilman and Underwood (2003) used measures of performance accuracy merely to check if more experienced participants performed more accurately. They did not control for the statistical effect of accuracy measures in their analysis of the association between expertise and eye movements. In the study by Arthur et al. (2016) it was not reported if or how measures of performance accuracy were used in the analyses. Only the study by Penttinen et al. (2015) controlled for differences in performance accuracy by excluding inaccurate performances. As performance accuracy appears to be associated with eye movements and as more experienced participants might perform more accurately, the different ways of using performance accuracy measures in the analyses of Goolsby (1994), Gilman and Underwood (2003), Arthur et al. (2016), and Penttinen et al. (2015) might explain the inconsistency in their results.

As a by-product, the present analyses provided findings on the effect of expertise on eye movements during sight-reading. The main purpose of the Gold-MSI score was to control for the effect of expertise when analyzing the association of eye movements and performance accuracy. However, there is no reason why the resulting statistical effect should not be interpreted. In the regression models, Gold-MSI score was negatively associated with the number of fixations and the gaze duration on note pairs. This effect was found while the statistical effects of performance accuracy, practice, and type of note pair were controlled for. Shorter fixations have previously been found to be associated with expertise in silent reading of music (Penttinen et al., 2013), sight-reading (Truitt et al., 1997), and musical pattern matching (Waters et al., 1997). According to Sheridan et al. (2020) the fewer and shorter fixations of experts can be explained with their parafoveal preview benefit. Experts can perceive information from areas that are currently not fixated but lie in the periphery. This preview enables them to spare fixations and perceive notes rapidly.

6.4 Limitations

For the present work, I developed a new experimental task – the musical complex span task. While it provided some interesting findings, several aspects of it are worth to be pointed out critically. I adapted the procedure of the complex span task to the skill level of the participant groups. Consequently, the computational simulations needed to be performed separately for hobby musicians and music students, and the results could not be directly compared between these two sub-samples. A joint simulation and direct comparison would have provided much stronger insights. This case shows how challenging it can be to develop experimental procedures for participants of varying expertise levels.

Moreover, the distractor task, i.e., the sight-reading, did not allow to measure the time of attentional capture of single distractors. Thus, the *Ta* parameter needed to be estimated by fitting TBRS*C simulations to empirical data. This approach is uncommon and of questionable validity. Oberauer and Lewandowsky (2011) claim that the validity of the distractor task is dependent on the degree of experimental control over participants' strategies and over the time they devote to the processing component. The present task provided limited experimental control of both aspects. Besides, the sight-reading was rather challenging. Thus, excluding trials with inaccurate distractor processing, as is common practice in complex span tasks (Conway et al., 2005), would have led to a large loss of data. Hence, the accuracy in the distractor task was not restricted to a high level. Consequently, though, the meaning of certain TBRS*C parameters in the present study differed from their usual meaning. For example, the *Ta* parameter can commonly be interpreted as the time needed to *successfully* process a distractor. In the present case, it was unclear if distractors were processed successfully. Thus, I rather interpreted the parameter as the time *used* for distractor processing. The fact that parameters can be theoretically interpreted is a strength of computational simulations. The change in parameter meaning due to the present design limits this strength.

In addition, the distractor task was domain-specific, i.e., its difficulty varied across participants of different expertise levels. This is usually avoided as an easier distractor task might provide additional opportunities for refreshing. Lastly, the long presentation of memoranda and the periods of free time during the task (e.g., during the count-in or during the saving of the eye tracking data) might have allowed participants to apply mnemonic strategies that might have affected the outcomes of the recall task.

Lastly, the failure to create experimental procedures with a sufficient difficulty was problematic for both the simple as well as the complex span task. First, if a task has an insufficient difficulty, cognitive processes might change, as task requirements render certain cognitive processes unnecessary. Second, statistical information on the differences between experimental conditions might be lost due to ceiling effects. This might occur especially in highly experienced participants, as they perform most accurately. In the context of the present work, this might have been especially problematic, as my main assumption was that there would be a pronounced difference between experimental conditions for highly experienced participants.

6.5 Future studies

It is central for the phenomenon of chunking if a stimulus is meaningful in the respective domain. However, in the domain of music, it is especially difficult to systematically vary the meaningfulness of stimuli. In music, meaning is not a fixed concept but is individual and fluent, depending on the tonal context. In addition, this tonal context might be interpreted differently, depending on musicians' knowledge and musical background. In other words, a melodic sequence that appears arbitrary to one musician might seem meaningful for another. It might be even difficult to find melodic cells that do not have a clear label for musicians with a certain knowledge. The *theory of atonal music* (Forte, 1973) provides a system that allows to label all kinds of note sequences. A musician who is highly familiar with this theory might even have been able to label the arbitrary melodic cells used in the present work.

This has implications for the research on chunking in the processing of musical notation. In other domains, it would be most appropriate to create stimuli according to a clearly defined experimental design. However, due to the fluent nature of meaning in musical stimuli, I rather suggest future studies to gather extensive information on participants' individual perspective on the employed stimuli. After completing a serial recall task, participants could be asked how they interpreted the tonal context of the stimuli, if they recognized melodic cells, and if they used labels to memorize them. This would allow to unravel the individual role of clearly recognizable tonal contexts and of melodic cells with clear labels in chunking during processing musical notation. Additionally, such an approach might provide further insights into the kinds of note sequences that are perceived as more meaningful for musicians with a certain background.

Moreover, if future studies want to avoid the limitations of the complex span task described above, I suggest that they use notes as memoranda and a spatial judgment task as the distractor task. In such a spatial judgment task, participants have to react on the location of an object on the screen with a key press. For example, a black square might appear on either the left or the right side of the screen and participants would have to indicate the location as fast as possible with pressing either a key on the left or right, respectively (see Portrat et al., 2016). In addition, memoranda should be presented rather briefly (1,000 ms) and there should not be any free periods in the task over and above some brief interstimulus intervals. This procedure would have numerous advantages compared to the one used in the present study. It would allow to use the same procedure for participants of varying expertise levels. The distractor task would have the same difficulty for all participants. The reaction time of distractor processing could be measured. The distractor task would be easy and thus, trials with inaccurate performance could be excluded. By varying list length (e.g., 9, 12, 15, and 18 memoranda), ceiling effects could be avoided. Furthermore, this task procedure would allow to compare different ways of estimating the *Ta* parameter in the context of the TBRS*C computational model. *Ta* could be estimated based on the measured reaction times and by fitting the simulated to the empirical recall data. If the latter method would provide similar estimates than the former one, it could be established as a valid estimation method.

Lastly, future studies should investigate the causal relationships between eye movements and performance accuracy. To this end, I propose to employ a classic sight-reading task in which short, simple melodies have to be performed at first sight with performance tempo being controlled by a metronome. To check the assumption that sight-reading with fewer fixations requires less cognitive resources, I suggest to use the *Index of Cognitive Activity* (ICA, Marshall, 2002). It has been found that the size of the pupil is never constant but shows a high-frequency tremor. The ICA computes the frequency of this tremor, which is independent of changes in light. The ICA has been found to indicate cognitive workload across a range of tasks (Bartels & Marshall, 2012). Thus, the ICA can be used to check if cognitive workload is reduced when few fixations are performed during sight-reading. In addition, the MidiAnalyzer could be used to derive the location of errors in the performances. AOIs could be defined at these locations. Thereby, it would be possible to investigate if errors affect eye movements.

Besides these rather specific suggestions, I identified two general issues in sight-reading research. These are (1) the inherent confounding of visual and musical information in the system of musical notation and (2) the dependency of eye movements on musical and graphical parameters. For both issues, I will now provide some general suggestions how they might be addressed in future studies.

In the system of musical notation, musical information is represented by visual symbols. Thus, both types of information are inherently confounded. However, this renders it rather difficult to investigate what drives the eye movements used to read the notes. It would be useful to develop methods to decouple the visual and the musical dimension. Figure 6.2 shows one approach of how this might be achieved. Using a tie, as shown in the top row, allows to create notes with a sound of equal duration that differ in their visual appearance; using accidentals, as shown in the bottom row, allows to create notes with a sound of equal pitch that differ in their visual appearance. Future studies might use this logic to create melodies that sound identical but look differently. However, it should be noted that different ways of notation are used in different musical contexts and hence, evoke different associations and expectations about the progression of the melodies. Thus, if this approach is employed by future studies, the experimental design needs to be planned carefully to avoid confounds.

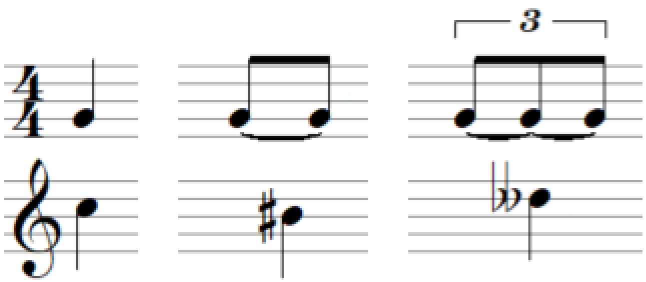

Figure 6.2. An approach to circumvent the confounding of visual and musical information in musical notation.

Another issue of sight-reading research is the dependency of eye movements on musical or graphical parameters. For example, the duration of fixations probably depends on the musical tempo, and the distance of saccades probably depends on the size of the staff. Accordingly, it might be helpful for future studies to develop music- and layout-dependent eye movement measures, such as the duration of fixations per beat duration, the number of fixations per beat, or the distance of saccades per bar size. Such measures could be intuitively interpreted, and they could be directly compared across studies.

6.6 Practical implications

From the age of nine to the age of fourteen, I took music lessons. Naturally, reading and performing notation was taught in these lessons. Then, with sixteen, I started to play in a band in which pieces were not written down. Rather, when one of us had come up with an idea for a melody, he performed it for the others, and we talked about it. After eight years of playing in that band, with the age of 24, I took lessons again. When my new teacher asked me if I was able to play notes, I denied. I had not done so for ten years and assumed that I had forgotten how to do it. When I started the first attempt, I was amazed. My eyes and my fingers new what to do by themselves. In quite a short time, I was able to perform complex notated pieces. In a sense, this anecdote exemplifies quite well the topic of the present work. LTM has an unlimited capacity. It can store huge amounts of information for very long periods. However, if certain information is not used, its accessibility ceases. Cues and external input can reactivate it and re-enable its usage. This notion supports the importance of institutionalized education. Learning in adolescence, when the brain is still flexible, builds a basis of knowledge that can be reactivated throughout life if needed. This demand for education is further supported by the idea, that expert skill is not limited to some few gifted individuals but develops gradually in everyone that attends to deliberate practice. In other words, if anyone is a potential expert, everybody deserves to be fostered to their full potential.

With respect to specific teaching methods, the present work supports the claim that "curricula should provide means to acquire [...] chunks in a given domain" (Gobet et al., 2001, p. 241). Chunking improves working memory, which, in turn, might facilitate domain-specific tasks and activities. So, while expert memory might develop as a by-product of domain activities, it should nevertheless be specifically fostered by training and education. While doing so, teachers should direct learners' attention to the crucial features of a given material (Gobet, 2005). In addition, it might be helpful to gather diagnostic feedback on the knowledge of chunks. For example, computer programs might be designed that show information of the given domain with increasing complexity. The task of students could be to simply name the depicted information as fast as possible. This would provide valuable insights, if chunks are known to students and how rapidly they can access them. This principle is already implemented in some commercial apps (for a list of examples, see https://nolaschoolofmusic.com/blog/best-sight-reading-music-note-apps-for-iphone-ipad) and is the basis of a test that has been developed at our lab for research purposes. As discussed by Anderson et al. (2000), this teaching method is opposed to the logic of situated learning and constructivism that knowledge cannot be decomposed and decontextualized.

With respect to sight-reading, the present study implies that teachers might look not only on students' movements on the instrument but might also observe their eyes. The eyes should move smoothly and steadily. If the eyes move chaotically with many fixations, the teacher should intervene. Then, it would be important to check if the student knows where to find the required information. Moreover, single phrases might be practiced separately and in a slower tempo. This strategy would allow to learn the patterns contained in the melody which would enable to read them with fewer fixations. Of course, the information that is provided by simply observing eye movements is not as detailed as the one provided by an eye tracker. Nevertheless, the eyes play such an important role during the performance of musical notes that it would not be warranted to completely ignore them as a teacher.

6.7 Conclusion

To conclude, the present work provided several methodological developments to the research on expertise in the processing of musical notation: (1) I demonstrated how musical stimuli might be systematically created based on music theory, (2) I showed how the TBRS*C computational model can be combined with Bayesian regression models to analyze group differences in cognitive processes, and (3) I designed an open-source software tool for the assessment of musical performance MIDI data.

In addition, this work contributed to the development of the theoretical understanding of expert processing of musical notation. I extended findings by Deutsch (1980) that a systematic tonal structure supports chunking, by developing the idea that, given a melodic sequence has a systematic tonal structure, chunking is further supported by the combination of a clear tonal context with melodic cells with clear labels. Based on the idea that working memory in complex span tasks involves a rapid switching between task components (Barrouillet & Camos, 2007) and a chunking process (Portrat et al., 2016), I developed a novel conceptualization of expert memory that comprises rapid LTM access and reliable chunking. Lastly, starting from initial findings on the association of eye movements with sight-reading accuracy (Chitalkina et al., 2021; Drai-Zerbib et al., 2012; Lim et al., 2019; Zhukov et al., 2019), I established the notion that reading with few fixations might be beneficial for sight-reading as fewer eye movements have to planned and less information from subsequent fixations needs to be integrated. It is these methodological and theoretical developments that constitute the essence of the present work and took us one step further towards a comprehensive understanding of expertise in the processing of musical notation.

7. References

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

Ahken, S., Comeau, G., Hébert, S., & Balasubramaniam, R. (2012). Eye movement patterns during the processing of musical and linguistic syntactic incongruities. *Psychomusicology: Music, Mind, and Brain*, *22*(1), 18–25. https://doi.org/10.1037/a0026751

Anderson, J. R. (1990). *Cognitive psychology and its implications* (3. ed.). *A series of books in psychology*. Freeman.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. https://doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., Reder, L. M., & Simon, H. A. (2000). Applications and Misapplications of Cognitive Psychology to Mathematics Education. *Texas Education Review*.

Arthur, P., Khuu, S., & Blom, D. (2016). Music sight-reading expertise, visually disrupted score and eye movements. *Journal of Eye Movement Research*, *9*(7).

Atkinson, R. C., & Shiffrin, R. M. (1968). Human Memory: A Proposed System and its Control Processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of Learning and Motivation: v.2. Psychology of Learning and Motivation: Advances in Research and Theory* (Vol. 2, pp. 89–195). Elsevier. https://doi.org/10.1016/S0079-7421(08)60422-3

Baddeley, A. D. (1990). *Human memory: Theory and practice*. Erlbaum.

Barrouillet, P., Bernardin, S., & Camos, V. (2004). Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, *133*(1), 83–100. https://doi.org/10.1037/0096-3445.133.1.83

Barrouillet, P., & Camos, V. (2007). The time-based resource-sharing model of working memory. In N. Osaka, R. H. Logie, & M. D'Esposito (Eds.), *Working memory: Behavioural and neural correlates* (pp. 59–80). Oxford University Press.

Bartels, M., & Marshall, S. P. (2012). Measuring cognitive workload across different eye tracking hardware platforms. In *Proceedings of the Symposium on Eye Tracking Research and Applications* (pp. 161–164). ACM. https://doi.org/10.1145/2168556.2168582

Bellezza, F. S., & Buck, D. K. (1988). Expert knowledge as mnemonic cues. *Applied Cognitive Psychology*, *2*(2), 147–162. https://doi.org/10.1002/acp.2350020206

Bertram, R., Helle, L., Kaakinen, J. K., & Svedström, E. (2013). The effect of expertise on eye movement behaviour in medical image perception. *PLoS ONE*, *8*(6), e66169. https://doi.org/10.1371/journal.pone.0066169

Binet, A. (1894). *Psychologie des grands calculateurs et joueurs d'échecs [Psychology of great mental calculators and chess players]*. Librairie Hachette.

Bloom, B. S. (1985). Generalizations about talent development. In B. S. Bloom (Ed.), *Developing talent in young people* (pp. 507–549). Ballantine Books.

Botvinick, M. M. (2005). Effects of domain-specific knowledge on memory for serial order. *Cognition*, *97*(2), 135–151. https://doi.org/10.1016/j.cognition.2004.09.007

Botvinick, M. M., & Bylsma, L. M. (2005). Regularization in short-term memory for serial order. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *31*(2), 351–358. https://doi.org/10.1037/0278-7393.31.2.351

Broadbent, D. E. (1958). *Perception and communication*. Pergamon Press. https://doi.org/10.1037/10037-000

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

Camos, V., Johnson, M., Loaiza, V., Portrat, S., Souza, A., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences*, *1424*(1), 19–32. https://doi.org/10.1111/nyas.13616

Cara, M. A. (2018). Anticipation awareness and visual monitoring in reading contemporary music. *Musicae Scientiae*, *22*(3), 322–343. https://doi.org/10.1177/1029864916687601

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency and the growth of short-term memory span. *Journal of Experimental Child Psychology*, *33*(3), 386–404. https://doi.org/10.1016/0022-0965(82)90054-6

Charness, N. (1976). Memory for chess positions: Resistance to interference. *Journal of Experimental Psychology: Human Learning and Memory*, *2*(6), 641–653. https://doi.org/10.1037/0278-7393.2.6.641

Chase, W. G., & Ericsson, K. A. (1982). Skill and working memory. *Psychology of Learning and Motivation*, *16*, 1–58.

Chase, W. G., & Simon, H. A. (1972). The mind’s eye in chess. In *Eighth Annual Carnegie Symposium on Cognition.* Symposium conducted at the meeting of Carnegie-Mellon University, Pittsburgh, Pennsylvania.

Chitalkina, N., Puurtinen, M., Gruber, H., & Bednarik, R. (2021). Handling of incongruences in music notation during singing or playing. *International Journal of Music Education*, *39*(1), 18–38. https://doi.org/10.1177/0255761420944036

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user’s guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786. https://doi.org/10.3758/BF03196772

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*(1), 87–185. https://doi.org/10.1017/S0140525X01003922

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? *Progress in Brain Research*, *169*, 323–338. https://doi.org/10.1016/S0079-6123(07)00020-9

Croonen, W. L. M. (1991). Recognition of Tone Series Containing Tonic Triads. *Music Perception: An Interdisciplinary Journal*, *9*(2), 231–250. https://doi.org/10.2307/40285530

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466. https://doi.org/10.1016/S0022-5371(80)90312-6

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review*, *3*(4), 422–433. https://doi.org/10.3758/BF03214546

de Groot, A. (1946). *Thought and choice in chess*. Mouton.

de Groot, A., & Gobet, F. (1996). *Perception and Memory in Chess: Heuristics of the Professional Eye*. Van Gorcum.

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389.

Drai-Zerbib, V., Baccino, T., & Bigand, E. (2012). Sight-reading expertise: Cross-modality integration investigated using eye tracking. *Psychology of Music*, *40*(2), 216–235. https://doi.org/10.1177/0305735610394710

Ericsson, K. A. (2018a). The Differential Influence of Experience, Practice, and Deliberate Practice on the Development of Superior Individual Performance of Experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 745–769). Cambridge University Press. https://doi.org/10.1017/9781316480748.038

Ericsson, K. A. (2018b). Superior Working Memory in Experts. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 696–713). Cambridge University Press. https://doi.org/10.1017/9781316480748.036

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist*, *49*(8), 725–747. https://doi.org/10.1037/0003-066X.49.8.725

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245. https://doi.org/10.1037/0033-295X.102.2.211

Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, *100*(3), 363–406. https://doi.org/10.1037/0033-295X.100.3.363

Ericsson, K. A., & Lehmann, A. C. (1996). Expert and exceptional performance: Evidence of maximal adaptation to task constraints. *Annual Review of Psychology*, *47*, 273–305. https://doi.org/10.1146/annurev.psych.47.1.273

Ericsson, K. A., & Staszewski, J. J. (1989). Skilled memory and expertise: Mechanisms of exceptional performance. In *Complex information processing: The impact of Herbert A. Simon* (pp. 235–267). Lawrence Erlbaum Associates, Inc.

Fadde, P. J., & Jalaeian, M. (2019). Learning with Zeal. In P. Ward, J. Maarten Schraagen, J. Gore, E. M. Roth, & J. Mishra (Eds.), *The Oxford Handbook of Expertise* (pp. 926–950). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198795872.013.40

Forte, A. (1973). *The structure of atonal music*. Yale Univ. Press.

Frey, P. W., & Adesman, P. (1976). Recall memory for visually presented chess positions. *Memory & Cognition*, *4*(5), 541–547. https://doi.org/10.3758/BF03213216

Gilchrist, A. L. (2015). How should we measure chunks? A continuing issue in chunking research and a way forward. *Frontiers in Psychology*, *6*, 1456. https://doi.org/10.3389/fpsyg.2015.01456

Gilman, E., & Underwood, G. (2003). Restricting the field of view to investigate the perceptual spans of pianists. *Visual Cognition*, *10*(2), 201.

Gobet, F. (1998). Expert memory: a comparison of four theories. *Cognition*, *66*(2), 115–152. https://doi.org/10.1016/S0010-0277(98)00020-1 (Cognition, 66(2), 115-152).

Gobet, F. (2005). Chunking models of expertise: implications for education. *Applied Cognitive Psychology*, *19*(2), 183–204. https://doi.org/10.1002/acp.1110

Gobet, F. (2012). Development of Expertise. In *Encyclopedia of the Sciences of Learning* (pp. 950–953). Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6_1913

Gobet, F. (2019). The Classic Expertise Approach and Its Evolution. In P. Ward, J. Maarten Schraagen, J. Gore, E. M. Roth, & J. Mishra (Eds.), *The Oxford Handbook of Expertise* (pp. 34–55). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198795872.013.2

Gobet, F., Lane, P. C., Croker, S., Cheng, P. C.-H., Jones, G., Oliver, I., & Pine, J. M. (2001). Chunking mechanisms in human learning. *Trends in Cognitive Sciences*, *5*(6), 236–243. https://doi.org/10.1016/S1364-6613(00)01662-4

Gobet, F., Lloyd-Kelly, M., & Lane, P. C. R. (2016). What's in a Name? The Multiple Meanings of "Chunk" and "Chunking". *Frontiers in Psychology*, *7*, 102. https://doi.org/10.3389/fpsyg.2016.00102

Gobet, F., & Simon, H. A. (1996). Templates in chess memory: A mechanism for recalling several boards. *Cognitive Psychology*, *31*, 1–40. https://doi.org/10.1006/cogp.1996.0011

Goolsby, T. (1994). Eye Movement in Music Reading: Effects of Reading Ability, Notational Complexity, and Encounters. *Music Perception: An Interdisciplinary Journal*, *12*(1), 77–96. https://doi.org/10.2307/40285756

Halpern, A. R., & Bower, G. H. (1982). Musical expertise and melodic structure in memory for musical notation. *The American Journal of Psychology*, *95*(1), 31–50.

Hayes, J. R. (1981). *The complete problem solver*. Franklin Inst. Press.

Holmqvist, K., & Andersson, R. (2018). *Eye tracking: A comprehensive guide to methods, paradigms and measures* (2nd edition). Lund Eye-Tracking Research Institute.

Jones, T., & Farrell, S. (2018). Does syntax bias serial order reconstruction of verbal short-term memory? *Journal of Memory and Language*, *100*, 98–122. https://doi.org/10.1016/j.jml.2018.02.001

Kalakoski, V. (2007). Effect of skill level on recall of visually presented patterns of musical notes. *Scandinavian Journal of Psychology*, *48*(2), 87–96. https://doi.org/10.1111/j.1467-9450.2007.00535.x

Kennison, S. M., & Clifton, C. (1995). Determinants of parafoveal preview benefit in high and low working memory capacity readers: Implications for eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(1), 68–81. https://doi.org/10.1037/0278-7393.21.1.68

Kopiez, R., & Lee, J. I. (2006). Towards a dynamic model of skills involved in sight reading music. *Music Education Research*, *8*(1), 97–120. https://doi.org/10.1080/14613800600570785

Kopiez, R., & Lee, J. I. (2008). Towards a general model of skills involved in sight reading music. *Music Education Research*, *10*(1), 41–62. https://doi.org/10.1080/14613800701871363

Land, M. F., & Furneaux, S. (1997). The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, *352*(1358), 1231–1239.

Lehmann, A. C. (2005). Vomblattspielen und Notenlesen. In T. H. Stoffer, R. Oerter, & N.-P. Birbaumer (Eds.), *Enzyklopädie der Psychologie Praxisgebiete Musikpsychologie: Bd. 1. Allgemeine Musikpsychologie* (pp. 883–890). Hogrefe.

Lehmann, A. C., Gruber, H., & Kopiez, R. (2018). Expertise in Music. In K. A. Ericsson, R. R. Hoffman, A. Kozbelt, & A. M. Williams (Eds.), *The Cambridge Handbook of Expertise and Expert Performance* (pp. 535–549). Cambridge University Press. https://doi.org/10.1017/9781316480748.028

Lim, Y., Park, J. M., Rhyu, S.-Y., Chung, C. K., Kim, Y., & Yi, S. W. (2019). Eye-Hand Span is not an Indicator of but a Strategy for Proficient Sight-Reading in Piano Performance. *Scientific Reports*, *9*(1), 17906. https://doi.org/10.1038/s41598-019-54364-y

Macnamara, B. N., Moore, A. B., & Conway, A. R. A. (2011). Phonological similarity effects in simple and complex span tasks. *Memory & Cognition*, *39*(7), 1174–1186. https://doi.org/10.3758/s13421-011-0100-5

Madell, J., & Hébert, S. (2008). Eye Movements and Music Reading: Where Do We Look Next? *Music Perception: An Interdisciplinary Journal*, *26*(2), 157–170. https://doi.org/10.1525/mp.2008.26.2.157

Marshall, S. P. (2002). The Index of Cognitive Activity: Measuring cognitive workload. In *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*. IEEE. https://doi.org/10.1109/HFPP.2002.1042860

Mathy, F., & Feldman, J. (2012). What's magic about magic numbers? Chunking and data compression in short-term memory. *Cognition*, *122*(3), 346–362. https://doi.org/10.1016/j.cognition.2011.11.003

Meinz, E. J., & Salthouse, T. A. (1998). The effects of age and experience on memory for visually presented music. *The Journals of Gerontology*, *53*(1), 60-69. https://doi.org/10.1093/geronb/53B.1.P60

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81.

Mishra, J. (2014). Factors Related to Sight-Reading Accuracy. *Journal of Research in Music Education*, *61*(4), 452–465. https://doi.org/10.1177/0022429413508585

Mishra, J. (2019). Musical Expertise. In P. Ward, J. Maarten Schraagen, J. Gore, E. M. Roth, & J. Mishra (Eds.), *The Oxford Handbook of Expertise* (pp. 573–593). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198795872.013.25

Müllensiefen, D., Gingras, B., Musil, J., & Stewart, L. (2014). The musicality of non-musicians: An index for assessing musical sophistication in the general population. *PLoS ONE*, *9*(2). https://doi.org/10.1371/journal.pone.0089642

Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the Time-Based Resource-Sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. https://doi.org/10.3758/s13423-010-0020-6

Pashler, H. E. (1999). *The psychology of attention*. *A Bradford book*. MIT Press.

Penttinen, M., Huovinen, E., & Ylitalo, A.-K. (2013). Silent music reading: Amateur musicians' visual processing and descriptive skill. *Musicae Scientiae*, *17*(2), 198–216. https://doi.org/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*, *33*(1), 36–50. https://doi.org/10.1177/0255761413515813

Portrat, S., Guida, A., Phénix, T., & Lemaire, B. (2016). Promoting the experimental dialogue between working memory and chunking: Behavioral data and simulation. *Memory & Cognition*, *44*(3), 420–434. https://doi.org/10.3758/s13421-015-0572-9

Puurtinen, M. (2018). Eye on Music Reading: A Methodological Review of Studies from 1994 to 2017. *Journal of Eye Movement Research, 11*(2). https://doi.org/10.16910/jemr.11.2.2

Ritchie, G., Tolan, G. A., & Tehan, G. (2015). Redintegration, task difficulty, and immediate serial recall tasks. *Canadian Journal of Experimental Psychology*, *69*(1), 54–63. https://doi.org/10.1037/cep0000031

Rosemann, S., Altenmuller, E., & Fahle, M. (2015). 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. https://doi.org/10.1177/0305735615585398

Royston, P., Altman, D. G., & Sauerbrei, W. (2006). Dichotomizing continuous predictors in multiple regression: A bad idea. *Statistics in Medicine*, *25*(1), 127–141. https://doi.org/10.1002/sim.2331

Saariluoma, P. (1989). Chess players' recall of auditorily presented chess positions. *European Journal of Cognitive Psychology, 1*(4), 309–320. https://doi.org/10.1080/09541448908403091

Sala, G., & Gobet, F. (2017). Experts' memory superiority for domain-specific random material generalizes across fields of expertise: A meta-analysis. *Memory & Cognition*, *45*(2), 183–193. https://doi.org/10.3758/s13421-016-0663-2

Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrenheit anhand einer deutschen Stichprobe. *Musicae Scientiae*, *18*(4), 423–447. https://doi.org/10.1177/1029864914541851

Shah, P., & Miyake, A. (1999). Models of Working Memory. An Introduction. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 1–27). Cambridge University Press.

Sheridan, H., Maturi, K. S., & Kleinsmith, A. L. (2020). Eye movements during music reading: Toward a unified understanding of visual expertise. In *Psychology of learning and motivation* (Vol. 73, pp. 119–156). https://doi.org/10.1016/bs.plm.2020.07.002

Simon, H. A., & Barenfeld, M. (1969). Information-processing analysis of perceptual processes in problem solving. *Psychological Review*, *76*(5), 473–483. https://doi.org/10.1037/h0028154

Snyder, B. (2016). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *The Oxford handbook of music psychology* (2nd ed., pp. 107–117). Oxford University Press.

Stanley, J., & Krakauer, J. W. (2013). Motor skill depends on knowledge of facts. *Frontiers in Human Neuroscience*, *7*, 503. https://doi.org/10.3389/fnhum.2013.00503

Tanabe, A., & Osaka, N. (2009). Picture span test: Measuring visual working memory capacity involved in remembering and comprehension. *Behavior Research Methods*, *41*(2), 309–317. https://doi.org/10.3758/BRM.41.2.309

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. https://doi.org/10.1080/713756756

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154. https://doi.org/10.1016/0749-596X(89)90040-5

Unsworth, N., Schrock, J. C., & Engle, R. W. (2004). Working Memory Capacity and the Antisaccade Task: Individual Differences in Voluntary Saccade Control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *30*(6), 1302–1321. https://doi.org/10.1037/0278-7393.30.6.1302

Waters, A. J., Townsend, E., & Underwood, G. (1998). Expertise in musical sight reading: A study of pianists. *British Journal of Psychology*, *89*(1), 123.

Waters, A. J., Underwood, G., & Findlay, J. M. (1997). Studying expertise in music reading: Use of a pattern-matching paradigm. *Perception & Psychophysics*, *59*(4), 477–488. https://doi.org/10.3758/BF03211857

Weisberg, R. W. (1999). Creativity and knowledge: A challenge to theories. In J. Sternberg (Ed.), *Handbook of creativity* (pp. 226–250). Cambridge University Press.

Wever, E. G., & Zener, K. E. (1928). The method of absolute judgment in psychophysics. *Psychological Review*, *35*(6), 466–493. https://doi.org/10.1037/h0075311

Wolf, T. (1976). A cognitive model of musical sight-reading. *Journal of Psycholinguistic Research*, *5*(2), 143–171. https://doi.org/10.1007/BF01067255

Zhang, J. D., Schubert, E., & McPherson, G. E. (2020). Aspects of music performance that are most highly related to musical sophistication. *Psychomusicology: A Journal of Research in Music Cognition*, *30*(2), 64–71. https://doi.org/10.1037/pmu0000252

Zhukov, K., Khuu, S., & McPherson, G. E. (2019). Eye-movement efficiency and sight-reading expertise in woodwind players. *Journal of Eye Movement Research*, *12*(2). https://doi.org/10.16910/jemr.12.2.6

A The MidiAnalyze program

MidiAnalyze – A Python package for the Analysis of Musical Performance MIDI Data

Author	Lucas Lörch; LucasLoerch@posteo.de; ORCID https://orcid.org/0000-0002-5665-3313
Download	https://doi.org/10.17605/OSF.IO/FKW4B
License	Creative Commons Attribution-NonCommercial 3.0 Unported License (CC BY-NC 3.0)
Cite	Lörch, L. (2021). MidiAnalyze. A Python Package for the Analysis of Musical Performance MIDI Data. https://doi.org/10.17605/OSF.IO/FKW4B

General idea and purpose

MidiAnalyze is a Python software package to analyze musical performances on MIDI instruments. It has been developed for usage in the context of scientific experiments. The program uses two kinds of MIDI files: *stimulus melodies*, i.e., MIDI files that hold the notes that participants were asked to perform, and *performances*, i.e., MIDI files that hold what the participants played. The functions of the program compare the performances with the stimulus melodies and store the resulting accuracy measures in a spreadsheet.

Technical specifications

MidiAnalyze was developed and tested with Python 3.7.1 on Spyder IDE 4.0.0b7 under Windows 10. It has not yet been tested on other operating systems or with other versions of Python. It requires the Python packages *music21*, *sys*, *os*, *glob* and *pandas*.

Instructions

1. Prepare software and MIDI files
   - Install Python and the packages music21, sys, os, glob and pandas. These packages are needed by MidiAnalyze.
   - Open the file MidiAnalyze v1-0.py in Python and run the whole code. Now the functions of MidiAnalyze are available.
   - Create one folder that contains MIDI files of all the original stimulus melodies, i.e., of the melodies participants had to perform during your experiment. These MIDI files need to be named according to the item name. For example, if one melody constitutes the first item of condition 1, the filename could be *condition1_item1.mid.* When choosing a name,

consider the overall number of conditions and items. If you would have, for example, more than 10 conditions and more than 10 items, you should name your files *condition01_item01.mid*

- Create one folder that contains all the experimental data, i.e. the MIDI files that you collected during participants' performances. These MIDI files need to be named according to the participant identifier and the item name. For example, if participant 15 performed item 1 of condition 1, you could name the file *participant15_condition1_item1.mid.* It is very important that the names of the MIDI files holding the original stimulus melodies and the names of the MIDI files holding the experimental performances are congruent. Only if this is the case, the stimulus melodies can be assigned correctly to the experimental performances. In the example above, if you would name your experimental file *participant15_condition_1_item_1.mid*, the program would not be able to assign the stimulus melody *condition1_item1.mid* to it, as the item identifier differs. Moreover, it is important that your MIDI files contain only one performance and that this performance is aligned with the beat. In some experiments, one might just let the recording running during the whole experiment, comprising the performance of multiple stimulus melodies. If this is the case, it is necessary to use some software that is able to edit MIDI and to extract the single melodies and export them as MIDI files.

2. Import the MIDI files to MidiAnalyze

- Run the command getyourMidiFiles(performanceDataPath, solutionDataPath, startItemName, endItemName, startSubjectName, endSubjectName) in Python.
  As *perfomanceDataPath*, enter the path to the folder with the performance data in quotation marks. Note that Python requires double-backslashes in filepaths under Windows. For example, if your performance data is in a folder "Experiment" on the desktop, the path would be "C:\\Users\\YourName\\Desktop\\Experiment\\".
  As *solutionDataPath* enter the path to the folder with the original stimulus melodies in quotation marks.
  As *startItemName* and *endItemName*, enter the position at which the item identifier starts and ends in the file names of your performance MIDI files. Note that in Python, the first position is indicated by 0. For example, if your file name would be "participant15_condition1_item1.mid", the item identifier (which is "condition1_item1") starts at the 14th position (at the c) so startItemName would be 14. The identifier ends at the 30th position (at the .) so endItemName would be 30.
  As *startSubjectName* and *endSubjectName* enter the position at which the participant identifier starts and ends in the filenames of your performance MIDI files in the same manner. So the full command could for example read
  getYourMidiFiles("C:\\Users\\YourName\\Desktop\\Experiment\\",
  "C:\\Users\\YourName\\Desktop\\Stimuli\\", 14, 30, 0, 13)
- If needed, all MIDI files can be quantized by running the command quantizeMidi(noteValue, quantizeOffsets, quantizeDurations) in Python.

As *noteValue*, indicate on which note value the quantization should be based (1=quarter, 2=eighth, 4=sixteenth, 0.5=half, 0.25=whole, 3=eighth triplets). Set the number in square brackets. If *quantizeOffsets* is set to TRUE, the beginning of each note is quantized. If *quantizeDurations* is set to TRUE, the end of each note is quantized. The quantization moves the beginning and/or the duration of each note to the closest note with the value specified. For example, if the specified note value is eighth note and quantizeOffsets and quantizeDurations is set to TRUE, both the beginning and the end of each note is moved to the closest eighth note. The whole command in this case would be quantizeMidi([2], TRUE, TRUE). Note that the MIDI information is only changed internally. The MIDI files themselves are not affected by the quantization.

3. Analyze the performances and export the resulting spreadsheet
   - Run the command analyzeMidi() in Python. MidiAnalyze calculates the accuracy measures and descriptives for all performance files and stores them in a spreadsheet.
   - Run the command exportResults(resultsFileName) in Python.
     As *resultsFileName*, indicate how you want to name the file that will be generated. Use quotation marks. MidiAnalyze then exports all results as a .csv spreadsheet in the folder with the original stimulus melodies. The full command could for example read exportResults("MyExperiment_results"). Then the file MyExperiment_results.csv would be created in "C:/Users/YourName/Desktop/Stimuli".

Functions

Utility functions

These functions handle the steps necessary to perform the analysis. They import the MIDI files and create or export the resulting spreadsheet.

Function	Description
indicateFilePaths("performanceDataPath", "solutionDataPath")	Checks and defines the two central file paths, i.e., the performanceDataPath which is the path to the MIDI files that resulted from participants' performance and the solutionDataPath which is the path to the MIDI files that hold the stimulus melodies.
importMidiFiles("performanceDataPath", startItemName, endItemName, startSubjectName, endSubjectName)	Imports all performance MIDI files. Creates a spreadsheet with a row for each file. If the filenames of the MIDI files contain item and

	participant identifiers the location of these identifiers in the filename can be indicated and then, these identifiers are also added to the spreadsheet. Creates the columns *MidiFile, item, participant* in the spreadsheet.
importCorrectSolutions(“solutionDataPath”)	Imports all stimulus melodies.
addCorrectSolutions()	Matches the stimulus melodies to the performances. Works only if the item identifier is identical in filenames of stimulus melodies and performances. Creates the column *correct* in the spreadsheet.
quantizeMidi([noteValue], quantizeOffsets, quantizeDurations)	Quantizes the performances. If quantizeOffsets is set to TRUE, the beginning of each note is quantized. If quantizeDurations is set to TRUE, the end of each note is quantized. The quantization value is specified by noteValue (1=quarter notes, 2=eighth, 4=sixteenth, 0.5=half, 0.25=whole, ect.). Beginning and/or end of the note are moved to the closest note value.
exportResults()	Creates a .csv file from the spreadsheet and stores it in the folder with the stimulus melodies.

Compare functions

These functions compare the performances with the stimulus melodies on three main parameters: the beginning (also called position), pitch and duration (also called note value) of the notes. These functions are the main part of the program as they provide an analysis of performance accuracy.

Function	Description
compareNumberOfNotes()	Creates a column *ommissionAddition* and compares the number of performed notes with the number of notes in the stimulus melodies. 0 means that the correct number of notes was performed. -3 means that the performance contained three notes less than the stimulus melody. +1 means that the performance contained one note more than the stimulus melodies.
compareNotePositions()	Compares the onset of each note in each performance with the stimulus melody and stores a relative accuracy value in the column *ACC_notePosition* (0.5 means that 50% of the performed notes started at the correct position)
compareNotePositionsOfSingleNotes()	For each note X in each stimulus melody, a variable *ACC_notePosition_noteX* is created in the spreadsheet. For each performance, the program indicates if it contains a note that starts at this positions (1) or not (0).
comparePitch()	Compares the pitch of each note in each performance that starts at a correct position and stores a relative accuracy value in the column *ACC_pitch* (0.5 means that 50% of the notes that started at the correct position had a correct pitch). The octave is not considered, i.e. if C4 had to be played and C5 was played, this counted as correct.
comparePitchWithOctave()	Same as comparePitch, but takes octaves into account, i.e. if C4 had to be played and C5 was

	played, this is counted as wrong. Stores a relative accuracy value in the column *ACC_pitchWithOctave*
comparePitchOfSingleNotes()	For each note Y in each performance that started at a correct position, a variable *ACC_pitch_noteY* is created in the spreadsheet. The program indicates if this note had a correct pitch (1) or not (0).
compareDuration()	Compares the duration of each note in the performances that starts at a correct position and stores a relative accuracy value in the column *ACC_duration* (0.5 means that 50% of the notes that started at the correct position had a correct duration).
compareDurationOfSingleNotes()	For each note Z in each performance that starts at a correct position, a variable *ACC_duration_noteZ* is created in the spreadsheet. The program indicates if this note had a correct duration (1) or not (0).

Describe functions

These functions do not compare but describe the performances. This can be especially useful to check the validity of the performance MIDI files by checking if certain measures such as the length of the performance or the pitch range take plausible values.

Function	Description
describeNumberOfNotes()	Creates a column *numberOfPerformedNotes* in the spreadsheet and indicates the number of notes in the performances.
describeNumberOfChords()	Creates a column *numberOfPerformedChords* in the spreadsheet and indicates the number of chords (simultaneously performed notes) in the performances.

describePitchSpan()	Creates a column *pitchRange* in the spreadsheet and indicates the lowest and highest pitches of the performances.
describeNoteValues()	Creates a column *performedNoteValues* and indicates a list of which note values were contained in the performances.
describeDuration()	Creates a column *durationOfPerformanceInQuarterNotes* and indicates the duration of the performances in quarter notes.
describeNumberOfBars()	Creates a column *numberOfPerformedBars* and indicates the length of the performances in bars. Works only if the MIDI files contain bar markers.
describeMeter()	Creates a column *meter* and indicates the meter of the performances.
describeInstrument()	Creates a column *instrument* and indicates the instrument information of the MIDI files.
describeClef()	Creates a column *clef* and indicates the clef of the performances.
describeTempo()	Creates a column *tempo* and indicates the tempo of the performances
describeKey()	Creates a column *key* and indicates the key of the performances.
describePerformedNotes()	Creates a column *performedNotes* that contains the whole performance in the format [{Note1: [{offset: 0}, {pitch:A4}, {notevalue:0.5}]},{Note2:...}] So for each note, its beginning (offset), its pitch and its duration (notevalue) is indicated.

Comprehensive functions

These functions integrate several of the previous function in order to support usability. They do not provide anything new, but only allow to call sets of functions with one command. In principle, all functions of the program can be called with the getYourMidiFiles and the analyzeMidi functions.

Function	Description
getYourMidiFiles (performanceDataPath, solutionDataPath, startItemName, endItemName, startSubjectName, endSubjectName)	Calls the functions indicateFilePaths, importMidiFiles, importCorrectSolutions and addCorrectSolutions.
compareMidi()	Calls the functions compareNumberOfNotes, compareNotePositions, comparePitch and compareDuration
describeMidi()	Calls the functions describeNumberOfNotes(), describePitchSpan(), describeNoteValues(), describeDuration(), and describePerformedNotes(). If the performance MIDI files contain information on instrument, clef, tempo, key, bars and if they contain chords, the respective describe functions are called.
analyzeMidi()	Calls both the describeMidi and the compareMidi functions

B Acknowledgments

First and foremost, I want to thank Stefan and Erkki for supporting me and making this work possible. They always were there when I needed them; they inspired me with their great thoughts and made it enjoyable to continue this work. They were the best supervisors I could wish for.

Moreover, I want to thank all my colleagues who also accompanied me through this project. I want to thank Benni for helping me with the eye tracking data, for the great conference parties, and for being a great office neighbor, Annso for going along with me since Bachelor times and always being a great colleague and friend, and Tim for showing me how great it is to work in science. I want to thank Nicola and Ios who always helped to navigate through the jungle of German bureaucracy with patience and friendliness, my Hiwis, namely Lisa, Leo, Anna, and Taniya, for their thorough work during data collection and pre-processing. Thanks to the colleagues from the "Päda" chair and from the junior professorship, that is, Oliver, Stefan, Tamara, Meike, Marc-Philipp, Belinda, Laura, Hanna Beißert, Julia, Saras, Sophie, Karina, Steffi, Hanna Bürkle, and all the new colleagues I do not yet know so well. Thank you for the many great lunch conversations, Kolloq discussions, and Christmas parties. I want to thank my fellow CDSS PhD students, Mirka, Sebastian, Stella, Sophie, Lilly, and all the others, for providing their input, sitting with me through lengthy courses, and helping me get through the initial phase of the project.

I want to thank all colleagues outside the University of Mannheim. My friends from Stockholm, especially Maria, Sheng-Ying, Inka, and Irene for their warm welcome and for making my stay there such a great time. Thanks to Benoît and Sophie, for their openness and for trusting me without even knowing me. They showed me that science should be a joint endeavor and not one of competition and hostility. Thanks to Martina Benz for enabling data collection at the Mannheim University of Music, and to Elke and Marjaana for their great feedback on my ideas.

I send heaps of love, thanks, and hugs to my family and friends. I want to thank my father for teaching me the importance of knowledge, effort, and endurance. I want to thank my mother for teaching me kindness, compassion, and creativity. I want to thank my sister for her cheerful nature and all the good times we had. I want to thank my brother David, and my "brothers" Sebi and Armin for I can always count on them, for alle the crazily amazing things we experienced, and for all those to come. Lastly, and most importantly, I want to thank Jana, for she is my love, my partner, and my best friend. She stood by my side during this project and hopefully always will.