

Chunking in tonal contexts: Information compression during serial recall of visually presented musical notes

Psychology of Music

1–18

© The Author(s) 2021



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/03057356211013396

journals.sagepub.com/home/pomLucas Lörch 

Abstract

Chunking is defined as information compression by means of encoding meaningful units. To advance the understanding of chunking in musical memory, the present study tested characteristics of melodic sequences that might enable a parsimonious memory representation, namely, the presence of a clear tonal context and of melodic cells with clear labels. Musical note symbols, which formed either triads (Experiment 1) or cadences (Experiment 2), were presented visually and sequentially to musically experienced participants for immediate serial recall. The melodic sequences were varied on the within-participant factors list length (long vs. short list) and tonal structure (chunking-supportive vs. chunking-obstructive). Chunking-supportive sequences contained tones from a single diatonic key that formed melodic cells with a clear label, such as “C major triad”. Transitional errors showed that participants grouped notes into melodic cells. Mixed logistic regression modeling revealed that recall was more accurate in chunking-supportive sequences and that this advantage was more pronounced for more experienced participants in the long list length condition of Experiment 2. The findings suggest that a clear tonal context and melodic cells with clear labels benefit chunking in melodic processing, but that the subtleties of the process are additionally influenced by type, size, and number of melodic cells.

Keywords

chunking, serial recall, short-term memory, musical notes, musical memory

The most common method to investigate musical memory is the auditory presentation of musical stimuli in a recognition task (Boltz, 1991). Its usage provided numerous findings on the memory for musical sounds (reviewed in Snyder, 2016), tonal cognition (reviewed in Bigand & Poulin-Charronnat, 2016), and the processing of melodies (reviewed in Schmuckler, 2016). The memory for visually presented musical note symbols, on the contrary, has received surprisingly little

University of Mannheim, Mannheim, Germany

Corresponding author:

Lucas Lörch, Fakultät für Sozialwissenschaften, University of Mannheim, A5, 6, B209, 68161 Mannheim, Germany.
Email: lloerch@mail.uni-mannheim.de

attention in the literature. One reason for this might be that the auditory presentation of music appears to be more natural and ecologically valid than the presentation of musical notes. However, a downside of using musical sound as an experimental stimulus is its multidimensional and complex nature, that might hinder the isolated analysis of specific cognitive processes.

One theoretic notion that appears repeatedly in the research on musical memory is the concept of *chunking*. The prevalent notion of chunking in cognitive psychology is that it denotes a compression of information in memory as a result of an automatic detection of meaningful structure (Mathy & Feldman, 2012). In music cognition, though, chunking is often rather conceived as a process of perceptual grouping (Bregman, 1990; Dowling, 1973). In his review on memory for music, Snyder (2016) writes “the first step in a listener’s construction of the form of a piece of music is the segmentation or chunking of the musical surface” (p. 114).

Discussing the multiple meanings of the term “chunking”, Gobet et al. (2016) suggested that segmentation and information compression need to be distinguished. Segmentation denotes an organization of information during encoding. This might facilitate access to the encoded information during recall. As such, segmentation does not rely on prior knowledge. For example, a musician might segment a sequence of notes into phrases and then use these phrases as cues to access the notes during recall. Information compression, though, is rather associated with using long-term knowledge to encode information in a different, more parsimonious manner. For example, if a sequence of notes corresponds to an etude a musician recently practiced, she might encode this sequence as “Etude in C#”. During recall, she might use her long-term knowledge on this etude to retrieve the single notes. As chunking has mainly been understood as a process of segmentation in the context of musical memory, there is a huge body of research on ways in which musical sounds are segmented (Deutsch, 2013), but rather limited knowledge on the compression of musical information.

Based on these considerations, the main motivation of the present study is twofold. First, it aims to contribute to the limited knowledge on memory for visually presented musical note symbols. Second, it strives to develop a more sophisticated understanding of chunking in musical information processing that combines ideas and methods from cognitive psychology with those of tonal cognition and melodic processing.

Memory for visually presented musical notes

In past studies on the recall of visually presented musical notes, musical experience was positively associated with recall accuracy (Meinz & Salthouse, 1998). Moreover, Halpern and Bower (1982) and Kalakoski (2007) found that the presence of a well-formed musical structure interacted with musical experience in their influence on recall accuracy. While well-formed musical stimuli were recalled more accurately in general, the advantage was more pronounced for more experienced participants.

The present study used these findings as a starting point but aimed to improve the understanding of chunking by providing a more detailed understanding of what constitutes a well-formed musical structure. Halpern and Bower (1982) and Kalakoski (2007) did not describe the employed musical structures in a formal way. To create their stimuli, Halpern and Bower (1982) asked professional musicians to write “good” melodies; Kalakoski (2007) used children’s songs. Accordingly, well-formed musical structure in their studies resulted from the intuition of composers. While this intuition might be based on formal rules, these rules were not made explicit. Therefore, it is unclear which specific attributes made the musical stimuli well-structured. The present study provides one approach for an explicit definition of well-formed musical structure by creating melodic sequences based on formal rules.

Chunking in musical information processing

Chunking is generally defined as the compression of information during encoding, resulting from the automatic detection of known structures (Mathy & Feldman, 2012). Based on these detected structures, multiple pieces of information are encoded as single, meaningful units (Miller, 1956). This process reduces the load on working memory and hence enables superior recall (Thalman et al., 2018).

In her study on chunking during processing of musical sound, Deutsch (1980) presented melodic sequences with a hierarchical melodic structure and random sequences consisting of the same notes to musically experienced participants. The structure of the hierarchical stimuli was strictly rule-based: one rule defined a sequence of root notes on a hierarchically higher level; another rule defined how these root notes were extended to form melodic cells on a hierarchically lower level. After participants listened to the melodic sequences, they had to recall them by writing down musical notes. Results showed that recall was superior in hierarchical stimuli. Deutsch (1980) concluded that chunking in music processing is supported by the presence of a systematic tonal structure. Such a structure allows to detect underlying rules and utilize them to represent the pitch information in a more compressed form. If the tonal structure is random, on the contrary, there is no rule that can be detected and hence, information compression is less effective.

However, besides this general effect of systematic compared to random tonal structures, little is known about more specific features of melodic sequences that foster chunking. There are many kinds of systematic tonal structures and the present study is based on the notion that some are more beneficial for chunking than others. To approach this issue, I compared systematically ill-structured with systematically well-structured melodic sequences. Well-structured sequences corresponded to musicians' knowledge of music theory and to their musical experience. More specifically, they (1) resided within a clearly recognizable tonal context and (2) contained melodic cells with clear labels within this tonal context. I expected that these features would foster chunking.

The beneficial effect of the former feature has been empirically supported by the finding that the presence of a tonal context supported the recognition of musical sound (Cuddy et al., 1979; Dewar et al., 1977). Theoretically, this notion is based on the model of musical activation (Bharucha, 1987). The model describes how the initial tones of a melody selectively activate tones and chords of the keys they belong to (Bigand & Poulin-Charronnat, 2016). Accordingly, I argue that if the beginning of a melody activates a key in which the whole melody can be represented, relevant information can be accessed rapidly and information compression is facilitated.

The latter claim, that chunking is supported by the presence of melodic cells with a clear verbal label, has been empirically supported by the finding that the presence of major triads starting on the tonic of a given key supported the memory for musical sounds (Croonen, 1991). Discussing Miller's famous article on chunking, Shiffrin and Nosofsky (1994) claimed that "a chunk is a pronounceable label that may be cycled within short-term memory" (p. 360). If the meaning of such a label is known to the individual, it might be used to represent its single elements in a parsimonious manner and access them during retrieval. Hence, I claim that while there might be other factors that support chunking in musical memory, the compression of musical information is supported if a melody contains melodic cells with a clear label, such as "C major triad".

In this article, I present two experiments which were created to test this theoretic conception of chunking in melodic processing. Both employed a serial recall task with sequentially presented musical note symbols. The stimuli of Experiment 1 consisted of three-note arpeggiated

triads, those of Experiment 2 consisted of four-note cadences. Henceforth, I will call these stimuli *triadic sequences* and *cadential sequences*, respectively. In both experiments, the tonal structure of melodic sequences was manipulated in order to support or hinder chunking with respect to the theoretical considerations presented above. *Chunking-supportive* sequences could be represented in a clearly recognizable tonal context, as they contained notes from a single diatonic key, started on the tonic of this key, and adhered to the conventions of the usage of accidentals in that key. In addition, items of this type consisted of melodic cells with a clear label, such as “major triad” or “authentic cadence”. *Chunking-obstructive* sequences could not be represented in a clearly recognizable tonal context as they contained notes from multiple harmonic minor (rather than diatonic) keys, did not start on the tonic notes of these keys, and violated conventions about enharmonic spelling in the respective keys. Finally, melodic cells in these sequences did not have clear labels.

I hypothesized that the presence of chunking-supportive melodic structures and the musical experience of the participants would have main and interacting effects on the serial recall accuracy. In line with previous findings (Halpern & Bower, 1982; Kalakoski, 2007; Kauffman & Carlsen, 1989; Meinz & Salthouse, 1998), I expected musical experience to have a positive effect on serial recall accuracy; in line with the theoretical considerations presented above, I expected chunking-supportive sequences to be recalled more accurately than chunking-obstructive sequences. As chunking is dependent on musical knowledge resulting from musical experience, I expected to find an additional advantage in chunking-supportive sequences for more experienced participants.

Method

Procedure

As both experiments are based on the same theoretic idea, use similar methods, and test analogous hypotheses, they will be presented not one after another, but in parallel. Both experiments were administered as online studies via the platform Unipark (<https://www.unipark.com>). Their procedure was identical. After participants gave informed consent, instructions were presented. Participants were informed that they would have to memorize the pitches of sequentially and visually presented quarter note symbols. They were also informed that the accidentals in the stimuli only affected the current note. The presence of systematic tonal structures was not mentioned in the instructions. Then, two training trials with random tones and the 32 randomly ordered trials of the recall task followed. One trial of the recall task consisted of a fixation cross (2,000 ms) followed by a series of single note symbols on a single-line staff in treble clef. Each note was presented for 1,000 ms and in between notes, there was a 500 ms inter-stimulus interval. Notes disappeared after presentation and subsequent notes appeared at the exact same horizontal position. Staff and clef were visible the whole time. After the last note of a trial, a prompt appeared, asking participants to scroll down and give their answer. For each serial position, a drop-down menu with note names was provided in which participants selected the note name they deemed correct at the given position. In case participants recalled the note symbol but did not know the respective note name, an image above the drop-down menus was presented that depicted all relevant note symbols together with their names. As the drop-down menus for all notes of one trial were depicted on the same web page, participants could choose the order in which they indicated their answers (e.g., starting with the last note).

After the recall task, the German version of the Gold Musical Sophistication Index (Gold-MSI) was administered (Schaal et al., 2014). I used the global scale of this questionnaire which is

termed *general musical sophistication* as the indicator of musical experience in all analyses. Finally, demographic data were collected. The experiment was advertised via Email and in social media groups. The advertisement called for participants who were able to identify note symbols or to play notes on an instrument. Participants did not receive any compensation for taking part in the study. The experiment was conducted in full accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki) as well as with German data privacy laws.

Design

In both experiments, the design was defined by two within-participant factors, namely, *tonal structure* (chunking-supportive vs. chunking-obstructive) and *list length* (6-note vs. 9-note in Experiment 1; 4-note vs. 8-note in Experiment 2). The number of trials was not constant across conditions, but the short list length conditions employed a larger number of trials (ten trials) than the long list length conditions (six trials). The purpose of this was to avoid that participants would get frustrated by the difficult task and abort the experiment. The design of the two experiments with exemplary items, the number of trials and the mean number of accidentals in each condition can be found in Table 1. Both studies were preregistered on aspredicted.org. The preregistrations can be found under <https://aspredicted.org/rz6qt.pdf> and <https://aspredicted.org/xn2my.pdf>.









Material

The procedure to create the memoranda of each trial consisted of three steps: (1) choose the first note of each trial; (2) choose further root note(s) with respect to the initial note; and (3) complete the melodic cells with respect to each root note. These three steps and the rules they followed are depicted in Table 2. In both experiments, only notes between B3 and Ab5 in treble clef were used. As can be seen in Table 2, the first notes of the trials were parallel between experiments and between the chunking-supportive and chunking-obstructive condition. In the chunking-supportive condition, the initial note defined the key on which this melodic sequence was based.

The rules for the second step, which was the selection of root notes with respect to the initial note, varied between experiments and conditions. In order to create a clear tonal context in chunking-supportive items, the Scale Steps IV and V of the major key defined by the initial note were used as root notes. In the final step of the creation of melodic sequences, root notes were extended to create melodic cells. In the chunking-supportive condition, root notes were extended by major triads (Experiment 1) or authentic cadences (Experiment 2). In the chunking-obstructive condition, triads and cadences with intervals of 8 and 9 semitones to the root notes were created. This was done as these intervals are not part of the diatonic scale and as the resulting melodic cells do not have a clear, conventional label. In addition to the rules presented in Table 2, there was one further systematic difference between conditions, which referred to the use of accidentals. In the chunking-supportive condition, accidentals were in coherence with the conventions for the given key, while in the chunking-obstructive condition, these conventions were violated, assuming one of the possible interpretations in harmonic minor.

To finalize the stimuli, note images of the melodic sequences were created with the program Forte 7 Basic (<https://www.fortenotation.com/en/>) and then graphically altered with the program GIMP (<https://www.gimp.org/>). The indication of meter and bar lines were removed and the staff with a 20×60 pixels single note was positioned in the center of a 640×420 pixels white image. All single images of each trial were compiled into an animated GIF file which was used in the online experiment. All stimuli, the raw data, and the analysis code of the present experiment can be found in the Online Supplemental Material.

Table 1. Experimental Design.

Experiment	List length	Chunking-supportive	Chunking-obstructive
Experiment 1 - triads	6-note	 Number of trials: 10 Mean number of accidentals: 2.4	 Number of trials: 10 Mean number of accidentals: 2.5
	9-note	 Number of trials: 6 Mean number of accidentals: 3.83	 Number of trials: 6 Mean number of accidentals: 3.66
Experiment 2 - cadences	4-note	 Number of trials: 10 Mean number of accidentals: 1.6	 Number of trials: 10 Mean number of accidentals: 1.6
	8-note	 Number of trials: 6 Mean number of accidentals: 3.83	 Number of trials: 6 Mean number of accidentals: 3.33
Characteristics		Contain tones from a single diatonic key Start on tonic of that key Adhere to conventions of accidentals used Melodic cells have a clear label	Contain tones from multiple non-diatonic keys Do not start on tonic of these keys Do not adhere to conventions of accidentals used Melodic cells do not have a clear label

The cells contain one exemplary stimulus, as well as the number of trials and the mean number of accidentals in the respective condition. The depicted notes were presented sequentially at the same horizontal position.

Participants


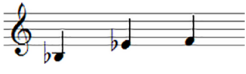
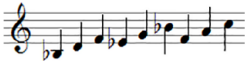
Table 3 depicts the characteristics of the samples. The mean scores on the Gold-MSI global scale were about 1 standard deviation above the mean of the norm sample (Gold-MSI norm sample: $M = 70.41$; $SD = 19.94$; Schaal et al., 2014). They corresponded to the 85th and 80th percentile of the norm sample in Experiment 1 and Experiment 2, respectively. Hence, although I addressed musically literate persons in general when advertising the study, both samples can be considered highly musically experienced.

Results

Exclusion of participants

For both experiments, four exclusion criteria were preregistered. According to these criteria, participants were excluded from further analyses (1) if they completed less than half of the recall task or of the Gold-MSI questionnaire; (2) if they completed the recall task with unrealistically short answering times (6-note: <15 s; 9-note: <25 s; 4-note: <12 s; 8-note: <22 s); (3) if their

Table 2. Creation of Melodic Sequences.

Step and exemplary item	Condition	Experiment 1 - triads	Experiment 2 - cadences
1. Choose the initial note for each trial 	Short list length	The 10 trials started on the notes between C4 and A4	
	Long list length	The 6 trials started on the notes between Bb3 and Eb4	
2. Choose further root note(s) with respect to the initial note 	Short list length/ chunking-supportive	+ 5 semitones	–
	Long list length/ chunking-supportive	+ 5 semitones + 7 semitones	+ 7 semitones
	Short list length/ chunking-obstructive	+ 1 semitones	–
	Long list length/ chunking-obstructive	+ 1 semitones + 2 semitones	+ 9 semitones
3. Complete the melodic cell(s) with respect to the root note(s) 	Chunking-supportive	+ 4 semitones + 7 semitones (major triad)	+ 5 semitones + 7 semitones + 0 semitones (authentic cadence)
	Chunking-obstructive	+ 8 semitones + 9 semitones	+ 8 semitones + 9 semitones + 0 semitones

The depicted exemplary item is a 9-note chunking-supportive triadic sequence.

Table 3. Characteristics of the Samples.

Measure	Experiment 1 - triads	Experiment 2 - cadences
Sample size	87	73
Age	$M = 31.53$; $SD = 13.44$; $min = 18$; $max = 74$	$M = 28.59$; $SD = 14.22$; $min = 18$; $max = 76$
Sex	60 female; 26 male	44 female; 27 male
Occupation	41 students (9 music, 31 other subjects) 46 employed (20 professional musicians, 26 other professions)	47 students (7 music students, 40 other subjects) 25 employed (3 professional musicians, 22 other professions)
Gold-MSI global scale	$M = 93.56$; $SD = 13.87$; $min = 58$; $max = 123$	$M = 88.6$; $SD = 15.13$; $min = 55$; $max = 114$

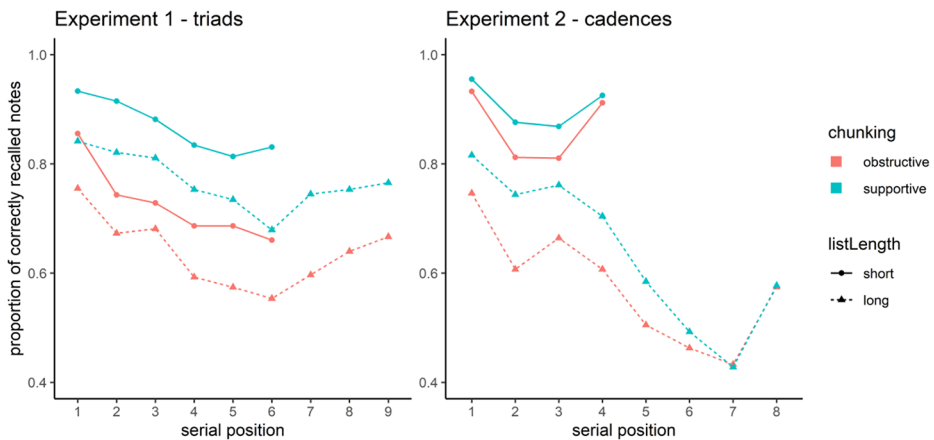
Gold-MSI: Gold Musical Sophistication Index. In Experiment 1, one participant did not indicate sex, age, and occupation. In Experiment 2, there were 2, 3, and 1 missing values on the variables sex, age, and occupation, respectively.

mean recall accuracy was below a critical value, which was defined as the median minus 2.5 times the median absolute deviation (Leys et al., 2013); or (4) if they gave the same answer to all items of the Gold-MSI questionnaire. Criteria 1 and 4 did not lead to the exclusion of any participants; criteria 2 and 3 led to the exclusion of eight and six participants in Experiment 1 and Experiment 2, respectively. The mean proportion of correctly recalled notes of participants excluded due to Criterion 2 was between 0.06 and 0.48. Hence, the fast answering times of the participants excluded due to this criterion probably were not a result of highly proficient performance, but of thoughtless answering.

Table 4. Descriptive Data of the Serial Recall Task.

Experiment 1 - triads				Experiment 2 - cadences			
List length	Measure	Chunking-supportive	Chunking-obstructive	List length	Measure	Chunking-supportive	Chunking-obstructive
6-note	PCN	0.87 (0.34)	0.73 (0.45)	4-note	PCN	0.91 (0.29)	0.87 (0.34)
	PCC	0.78 (0.41)	0.57 (0.49)		PCC	0.80 (0.40)	0.73 (0.45)
	PTE	0.07 (0.26)	0.16 (0.37)		PTE	0.06 (0.24)	0.09 (0.29)
9-note	PCN	0.77 (0.42)	0.64 (0.48)	8-note	PCN	0.64 (0.48)	0.57 (0.49)
	PCC	0.69 (0.46)	0.50 (0.50)		PCC	0.47 (0.50)	0.37 (0.48)
	PTE	0.08 (0.27)	0.15 (0.36)		PTE	0.18 (0.38)	0.23 (0.42)

Means and standard deviations of the proportion of correctly recalled notes (PCN), the proportion of correctly recalled melodic cells (PCC), and the proportion of transitional errors (PTE).

**Figure 1.** Proportion of Correctly Recalled Notes at Each Serial Position.

Recall accuracy

The analysis of the present experiment used three measures, namely (1) the proportion of correctly recalled notes (PCN), (2) the proportion of correctly recalled melodic cells (PCC), and (3) the proportion of transitional errors (PTE). A melodic cell was defined as being correctly recalled if all of its notes were recalled at their correct serial positions. The PTE describes the proportion of errors at a certain serial position given correct recall at the previous serial position (Johnson, 1966). Table 4 shows descriptive data for these three measures. The table shows that, descriptively, the recall of notes and melodic cells was more accurate and the PTE smaller in chunking-supportive sequences and in short list length conditions.

Figure 1 depicts serial position curves, that is, the proportion of correctly recalled notes at each serial position. The plots indicate the typical primacy and recency effects (Oberauer, 2003). Most interestingly, the shape of the curves in chunking-supportive and chunking-obstructive conditions were similar in both experiments. They were shifted vertically, but except for minor divergences in 6-note triadic sequences, their overall contour was basically identical.

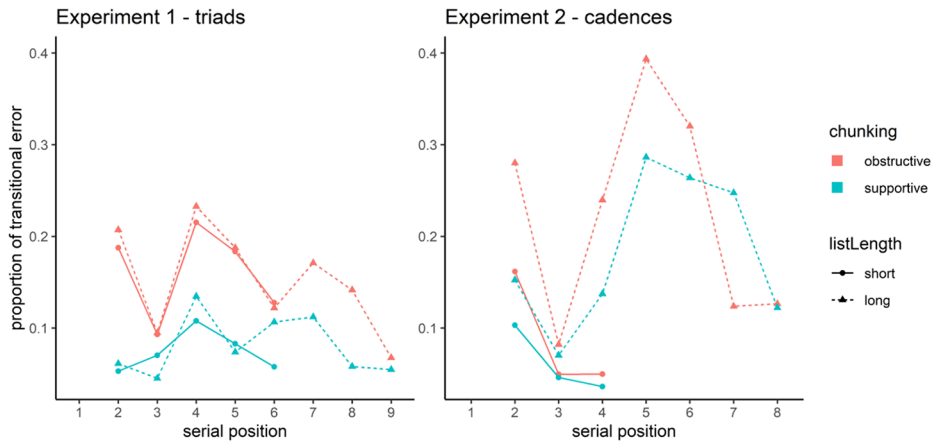


Figure 2. Proportion of Transitional Errors at Each Serial Position. In Experiment 1, Melodic Cells Started at Positions 1, 4, and 7; In Experiment 2, They Started at Positions 1 and 5.

This suggests that the cognitive processes involved in recall were similar between conditions, but that chunking-supportive items allowed a more parsimonious memory representation.

Group boundaries

To investigate how notes were grouped in memory, I examined the PTE in more detail. This measure indicated how dependent the recall of a note at a certain serial position was on the recall of its predecessor. Therefore, the PTE was not defined for the first serial position. Two aspects of this measure are of importance, namely, its level and its local extrema across serial positions. Concerning the level of the PTE, 0.5 denotes an important threshold. If the PTE is at 0.5, this indicates that the probability of an error at a certain serial position is at chance level given the previous item is recalled. In other words, the recall probabilities of the two items are independent of each other. If the PTE is above or below 0.5, this indicates a dependency between items: given the recall of the previous item, the proportion of error is more or less likely than chance level, respectively. In the context of chunking, a PTE below 0.5 is most crucial, as it indicates a chaining or grouping of items in memory. The second important aspect of the PTE is its local maximum or minimum across serial positions. Such an extremum at a certain serial position indicates that the respective item is especially independent or dependent of the recall of its predecessor. This indicates how items were grouped during encoding (Gilchrist, 2015).

Figure 2 depicts the PTE across serial positions. The plot indicates that the PTE did not exceed 0.5, that is, given the recall of a note, the recall of the following was more probable than chance. Apparently, the notes of the melodies were not represented in memory individually, but were chained to their predecessors. In addition, in chunking-supportive sequences, the PTE had its local maximum at the first note of melodic cells. This suggests that the recall of the first note of a melodic cell was less dependent on the recall of the previous note in chunking-supportive sequences, which can be interpreted as evidence for the encoding of melodic cells as units.

Last, the local minima in the chunking-obstructive conditions systematically occurred at the last note of semitone pairs across both experiments and list length conditions. Table 1 shows that semitone pairs were present in the chunking-obstructive condition of Experiment 1

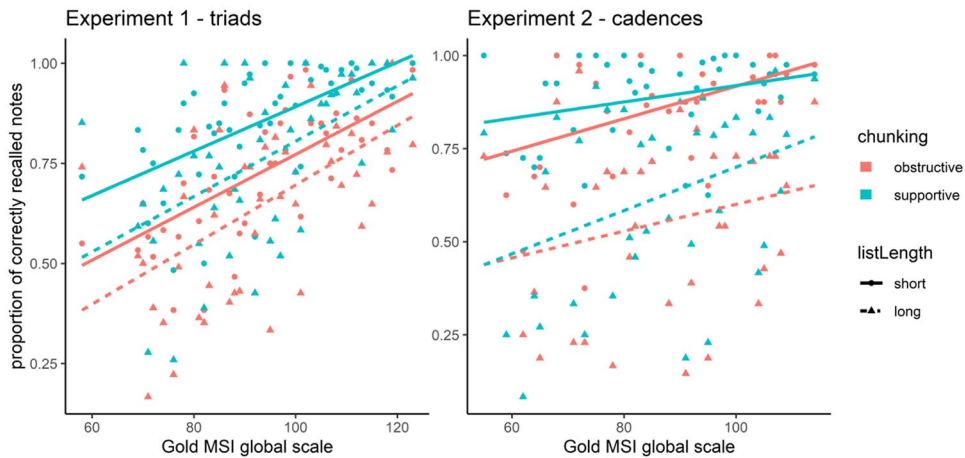


Figure 3. The Relation of Recall Accuracy and Gold-MSI Global Scale Score, Shown Separately for the Crossed Factors Tonal Structure and List Length.

at serial positions 2–3, 5–6, and 8–9 and in Experiment 2 at serial positions 2–3 and 6–7. In Figure 2, we see that in Experiment 1, the curve in the chunking-obstructive condition has its local minimum at positions 3, 6, and 9 and in Experiment 2 at positions 3 and 7. This means that the recall of the second note of a semitone pair was especially dependent on the recall of its predecessor, which can be interpreted as evidence that semitone pairs were encoded as units.

The interaction of musical experience with tonal structure

The main theoretical assumption of this study is that the recall advantage in chunking-supportive sequences is more pronounced for more experienced participants. Figure 3 depicts the relation of the Gold-MSI global scale score, which was the indicator of musical experience, and recall accuracy in chunking-supportive and chunking-obstructive sequences across both list length conditions. The plots show that there was an overall positive relationship of musical experience and recall accuracy. Unexpectedly, the plot suggests a three-way interaction between musical experience, tonal structure, and list length. The plots show that the pattern of interaction between musical experience and tonal structure varied across levels of the factor list length. Lastly, we see that the data points crowd at the top of the plot, which means that many participants solved the task with perfect or near-perfect accuracy. This pattern is more pronounced in Experiment 1. In Experiment 2, it rather seems to concern the short list length condition.

Mixed logistic regression modeling

In the mixed logistic regression analysis, the dependent measure was a variable indicating whether a note was recalled at the correct serial position or not (0/1). Predictors were the factors tonal structure (chunking-supportive vs. chunking-obstructive) and list length (short vs. long), the numerical variables serial position and Gold-MSI global scale score, and an interaction between Gold-MSI global scale score and tonal structure. By-participant random intercepts were used as random effects. The Gold-MSI score was scaled and centered. The serial position was included as an integer variable starting at zero for the first serial position.

Table 5. Model Comparison as Well as Estimates, Standard Errors (SE), and z Values for the Parameters of the Mixed Logistic Regression Models.

Parameter	Experiment 1 - triads			Experiment 2 - cadences		
	Estimate	SE	z value	Estimate	SE	z value
Comparison with null model	$\Delta AIC = 1,074$ $\chi^2(8) = 1,089.4$ $p < .001$			$\Delta AIC = 1,899$ $\chi^2(8) = 1,915.5$ $p < .001$		
Intercept	1.54***	0.12	12.54	2.76***	0.15	17.90
Long list length	-0.34***	0.05	-6.48	-1.60***	0.08	-19.14
Gold-MSI score	0.56***	0.12	4.64	0.60***	0.15	3.99
Chunking-supportive	1.22***	0.06	18.78	0.42***	0.10	4.17
Serial position	-0.11***	0.01	-12.98	-0.22***	0.01	-17.36
Long list length by Gold-MSI score	0.02	0.05	0.35	-0.27***	0.08	-3.51
Chunking-supportive by long list length	-0.38***	0.08	-4.51	-0.06	0.12	-0.50
Chunking-supportive by Gold-MSI score	0.32***	0.06	5.07	-0.12	0.09	-1.30
Chunking-supportive by long list length by Gold-MSI score	-0.17*	0.08	-2.05	0.32**	0.11	2.90

The Gold-MSI score was scaled and centered; the predictor serial position was an integer variable starting at zero. AIC: Akaike information criterion; SE: standard error; Gold-MSI: Gold Musical Sophistication Index.

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

As a first step in the analysis, I checked if a three-way interaction between Gold-MSI global scale score, tonal structure, and list length improved model fit. Compared to the models with a two-way interaction between Gold-MSI global scale score and tonal structure, the models with a three-way interaction had a superior fit to the data in both experiments (Experiment 1: $\Delta AIC = 15$, $\chi^2(3) = 21.287$, $p < .001$; Experiment 2: $\Delta AIC = 8$, $\chi^2(3) = 14.46$, $p < .01$). Accordingly, the models with three-way interaction were employed in subsequent analyses. As a next step, I checked the parameter estimates of these theoretically meaningful models and compared their model fit with the respective null models that included only by-participant random intercepts. The results are depicted in Table 5. In both experiments, the theoretically meaningful models showed a highly superior fit to the data than the null model. All predictors were significant, except for the two-way interaction terms in Experiment 2.

Parameter estimates in logistic regression models are commonly transformed from the log-odds scale to the probability scale in order to interpret them. However, as this transformation is non-linear, it might distort the interactional pattern indicated by the model (Wagenmakers et al., 2011). Accordingly, neither the parameter estimates of the model nor common post hoc tests provide valid information on the pattern of interaction on the probability scale.

Analyzing the interaction with bootstrapping

As a solution to this problem, I decided to use bootstrapping (Mooney & Duval, 2006). This allowed to obtain a distribution of parameter estimates on the probability scale and thus, to estimate how the interaction of Gold-MSI global scale score and tonal structure on the probability scale varied across levels of the factor list length. The results of this analysis, that is, the means and standard deviations of the slopes of the Gold-MSI score on the probability scale in the four conditions are depicted in Figure 4.

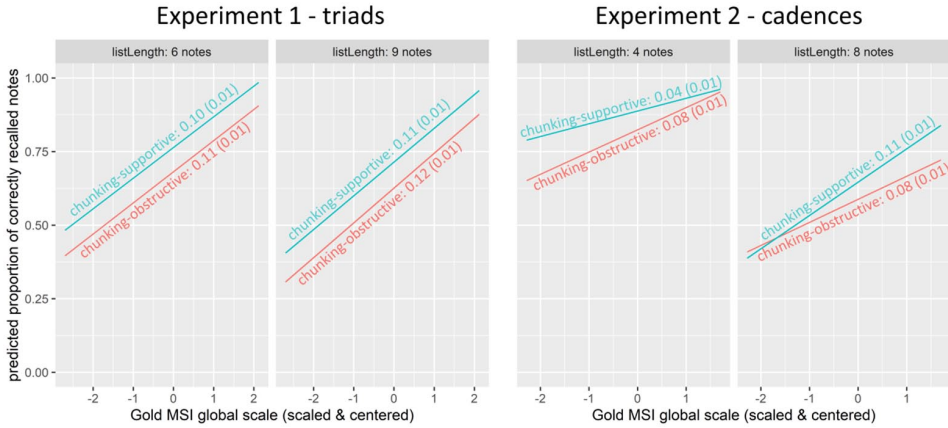


Figure 4. Predicted Pattern of Interaction on the Probability Scale.

Captions indicate the means and standard deviations of the slopes of the Gold-MSI global scale score on the probability scale obtained by bootstrapping.

In Experiment 1, the bootstrapping analysis indicated that on the probability scale, there was no three-way interaction, but merely main effects. The slopes of the Gold-MSI global scale score varied slightly, but the distributions of the slopes that were obtained by bootstrapping largely overlapped. This means that there was no significant variation in the slopes across conditions. As expected, more experienced participants were predicted to solve the task more accurately: for an increase of 1 standard deviation in the Gold-MSI global scale score, the model predicted an increase of 0.10 to 0.12 in the proportion of correctly recalled notes. The expected main effect of the factor tonal structure was also supported by the analysis. According to the model, the proportion of correctly recalled notes increased by 0.1 for chunking-supportive sequences compared to chunking-obstructive sequences. Last, the model indicated that 6-note sequences were recalled more accurately than 9-note sequences, with the proportion of correctly recalled notes differing by 0.05.

In Experiment 2, on the contrary, the significant three-way interaction on the log-odds scale did translate to the probability scale. The way in which tonal structure was predicted to interact with the Gold-MSI global scale score varied across levels of the factor list length. It becomes apparent that the hypothesized interaction was supported in the long list length condition. While the model indicated that 8-note chunking-supportive sequences were recalled more accurately overall, an increase of 1 standard deviation in the Gold-MSI global scale score was predicted to increase the proportion of correctly recalled notes additionally by 0.03. In the short list length condition, the predicted interaction was in the opposite direction than hypothesized. The advantage in 4-note chunking-supportive sequences was predicted to decrease by 0.04 with an increase of 1 standard deviation in the Gold-MSI global scale score.

Discussion

The main idea of a chunking mechanism in human information processing is that “material to be encoded is automatically analyzed for structure and redundancy, and compressed as much

as each pattern allows” (Mathy & Feldman, 2012, p. 357). In the processing of musical information, chunking has been found to depend on the presence of systematic tonal structures (Deutsch, 1980). In the present study, I tested two features of tonally structured melodic sequences which I expected to support chunking, namely, the presence of a clear tonal context and of melodic cells with a clear label. By testing these features, I aimed to deepen the understanding of what constitutes a “good melody” (Halpern & Bower, 1982) or a “musically well-formed note pattern” (Kalakoski, 2007).

In two experiments, participants had to recall the note names of musical note symbols that were presented visually and sequentially. In Experiment 1, the sequences of notes consisted of three-note triads; in Experiment 2, they consisted of four-note cadences. All note sequences had a systematic tonal structure, but it was varied if this structure supported or obstructed chunking. In chunking-supportive sequences, the notes formed major triad arpeggios (Experiment 1) or authentic cadences (Experiment 2) from a single diatonic key. Chunking-obstructive sequences consisted of notes from multiple non-diatonic keys which formed melodic cells with an unclear label. In addition, the length of melodic sequences was varied. The general musical sophistication scale of the Gold-MSI questionnaire, which was used as an indicator of participants’ level of musical experience, showed that the sample in both experiments was highly experienced.

I hypothesized that information compression would be superior in chunking-supportive sequences and hence, that they would be recalled more accurately than chunking-obstructive sequences. In addition, I expected information compression to be superior for more experienced participants. Therefore, I assumed them to recall melodic sequences more accurately and to benefit additionally from the presence of chunking-supportive tonal structures.

The influence of tonal structure on chunking processes

In summary, the data of the present experiment largely supported this hypothesis. Throughout both experiments, analyses showed a clear recall advantage for chunking-supportive sequences. In 8-note cadential sequences, the hypothesized interaction of musical experience and tonal structure was found. Moreover, the pattern of transitional errors supported the notion that participants grouped the sequences into melodic cells in chunking-supportive sequences. In these sequences, the PTE showed a local maximum at the first note of melodic cells. This means that the recall of the first note of a melodic cell was less dependent on the recall of the previous note, which can be interpreted as marking the beginning of a group in memory. Overall, there was a larger PTE in chunking-obstructive sequences. It appears that the recall of a note in these sequences was generally less dependent of the recall of the previous note, marking a less reliable grouping of notes during encoding.

These findings need to be interpreted on the background that *all* melodic sequences were systematically structured. Both chunking-supportive and chunking-obstructive sequences were created based on rules extending root notes with fixed intervals to form melodic cells. The fact that chunking-supportive sequences nevertheless were recalled more accurately and grouped into melodic cells more reliably suggests that indeed, certain systematic tonal structures are more beneficial for chunking than others. Combining the present findings with the ones by Deutsch (1980), the following general claim on chunking in music processing can be made: A melodic sequence can be chunked more efficiently if it has a systematic tonal structure than if it has a random tonal structure. Given a melody has a systematic tonal structure, it can be chunked more efficiently if it resides in a clear tonal context and contains melodic cells with

clear labels. Future studies might verify this claim by using random, systematically ill-structured and systematically well-structured melodic sequences in the same experiment.

The present results are in line with the findings by Halpern and Bower (1982) and Kalakoski (2007) that visually presented note sequences with a well-formed musical structure are recalled more accurately. In addition, the present work provided an explicit definition of what might characterize such a well-formed melodic sequence, namely that (1) it starts on the root note of a major scale and contains only notes from that scale and that (2) the notes form melodic cells with clear labels such as major triads. Based on this insight, future studies might develop and test further possible characteristics of well-formedness in musical stimuli.

In contrast to these expected findings, though, the regression models revealed that the interaction between musical experience and tonal structure varied across experiments and levels of the factor list length. In Experiment 1, musical experience did not interact with tonal structure. In short sequences of Experiment 2, the pattern was in the opposite direction than expected: the difference in recall accuracy between chunking-supportive and chunking-obstructive sequences decreased for more experienced participants.

The high level of musical experience of the samples provides a possible explanation for these unexpected effects. Both the recall test as well as the hypotheses were generated with reference to a sample containing novice and intermediate musicians. The sample, though, actually contained intermediates and professionals. This might have influenced the results in two ways. First, it might have had an impact on how participants encoded the note sequences. Second, it might have caused a statistical artifact, that is, a ceiling effect.

Concerning the former, it might be assumed that a pronounced level of musical experience enables to encode note sequences in terms of intervals. By encoding intervals instead of notes, highly experienced participants in the present study might have detected the systematic structure of chunking-obstructive sequences. As a result, they would have needed to recall merely the first note and then would have been able to derive the whole sequence. This might have mitigated the differences in recall accuracy between the conditions.

This logic provides a plausible explanation for the reversed interactional pattern in short cadential sequences. These sequences had a rather simple structure as they were only four notes long and as the last note was identical to the first note. Accordingly, in this condition, it seems likely that highly experienced participants recognized the systematic structure of both chunking-supportive and chunking-obstructive sequences, recalling them equally accurate. In addition, the encoding of melodies in terms of intervals would account for the finding that participants seem to have built groups of semitones in chunking-obstructive sequences as indicated by the transitional error plots.

Moreover, it might be argued that the high level of musical experience enabled most participants to detect the presence of major triads and authentic cadences. The main idea of chunking is that a certain threshold of experience is needed to recognize meaningful structures and use them for encoding. If most of the participants of a sample are above this threshold, the advantage in well-structured stimuli does not vary with experience and no interaction is found. This might be the reason for the missing interaction in triadic sequences. The transitional error profiles in Figure 2 imply that the encoding of triads in chunking-supportive sequences was prevalent across the whole sample.

Using questionnaires, structured interviews, or think-aloud protocols to collect data on participants' self-reported encoding strategies might have helped to interpret the recall data in the present experiment. Future studies might collect such data in recall tasks to check if participants recognized the systematic structure of control stimuli and if they detected the presence of melodic cells and used them as an aid to memory.

Lastly, both the missing as well as the reversed interaction could be a statistical artifact resulting from a ceiling effect. Ceiling effects lead to the loss of information on the true difference between data points and thereby cause the intercept of regression lines to increase and the slope to decrease (McBee, 2010). I expected a pattern of interaction in which the top regression line has a similar intercept but a steeper slope than the bottom regression line. If the intercept of the top line in such a pattern would increase and the slope would decrease due to a ceiling effect, this would lead to the lines being parallel and eventually, to a reversal of the interactional pattern.

Figure 3 supports the notion that the pattern of interaction was influenced by a ceiling effect. The data points in both plots crowd at the top, that is, multiple participants reached perfect or near-perfect recall in all sequences of one condition. In Experiment 1, we see that perfect or near-perfect recall rather occurred for highly experienced participants in chunking-supportive sequences. Thus, information on the true difference between chunking-supportive and chunking-obstructive sequences was lost for highly experienced participants. This might have caused the regression lines of the chunking-supportive conditions to have a larger intercept and a flatter slope and hence to be parallel to the regression lines in the chunking-obstructive conditions.

In Experiment 2, perfect recall apparently occurred in short sequences and across a wider range of Gold-MSI scores. Therefore, information was lost on the difference between 4-note chunking-supportive and chunking-obstructive sequences for participants of various experience levels. However, chunking-supportive sequences seem to have been affected to a larger extent than chunking-obstructive ones. Thus, it can be concluded that the regression lines of both short chunking-supportive and short chunking-obstructive sequences were affected by the ceiling effect but that the impact was more extreme for chunking-supportive sequences causing the interactional pattern to reverse.

To check if this explanation is valid, future studies might recruit a sample that contains musicians of lower musical experience or might increase the difficulty of the task by introducing longer sequences or shorter presentation times. Both means should lead to a positive interaction between musical experience and well-formed musical structure. As a control condition, researchers might even recruit a group of non-musicians, that is, participants who cannot read notes at all. As these participants would rely on encoding notes in a visual rather than musical manner, comparing their recall to the one of musicians would allow to ensure that musicians actually encode the notes as musical information.

Ceiling effects generally occur if the difficulty of a task is too low, given a certain level of experience of the participants. One factor that influences task difficulty is list length. It has long been known that recall accuracy decreases for longer lists (Strong, 1912). In the present study, though, a ceiling effect seems to have occurred in 9-note triadic sequences, but does not seem to have occurred in 8-note cadential sequences. Hence, we need to consider further factors that might have influenced difficulty in the present experiment. One of these potential factors is chunking. If the features of a task enable highly effective chunking processes, recall accuracy increases; the task becomes easier. Therefore, I deem it crucial to discuss in greater detail how specific features of the two experiments, that is, the specific combinations of type, size, and number of melodic cells, might have influenced chunking processes.

The detection of melodic cells

In planning the present study, I assumed that the presence of chunking-supportive tonal structures would be the only feature of the sequences that influences chunking processes. Triads,

cadences, and different list lengths were introduced merely to increase the validity of the findings. I sought to prove my theoretic idea in various contexts. However, chunking processes apparently were influenced in a complex manner by the specific combinations of triads, cadences, and list lengths.

The detection of melodic cells with clear labels was crucial for chunking in the present experiments, as it enabled the representation of multiple notes with a single label. The ease of detecting a melodic cell, though, might depend on its type. Major triads, for example, might appear more frequently throughout a musical piece, as authentic cadences commonly are used only at the end of a phrase. Therefore, major triads might be more readily available in memory and hence easier to detect. This provides one explanation why there was a ceiling effect in 9-note triadic sequences, but not in 8-note cadential sequences.

Another factor that might influence the ease of detecting a melodic cell is its size. There is evidence that humans tend to build groups of three elements in memory (Allen & Crozier, 1992). This tendency might make it more difficult to detect melodic cells that consist of more than three notes. Last, the type and number of melodic cells might jointly influence how easy melodic cells can be detected. In other words, for some types of melodic cells, it might be especially difficult to detect their presence if they appear repeatedly. In the present study, several findings suggest that authentic cadences were especially difficult to detect in 8-note sequences. Both the PTE and the proportion of recalled notes suddenly changed at the fifth serial position, indicating a less accurate recall and a less reliable grouping in the second authentic cadence. As cadences are structures that are mainly used to end a melodic phrase, a sequence of multiple cadences directly following each other might be rather rare in Western tonal music. In addition, contour changes were not coherent with the beginning of melodic cells in 8-note cadential sequences. Both aspects might have contributed to the difficulty of detecting authentic cadences in 8-note sequences. Unfortunately, the size, number, and type of melodic cells were confounded in the present experiments and the samples differed in their characteristics. Hence, we cannot know how these factors individually affected chunking processes. However, my considerations might provide valuable impulses for future studies.

Limitations and conclusion

One aspect that limits the scope of the present study is that it was administered as an online experiment. This entails the danger that participants were not concentrated or that they did not follow experimental instructions. For example, participants might have cheated by writing down memoranda during presentation. However, the present study used a within-participants design. Chunking-supportive as well as chunking-obstructive sequences were presented in random order. Cheating or being distracted would have had the same effect on both types of sequences, reducing the difference between them. The fact that there was a clear recall advantage in chunking-supportive sequences provides a strong argument that participants completed the task as instructed.

In summary, the main effect of tonal structure in Experiment 1, the hypothesized interaction in the long list condition of Experiment 2, and the pattern of transitional errors support the notion that chunking in melodic processing is fostered by the presence of a clear tonal context and of melodic cells with clear labels. In addition, the ease of detecting melodic cells might be jointly influenced by the type, size, and number of melodic cells. Future studies can build upon this work by disentangling how these factors influence information compression individually.

Chunking is a cognitive process that depends on the specifics of the domain in which it occurs. Thus, it is most crucial to conceive it within the regularities and constraints of that domain. Only by revealing how these regularities and constraints are utilized by the memory system to compress information, we can develop a nuanced understanding of the chunking phenomenon.

Authors' note

Parts of this work were presented at the conference of the German Society for Music Psychology (DGM) in Eichstätt, Germany, September 2019, and at the Conference of Experimental Psychologists (TeaP) in London, United Kingdom, April 2019. The studies were preregistered on aspredicted.org. The preregistrations can be found under <https://aspredicted.org/rz6qt.pdf> and <https://aspredicted.org/xn2my.pdf>.

Acknowledgements

I want to thank Stefan Münzer, Erkki Huovinen, and the participants of my seminar “Experimentalpraktikum 2018” for their support of this research.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the University of Mannheim's Graduate School of Economic and Social Sciences.

ORCID iD

Lucas Lörch  <https://orcid.org/0000-0002-5665-3313>

Supplemental material

Supplemental material for this article is available online and comprises stimuli, datasets and analysis code.

References

- Allen, P. A., & Crozier, L. C. (1992). Age and ideal chunk size. *Journal of Gerontology*, *47*(1), 47–51. <https://doi.org/10.1093/geronj/47.1.p47>
- Bharucha, J. J. (1987). Music cognition and perceptual facilitation: A connectionist framework. *Music Perception: An Interdisciplinary Journal*, *5*(1), 1–30.
- Bigand, E., & Poulin-Charronnat, B. (2016). Tonal cognition. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford library of psychology: The Oxford handbook of music psychology* (2nd ed., pp. 59–71). Oxford University Press.
- Boltz, M. (1991). Some structural determinants of melody recall. *Memory & Cognition*, *19*(3), 239–251. <https://doi.org/10.3758/BF03211148>
- Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. MIT Press.
- 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>
- Cuddy, L. L., Cohen, A. J., & Miller, J. (1979). Melody recognition: The experimental application of musical rules. *Canadian Journal of Psychology*, *33*(3), 148–157. <https://doi.org/10.1037/h0081713>
- Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389.
- Deutsch, D. (2013). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 183–248). Academic Press. <https://doi.org/10.1016/B978-0-12-381460-9.00006-7>
- Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(1), 60–67.
- Dowling, W. J. (1973). Rhythmic groups and subjective chunks in memory for melodies. *Perception & Psychophysics*, *14*(1), 37–40. <https://doi.org/10.3758/BF03198614>
- 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>
- 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>

- 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. <https://doi.org/10.2307/1422658>
- Johnson, N. F. (1966). On the relationship between sentence structure and the latency in generating the sentence. *Journal of Verbal Learning and Verbal Behavior*, 5(4), 375–380. [https://doi.org/10.1016/S0022-5371\(66\)80047-6](https://doi.org/10.1016/S0022-5371(66)80047-6)
- 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>
- Kauffman, W. H., & Carlsen, J. C. (1989). Memory for intact musical works: The importance of musical expertise and retention interval. *Psychomusicology: A Journal of Research in Music Cognition*, 8(1), 3–20. <https://doi.org/10.1037/h0094235>
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <https://doi.org/10.1016/j.jesp.2013.03.013>
- 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>
- McBee, M. (2010). Modeling outcomes with floor or ceiling effects: An introduction to the Tobit model. *Gifted Child Quarterly*, 54(4), 314–320. <https://doi.org/10.1177/0016986210379095>
- 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*, 101(2), 343–352.
- Mooney, C. Z., & Duval, R. D. (2006). *Bootstrapping: A nonparametric approach to statistical inference: Quantitative applications in the social sciences (Vol. 95)*. SAGE.
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49(4), 469–483. [https://doi.org/10.1016/S0749-596X\(03\)00080-9](https://doi.org/10.1016/S0749-596X(03)00080-9)
- Schaal, N. K., Bauer, A.-K. R., & Müllensiefen, D. (2014). Der Gold-MSI: Replikation und Validierung eines Fragebogeninstrumentes zur Messung Musikalischer Erfahrung anhand einer deutschen Stichprobe [The Gold-MSI: Replication and validation of a questionnaire to measure musical sophistication using a German sample]. *Musicae Scientiae*, 18(4), 423–447. <https://doi.org/10.1177/1029864914541851>
- Schmuckler, M. A. (2016). Components of melodic processing. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford library of psychology: The Oxford handbook of music psychology* (2nd ed., pp. 94–106). Oxford University Press.
- Shiffrin, R. M., & Nosofsky, R. M. (1994). Seven plus or minus two: A commentary on capacity limitations. *Psychological Review*, 101(2), 357–361. <https://doi.org/10.1037/0033-295X.101.2.357>
- Snyder, B. (2016). Memory for music. In S. Hallam, I. Cross, & M. Thaut (Eds.), *Oxford library of psychology: The Oxford handbook of music psychology* (2nd ed., pp. 107–117). Oxford University Press.
- Strong, E. K., Jr. (1912). The effect of length of series upon recognition memory. *Psychological Review*, 19(6), 447–462. <https://doi.org/10.1037/h0069812>
- Thalman, M., Souza, A. S., & Oberauer, K. (2018). How does chunking help working memory? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 37–55. <https://doi.org/10.1037/xlm0000578>
- Wagenmakers, E.-J., Kryptos, A.-M., Criss, A. H., & Iverson, G. (2011). On the interpretation of removable interactions: A survey of the field 33 years after Loftus. *Memory & Cognition*, 40(2), 145–160. <https://doi.org/10.3758/s13421-011-0158-0>