QJEP

Temporal grouping effects in verbal and musical short-term memory: Is serial order representation domain-general?

Simon Gorin[1,2] (iD)

Abstract

The question of the domain-general versus domain-specific nature of the serial order mechanisms involved in short-term memory is currently under debate. The present study aimed at addressing this question through the study of temporal grouping effects in short-term memory tasks with musical material, a domain which has received little interest so far. The goal was to determine whether positional coding—currently the best account of grouping effect in verbal short-term memory—represents a viable mechanism to explain grouping effects in the musical domain. In a first experiment, non-musicians performed serial reconstruction of 6-tone sequences, where half of the sequences was grouped by groups of three items and the other half presented at a regular pace. The overall data pattern suggests that temporal grouping exerts on tone sequences reconstruction the same effects as in the verbal domain, except for ordering errors which were not characterised by the typical increase of interpositions. This pattern has been replicated in two additional experiments with verbal material, using the same grouping structure as in the musical experiment. The findings support that verbal and musical short-term memory domains are characterised by similar temporal grouping effects for the recall of 6-item lists grouped by three, but it also suggests the existence of boundary condition to observe an increase in interposition errors predicted by positional theories.

Daily life activities such as remembering a phone number, having a discussion, or listening attentively to a piece of music all require the processing of serially organised information that unfolds over time and draws on short-term memory (STM) resources. The question of whether the mechanisms contributing to the maintenance of serially organised memoranda are domain-general or domain-specific is currently under debate in the STM literature (Hurlstone et al., 2014; Jones et al., 1995; Logie et al., 2016; Majerus, 2013; Soemer & Saito, 2016; Vandierendonck, 2016). Previous research comparing serial order STM for verbal and visuospatial items supported the view that the representation of serial order in STM is supported by domain-general mechanisms (for a review, see Hurlstone et al., 2014). However, the extent to which the domain-generality hypothesis applies to STM for music remains

unanswered. Given its inherent rhythmic and sequential structure, music represents an appropriate candidate to further our understanding of the ordering mechanisms involved in STM, as well as to address the question of the domain-generality of these mechanisms.

Contrary to the verbal domain, there are only a few models of musical STM (see Berz, 1995; Ockelford, 2007), none of which provide a comprehensive account of

[1]Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, Genève, Switzerland
[2]Faculty of Psychology, UniDistance Suisse, Brigue, Switzerland

Corresponding author:
Simon Gorin, Faculté de Psychologie et des Sciences de l'Éducation, Université de Genève, 1205 Genève, Switzerland.
Email: gorinsimon@gmail.com

the processes responsible for representing serial order of musical sequences. For instance, Ockelford (2007) suggested that serial order is coded in musical working memory through the action of a tagging mechanism in which each item serves as a retrieving cue for the next item (for more details regarding the notion of tagging, see Kieras et al., 1999). However, to the best of our knowledge, there is no direct empirical evidence that such a tagging mechanism plays a role in the representation of musical order in STM. Moreover, the tagging notion relies on a chaining account of serial order representation that has been challenged by a recent study on serial order STM for music (see Gorin et al., 2018a).

Several serial order effects considered as benchmark phenomena in the verbal STM domain (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008) have also been observed in a recent series of musical STM experiments (Gorin et al., 2016, 2018a, 2018b). The authors interpreted these results as evidence for the existence of domain-general processes to represent serial order information in the musical domain. This interpretation is also in line with the notion that verbal and musical STM systems involve common sequential processes even though they rely on different representational stores (Williamson et al., 2010). Thus, these results suggest that basic ordering principles are at work in the two domains. In addition, they justify the use of verbal order theories as a framework for exploring the nature of ordering mechanisms in musical STM and assessing the generality of these mechanisms.

The best account of benchmark order phenomenon in the verbal domain comes from models relying on positional codes to represent serial order information (see, for example, Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998; Lewandowsky & Farrell, 2008). In positional models, serial order is represented by associations between items and independent markers representing positions. The main strength of this class of models is its ability to account for temporal grouping effects. Temporal grouping is characterised by the insertion of additional pauses between some items during sequence presentation, inducing the perception of temporally distinct sub-groups of items. With verbal material, such manipulations lead to the well-replicated phenomena that constraint serial order models of STM (see Frankish, 1985, 1989; Hartley et al., 2016; Henson, 1996; Hitch et al., 1996; Maybery et al., 2002; Ng & Maybery, 2002, 2005; Ryan, 1969a, 1969b). For grouped sequences, a recall advantage as well as a multiply-bowed shape serial position curve are usually observed. An increase in the proportion of interposition errors, or between-group displacements of items that keep their initial within-group serial position, is also characteristic of the recall of grouped sequences. For instance, in a 6-item sequence composed of two groups of three items, an interposition error would be to recall the item from Position 2 (i.e., Position 2 in the first group) at Position 5 (Position 2 in the second group).

The study of temporal grouping effects is of particular interest to help determine the precise nature of serial order representation in STM. For example, models relying on ordinal codes such as activation gradients to represent serial order (see, for example, Farrell & Lewandowsky, 2002; Page & Norris, 1998) can accommodate the main effects induced by temporal grouping manipulations, the recall advantage, and the scalloped serial position curve. However, to account for a wider range of effects induced by temporal grouping (i.e., an increase in interposition errors in addition to the recall advantage and the scalloped appearance of the serial position curve), it is necessary to assume the existence of positional codes. Positional models accommodate the increase in interposition errors by representing serial order in a hierarchical manner. Items are associated with positional markers representing within-group positions, as well as markers representing the position of the groups/items in the sequence (see Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998). The hierarchical representation of serial order makes the items in grouped sequences more distinctive than in ungrouped ones, accounting for the recall advantage and the multiply-bowed shape of the serial position curve. Moreover, this hierarchical representation of serial order increases the similarity between items in different groups that share the same within-group position, thus accounting for the increase in interposition errors observed in grouped sequences (see Figure 1 markers for a graphical example).

In the musical domain, a great deal of work has been devoted to the study of the psychophysical and musical components influencing how adult listeners process and maintain musical information in STM (for a review, see Deutsch, 2013a, 2013b). However, little is known about the cognitive mechanisms involved in the short-term maintenance of musical information, and particularly those required to represent and maintain the order of a series of tones. In non-musicians, serial order reconstruction of verbal and musical sequences is characterised by similar serial order effects, suggesting that verbal ordering principles could be extended to the musical domain (Gorin et al., 2018a). In another study using a serial recognition task, researchers observed temporal grouping effects that are comparable to those usually observed in verbal STM tasks, suggesting that the positional markers described in verbal STM models of serial order could play a role in STM for music (Gorin et al., 2016). More precisely, the authors showed that in non-musicians, the rate of correct serial recognition for matching probes is higher for grouped versus ungrouped sequences, and that recognition as a function of position adopted a shape reflecting the grouping structure used in the experiment, replicating previous results obtained with musicians (Deutsch, 1980).
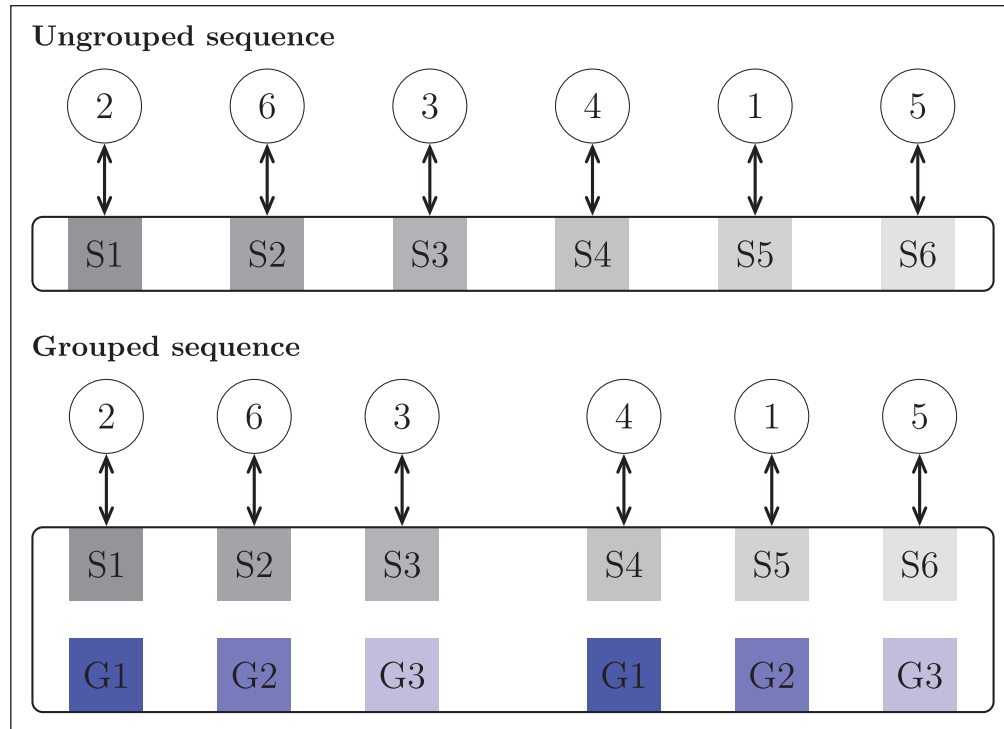
Figure 1. Schematic representation of positional markers. Top: ungrouped sequence where six digit items are associated to positional markers representing positions in the sequence (grey shades). Bottom: grouped sequences where items are associated to markers representing position of items in the sequence (grey shades) and within the groups (blue shades). Darkest and lightest shades represent start and end of the sequence, respectively. As one can see, the similarity between items at Positions 1 and 4 is low in ungrouped sequences. But the similarity between these items increases in grouped sequences because of the additional marker representing within-group positions.

However, the conclusions drawn in Gorin et al. (2018b) are limited as the assessment of temporal grouping on interposition errors was not possible due to the use of a recognition procedure. At the same time, there is evidence for the existence of interposition-like errors in serial order production tasks requiring experts to retrieve and play short musical excerpts on the piano from memory (Mathias et al., 2015). In comparison to shorter musical excerpts, for which the similarity between elements sharing the same metrical accent (strong or weak) in the sequence is reduced, longer musical excerpts are characterised by increased long-distance transpositions between positions with the same metrical accent. Interestingly, this phenomena can be accounted for by a model of musical sequence production assuming that to-be-produced musical events are represented hierarchically (see Mathias et al., 2015; Palmer & Pfordresher, 2003; Pfordresher et al., 2007). This model represents musical events according to their serial position and their metrical status strength, which is similar to the hierarchical coding of serial order proposed in positional models of verbal STM described above (see Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998).

As mentioned earlier, a growing body of evidence shows that benchmark serial phenomena characterising verbal STM are also observed in visuospatial STM tasks (for a review, see Hurlstone et al., 2014), and to some extent to the musical domain as well (Gorin et al., 2018b). This evidence supports the first account that the processing of serial order information is supported by processes shared across domains. At the same time, some authors consider that the presence of similar ordering phenomena across domains is also compatible with the existence of domain-specific mechanisms, but with functional similarities (see, for example, Logie et al., 2016; Saito et al., 2008). Indeed, observing the same serial order phenomena across STM domains is compatible with both a single domain-general mechanism and with domain-specific mechanisms coding serial order in a similar manner, and only this second account assumes that the existence of functionally similar domain-specific mechanisms can account for both differences and similarities across domains (Logie et al., 2016). Another account would be that serial order mechanisms are shared across modalities (e.g., auditory or visual) but not specific domains (e.g., verbal, visual and musical). In other words, we could envisage that both verbal and musical materials similarly draw on auditory STM resources, the latter being underpinned by auditory domain-general processes responsible for coding serial order information for both types of

material. This is in line with a recent proposal from Hartley et al. (2016) suggesting that cross-domain sequential principles are responsible for processing order information but function in parallel with domain-specific mechanisms responsible for perceptual input. They proposed a stimulus-driven mechanism responsible for processing and encoding order information in auditory–verbal sequences based on the activity of neuronal oscillators tracking amplitude variations of the speech envelope at different timescales. Interestingly, it has been suggested that the encoding of rhythmic features in both speech and music could be governed by a similar stimulus-driven oscillatory mechanism (see, for example, Musacchia et al., 2014). Thus, considering the evidence for domain-specificity in processing musical information (Peretz & Coltheart, 2003; Zatorre et al., 2002), a more parsimonious account would be that domain-specific features interact with domain-general ordering mechanisms (see Majerus, 2013).

To summarise, the present study aimed at investigating the effects of temporal grouping on immediate serial reconstruction of tone sequences. Through the comparison of the temporal grouping effects observed for tone sequences with those reported in the verbal STM literature, our goal was to (1) improve our understanding of the mechanisms underlying the representation of serial order in musical STM and (2) address the question of the domain-generality of serial order processes in STM. We conducted a first preregistered experiment comparing the forward reconstruction of serial order information between ungrouped 6-tone sequences and the same sequences grouped in two groups of three items.[1] Based on the results obtained in that first experiment, a non-preregistered follow-up online experiment requiring the serial recall of 6-letter grouped and ungrouped sequences has been conducted to allow a direct comparison with the data obtained in Experiment 1. Due to the presence of ceiling effect that limited the comparison of temporal grouping effects in the musical (Experiment 1) and verbal (Experiment 2) domains, another non-preregistered online experiment was conducted to account for ceiling effect. Overall, these experiments support the close similarity between the temporal grouping effects observed in the verbal and musical domains.

Experiment 1: forward reconstruction of musical order

Method

Sampling plan. There is currently a trend in the field of psychological sciences favouring the use of Bayesian statistical techniques to design experiments and make statistical inferences. Bayesian statistics provide several advantages (for a review, see Dienes, 2016; Wagenmakers et al., 2018). For instance, Bayesian statistical analyses allow the monitoring of statistical evidence during data collection,

are not influenced by the intention with which data are collected, and are not sensitive to optional stopping rules (Berger & Berry, 1988; Rouder, 2014). With these considerations in mind, we used the following sampling plan for Experiment 1 (for a similar rationale in determining sampling plan, see Wagenmakers et al., 2015). We first recruited 20 participants and conducted the planned analyses. If for these analyses (see the "Analysis plan" section for more details), we obtained strong level of statistical evidence for either an alternative (H_1) or the null (H_0) hypothesis with a Bayes factor (BF) of 10 or more, data collection would be stopped. If that criterion was not met for at least one of our planned analyses, we would recruit more participants while monitoring BF values. In other words, we ran the same analyses after each batch of five participants and continued until we reached strong statistical evidence for all the planned analyses (H_0 or H_1). However, due to resource limitations, we planned to stop data collection after the recruitment of 50 participants, even though we did not meet the criterion of statistical evidence for all the planned analyses.

Participants. The experiment was approved by the ethics committee of the Faculty of Psychology and Sciences of Education of the University of Geneva. Fifty-eight first-year psychology students from the University of Geneva took part in Experiment 1 in exchange for partial course credit. The final sample was composed of 50 participants (45 females; age n years: $M=21.78$, $SD=1.95$; education level in years: $M=13.00$, $SD=1.12$; musical theory learning in years: $M=0.35$, $SD=0.85$; musical practice in years: $M=0.69$, $SD=1.04$) after the exclusion of eight participants who did not meet the inclusion criteria (see the demographic data file on the OSF repository associated to this manuscript for more details).

Inclusion and exclusion criteria. As we were interested in musical STM for serial order processing in participants with no musical expertise, participants must have had no more than 3 years of experience in studying music theory or practicing a music instrument (including singing) at the time of the experiment. We excluded participants with neurological or speech disorders (e.g., dyslexia) from the sample. Finally, we excluded the data from any participants with performance equal to or lower than the .17 chance-level in at least one of the experimental conditions from the analysis. To adhere to the sampling plan, excluded participants were replaced by recruiting other participants.

Stimuli. The stimuli consisted of 60, 6-tone sequences. To reduce the possibility that using a limited set of six tones could increase proactive interference, we used a set of 14 different tones consisting of all the diatonic steps of the C major scale (ranging from C4 to B5). The tones were pure sine waves generated with Audacity (Audacity Team,

2017) and saved as .wav files, each lasting for 500 ms with a rise and fall period of 10 ms. The tone sequences were generated using pseudo-random permutations following three rules adapted from previous studies on verbal STM for serial order (see, for example, Hartley et al., 2016):

1. No more than two consecutive tones that are also consecutive in the tone set (e.g., C4–E4–G4 or B4–D5–F5 was not legal);
2. No more than two consecutive intervals in the same direction (e.g., C4 [↗] E4 [↗] D5 [↘] G4 was permitted but not C4 [↗] E4 [↗] D5 [↗] F5);
3. No tone at the same serial position in successive trials.

As the tones used cover two octaves, we constrained interval sizes to a maximum of seven semitones to avoid the presence of unfamiliar large intervals. We also ensured that the sequences were highly related to a major scale. In other words, each sequence has a maximum key correlation of at least .70 with the tone distribution profile of at least one of the major scales. The maximum key correlation was determined using the Krumhansl & Schmuckler key-finding algorithm (Krumhansl, 1990).

To have matched stimuli between the two grouping conditions, we reused the 30 sequences from the ungrouped trials but played them in reverse serial order and presented them from last to first in the grouped trials. To prevent unwanted effects resulting from the use of a fixed set of tone sequences, a new set of pseudo-randomly created tone sequences was generated in advance for each participant. To ensure that each created sequence was used both in an ungrouped and a group trial, even-numbered participants had the ungrouped and grouped sequences corresponding to the grouped and ungrouped sequences, respectively, of the preceding odd-numbered participant in the experiment.

Experimental design. The experiment was based on a 2-factor within-participants design. The two types of sequences were presented in two different blocks with the ungrouped sequences always presented first. This was done to avoid that presenting the grouped sequences first could lead to the use of subjective grouping strategies for ungrouped trials (for a similar procedure, see Farrell & Lewandowsky, 2004; Hartley et al., 2016). For ungrouped trials, the tones were presented at a regular pace.

Procedure. The procedure consisted of the auditory presentation of 60 trials in total. Stimuli were played at a comfortable auditory level through headphones connected to a portable workstation. Each trial began with a countdown from 3 to 1 displayed at the centre of the computer screen at a pace of 500 ms. The tone sequence was played consecutively on a blank screen displayed for 500 ms. In ungrouped trials, the tones were presented with a regular interstimulus interval (ISI) of 150 ms. In grouped trials, the ISI was 75 ms for within-group items (Positions 1–2, 2–3, 4–5, and 5–6) and 450 ms for between items forming group boundaries (Positions 3–4). Immediately after the presentation of a sequence, a virtual keyboard was displayed on the screen and the participants used the touch screen to reconstruct the sequence. The participants were forced to reconstruct the sequences in forward serial order. To do this, they had to find and validate the tone corresponding to the first position, then proceed to the second position, and so on until reconstructing the whole sequence.

The virtual keyboard was used again to reconstruct the tone sequences (Figure 2). A layer of six white keys representing the six tones heard in the to-be-reconstructed sequence were displayed horizontally on the screen. The tones were organised in ascending order, from the lowest on the left to the highest on the right. Each time a key was pressed on the touch screen, the corresponding tone was played through the headphones. Touching a key activated the associated tone by changing the colour of the key to green (see panels 1, 5, 7, or 10 in Figure 2). Once the participant retrieved the tone for the current position and activated the key, they had to press the "validate" button to proceed to the next position (see panels 4, 6, 8, or 12 in Figure 2). After a tone has been assigned to a position, the corresponding key changed to grey to indicate that the tone could not be used anymore and the auditory feedback for that key was turned off. It was possible to change the "active" tone before validating a position (see panels 10–12 in Figure 2) but not once the position was validated. If for any position the participant did not remember the corresponding tone or did not want to guess, it was possible to answer "I don't know" by selecting the "?" button before validating the position (see panel 11 in Figure 2). Finally, at any time during the reconstruction process, participants had the opportunity to hear the reconstructed sequence up until then (see panel 9 in Figure 2).

Hypotheses

The experiment had the following aims: (1) to better understand the nature of ordering mechanisms of in musical STM through the study of temporal grouping effects in non-musicians, which in turn would allow (2) to assess the domain-generality hypothesis of serial order in STM. To achieve this, we compared recall performance for ungrouped and grouped 6-tone sequences, focusing on serial recall accuracy, the shape of the serial position curves, response latencies, and the rates of interposition errors. According to the domain-generality hypothesis of serial order STM, it was predicted to observe higher recall accuracy for grouped than ungrouped sequences. We also predicted the presence of a multiply-bowed serial position curve for grouped sequences. Finally, we expected to
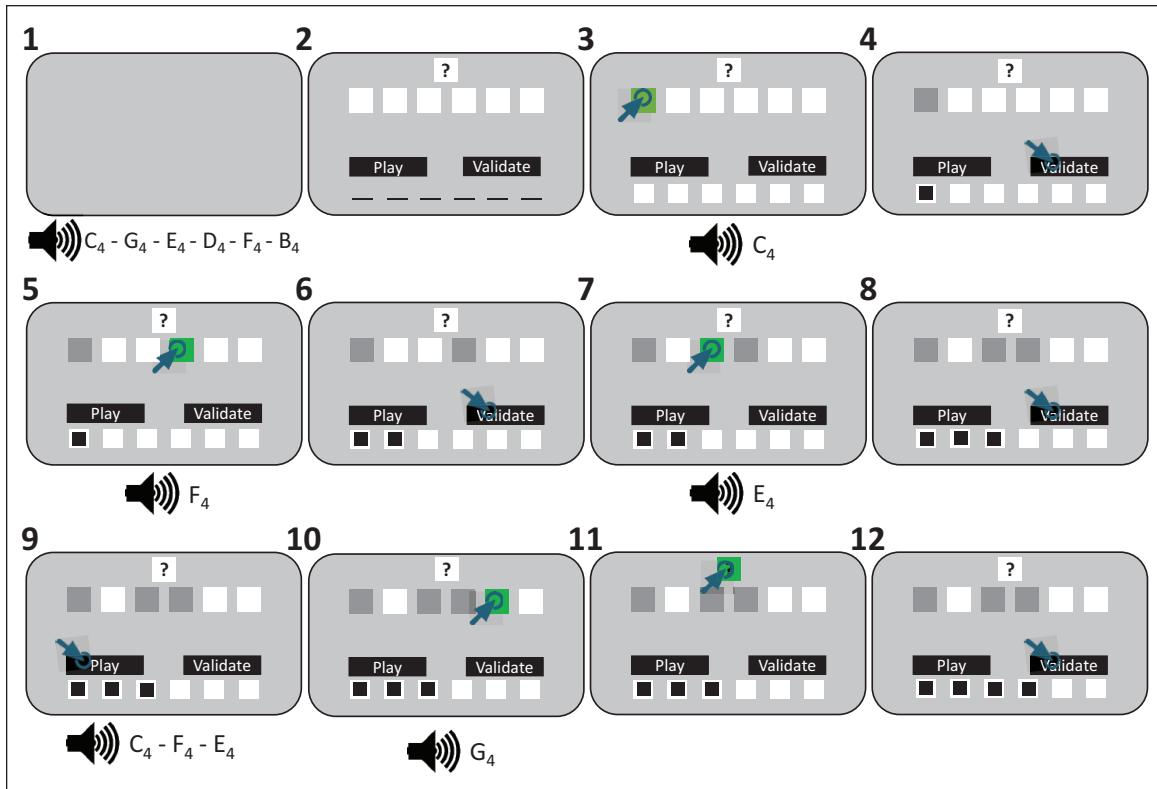
Figure 2. Graphical representation of the functioning of the serial order reconstruction task for tone sequences in Experiment 1. See the main text for more details about its functioning.

observe more interposition errors in grouped than ungrouped sequences.

Analysis plan

We used the open-source program JASP (version 0.14, JASP Team, 2018) with default settings for all planned (described here below) and exploratory analysis reported. For Bayesian *t*-tests, the prior was represented as a Cauchy distribution with an *r* scale of 0.707. For Bayesian analysis of variance (BANOVA), the prior also consisted of a Cauchy distribution, with an *r* scale of .5 and 1 for fixed and random effects, respectively.

Recall accuracy and serial position curve. We analysed serial position curves by averaging the recall accuracy as a function of serial position and grouping condition for each participant. Then, we performed a 2 × 6 repeated-measures BANOVA, with a 2-level type of sequence factor (ungrouped vs. grouped) and a 6-level serial position factor (from 1 to 6).

In case of an interaction between the two factors (i.e., the full model is the best model and is supported by a BF of at least 10, relative to the second-best model), we assessed the presence of mini-primacy and mini-recency effects in grouped sequences by comparing recall accuracy

between Positions 1 and 2 (H1: 1 > 2), Positions 2 and 3 (H1: 2 < 3), Positions 4 and 5 (H1: 4 > 5), and Positions 5 and 6 (H1: 5 < 6) via Bayesian paired samples *t*-tests.

Transposition gradients. We analysed transposition gradients by computing the proportion of transposition errors as a function of displacement separately for each condition and for each participant. To achieve this, we performed a 2 × 5 repeated-measures BANOVA with a 2-level type of sequence factor (ungrouped vs. grouped) and a 10-level displacement distance factor (from –5 to 5, excluding 0). If the full model turned out to be the best model (i.e., BF > 10 compared with the second best model), we analysed the interaction by focusing on the rate of adjacent displacements and interposition errors (see the next analysis for more details).

Interposition errors and adjacent displacement rates. The rate of interposition errors and displacements to adjacent serial positions was determined by calculating the proportion of errors involving between-group displacement of items keeping their initial within-group position (i.e., absolute distance of three positions) and the proportion of serial order transpositions characterised by an absolute displacement distance of one serial position among all serial order errors and separately for each type of sequence (ungrouped
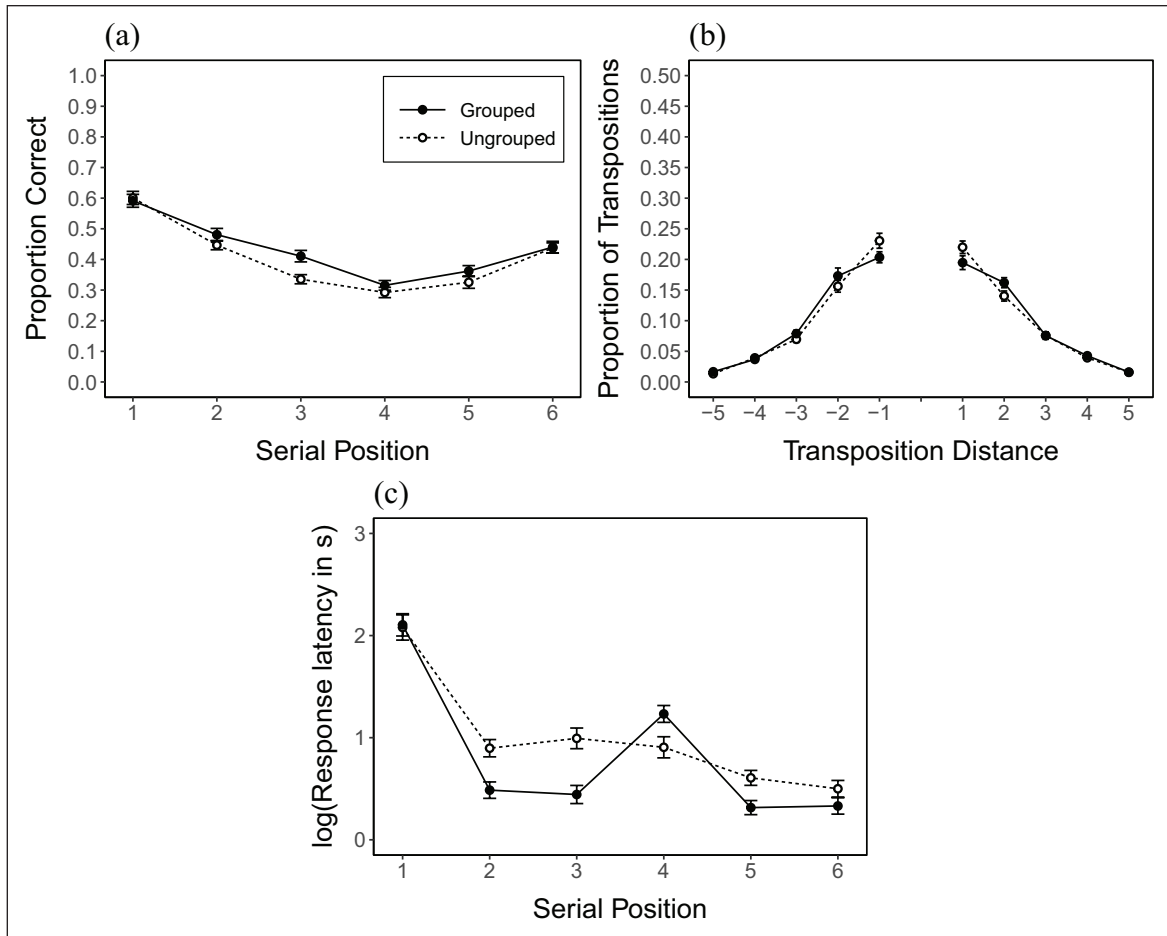
Figure 3. (a) Serial position curves, (b) transposition gradients, and (c) response latencies as a function of the type of grouping, from Experiment 1. Error bars represent confidence interval computed on data corrected for between-subject variability (Morey, 2008), following Baguley (2012, formula 8) recommendations.

vs. grouped). Then, the two grouping conditions were compared based on the observed rate of interposition errors (H1: interpositions in grouped sequences > interpositions in ungrouped sequences) and adjacent displacement (H1: adjacent displacement in grouped sequences < adjacent displacement in ungrouped sequences) via Bayesian paired samples *t*-tests.

Results

Planned analyses. The 2×6 repeated-measures BANOVA performed on recall accuracy as a function of serial position (1–6) and grouping condition (grouped vs. ungrouped), revealed that the best model is the model with the two main effects (see Figure 3a). This model is preferred over the second best, full model by a factor of 1.80 (see "Serial position curves" rows in Table 1). As the preference was characterised only by anecdotal evidence, we conducted an analysis of effect. This was done with JASP via a method averaging evidence across all the models containing the effect of interest. The data

provided decisive evidence in favour of the presence of a serial position effect ($BF_{Inclusion} = \infty$), very strong evidence in favour of a grouping effect ($BF_{Inclusion} = 31.28$), and anecdotal evidence in favour of the presence of the interaction ($BF_{Inclusion} = 2.15$). As initially planned, we did not analyse mini-primacy and mini-recency effects in grouped sequences as the presence of the interaction was not supported by the data.

The 2×10 repeated-measures BANOVA performed on the proportion of transposition errors as a function of transposition distance (−5 to 5, excluding 0) and grouping condition (grouped vs. ungrouped), revealed that the best model to explain the data is the full model (see Figure 3b). This model is preferred over the second best containing only the effect of distance by a factor of 173.36, representing decisive support for the best model (see "Transposition gradients" rows in Table 1). Given the clear support for an interaction between grouping condition and transposition distance, we compared the rate of adjacent transpositions and interpositions between the two grouping conditions as initially planned.

Table 1. Results of the Bayesian repeated-measures analyses of variances for the serial position curve, transposition gradients, and response latencies from Experiment 1.

Analysis	Models	P(M)	P(M | data)	BF_M	BF_{10}	Error %
Serial position curves	Null model (incl. subject)	0.20	$4.42e^{-83}$	1.77^{-82}	1.00	
	Condition + Position	0.20	0.63	6.80	$1.42e^{82}$	1.23
	Condition + Position + Condition × Position	0.20	0.35	2.15	$7.91e^{81}$	1.25
	Position	0.20	0.02	0.09	$4.72e^{80}$	1.50
	Condition	0.20	$6.97e^{-83}$	$2.79e^{-82}$	1.58	6.19
Transposition gradients	Null model (incl. subject)	0.20	0.00	0.00	1.00	
	Condition + Distance + Condition × Distance	0.20	0.99	648.69	∞	1.18
	Distance	0.20	$5.73e^{-3}$	0.02	∞	0.78
	Condition + Distance	0.20	$3.96e^{-4}$	$1.58e^{-3}$	∞	1.23
	Condition	0.20	0.00	0.00	0.07	6.26
Response latencies	Null model (incl. subject)	0.20	$1.12e^{-110}$	4.83^{-110}	1.00	
	Condition + Position + Condition × Position	0.20	1.00	$1.55e^8$	$8.28e^{109}$	1.23
	Condition + Position	0.20	2.65^{-8}	1.06^{-7}	$2.20e^{102}$	1.22
	Position	0.20	2.61^{-11}	1.04^{-10}	$2.16e^{99}$	1.50
	Condition	0.20	$5.82e^{-110}$	$2.33e^{-109}$	4.82	6.17

BF: Bayes factor.
All models include subject and for each analysis models are compared with the null model; Condition: temporal grouping effect; Position: serial position effect; Distance: transposition distance effect.

The directed Bayesian paired samples *t*-test comparing the rate of interposition errors between the two grouping conditions (H_1: ungrouped > grouped) provided anecdotal evidence in favour of the null model ($BF_{01} = 2.16$). Next, we compared the rate of adjacent transpositions between the two conditions via another Bayesian paired samples *t*-test (H_1: ungrouped < grouped). The results provided decisive evidence in favour of the presence of less adjacent transpositions in grouped than in ungrouped trials ($BF_{10} = 623.10$).

Exploratory analyses. Since the present study focused on exploring the nature of serial order representation in musical STM, it is critical to ensure that contour was not the dominant component in representing the sequences. Contour is a critical aspect of melodic representation, particularly in non-experts (see Dowling, 1978; Dowling & Tillmann, 2014). Thus, it is a possibility that the participants focused more on contour than item positions. If the grouping manipulation boosted recall performance, this could have influenced only contour-based representation. We then re-scored recall performance by considering an interval as correct when its direction (up or down) was the same as for the corresponding interval in the target sequence. Next, we compared the rate of above-chance correct recall for item position and contour scoring methods (subtracting 0.17 and 0.5 chance-level to item and contour scoring, respectively). The results of an undirected Bayesian paired samples *t*-test conducted on chance-corrected item position and contour scores provided decisive

Table 2. Proportions of within- and between-group transposition errors, as a function of grouping condition, from Experiment 1.

Grouping type	Within groups	Between groups		
		Boundary	Interpositions	Others
Ungrouped	.40	.10	.15	.35
	(.05)	(.03)	(.04)	(.07)
Grouped	.40	.08	.15	.37
	(.07)	(.04)	(.04)	(.10)

Values in parentheses are standard deviations.

evidence ($BF_{10} = 9.22e^5$) in favour of better performance when using the item position ($M = 0.24, SD = 0.11$) than the contour scoring method ($M = 0.18, SD = 0.09$).

To gain a better idea of the origin of the decrease of adjacent transposition errors in grouped sequences, we compared the rates of within and between-group displacements—the latest differentiating interpositions, non-interposition, and group-boundary displacements—between the two conditions of grouping (see Table 2). Exploratory comparisons performed via undirected Bayesian paired samples *t*-test suggest a moderate level of absence of difference between the two conditions regarding the rate of interpositions ($BF_{01} = 3.69$), within-group transpositions ($BF_{01} = 6.48$), and other between-group transpositions ($BF_{01} = 3.02$). Interestingly, the results revealed decisive evidence that a difference between the rate of displacements involving group boundaries ($BF_{10} = 153.29$) was present.

Finally, we took advantage of the changes introduced in the task, which made response behaviours more comparable with those characterising verbal serial recall, to perform an exploratory analysis of response latencies. This analysis is of interest because temporal grouping exerts an important effect on the pattern of recall timing, which is well accommodated by a two-dimensional representation of positional information (Lewandowsky & Farrell, 2008). In ungrouped sequences, response timing is characterised by a long latency for the initiation of the recall, followed by an inverted U-shaped response timing (Farrell & Lewandowsky, 2004). For grouped sequences, additional long latency is observed at the beginning of temporal groups, reflecting the temporal structure of the sequence (Farrell, 2008; Maybery et al., 2002). To determine the presence of such a pattern in the present study, we performed a BANOVA on the log of response latency (i.e., timing relative to the previous response or the last presented tone for the first responded item) for correct responses as a function of serial position (1–6) and grouping condition (grouped vs. ungrouped). The results revealed that the full model is the best model (see Figure 3c), preferred over the second best model containing only the effect of serial position by a factor of $3.77e^7$, representing decisive evidence supporting the presence of the two main effects and their interaction (see "Response latencies" rows in Table 1).

Discussion

Experiment 1 aimed to better understand the nature of serial order representations in musical STM. To achieve this goal, we tested whether, with tone sequences, temporal grouping exerts the same effects on recall accuracy, transposition errors, and response latencies as those reported with verbal material. We presented participants with ungrouped tone sequences and grouped tone sequences consisting of two groups of three items. The evidence that temporal grouping increased recall accuracy was strong. The effect of grouping on the shape of the serial position curve was anecdotal, with only limited scalloping. Analysis of response latencies showed a typical inverted U-shaped profile with a long latency for the first output item in the ungrouped condition, whereas we observed an increase in latency for the first output item in each group in the grouped sequences (for similar results in the verbal domain, see Farrell, 2008; Maybery et al., 2002). However, while temporal grouping reduced the rate of adjacent transpositions for items at group boundaries, a typical pattern in verbal STM for serial order (Henson, 1999; Maybery et al., 2002), we observed evidence against an increase in interposition errors in grouped sequences.

This experiment confirmed, using a serial recall procedure, the results of Gorin et al. (2018b) that temporal grouping provides an advantage in short-term recognition of musical stimuli. The pattern of grouping effects observed in this experiment is very similar to what is typically reported for similar verbal STM tasks: grouping induces scalloping of the serial position curve and provides a recall advantage (Frankish, 1985; Hitch et al., 1996; Ryan, 1969a), leads to a decrease in adjacent transpositions (Maybery et al., 2002), and response latency is longer at the beginning of groups (Farrell, 2008; Maybery et al., 2002). However, we did not observe the classical increase in interposition errors, which is a benchmark of temporal grouping and is considered as evidence for the existence of two-dimensional positional markers coding the positions of items within the groups and the positions of the groups or items in the sequence, respectively (Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998).

The results reported here mirror those observed with visuospatial material and where benchmarks of temporal grouping effects were also observed, except for the increase in interposition errors (Hurlstone, 2019). The authors accounted for the difference by proposing a model of serial order coding positional information in a slightly different way with regard to the type of material. For verbal information, two-dimensional markers code group positions in the sequence and item positions within the groups. For visuospatial material, the two-dimensional markers code group and item positions in the sequence. A straightforward account of the results reported here would be to assume that the same positional coding scheme is used for visuospatial and musical material, but that the increase in interposition errors in grouped sequence is specific to the positional coding scheme used for verbal information.

At the same time, the observation of interposition errors in the verbal domain is limited to a very specific context where the items are presented in a sequence of three groups of three items (e.g., Hartley et al., 2016; Henson, 1996; Hurlstone, 2019; Ng & Maybery, 2002, 2005; Ryan, 1969b). To the best of our knowledge, in the literature on temporal grouping effects with verbal sequences of six items (e.g., two groups of three items, see Farrell, 2008; Hitch et al., 1996; Maybery et al., 2002; Parmentier & Maybery, 2008),[2] there is no study reporting an increase in interposition errors in grouped sequences. Consequently, inferring the nature of serial order representation in the musical domain based on the assumption that in the verbal domain grouping sequences of nine or six items in groups of three should lead to the same pattern of grouping effects may represent a shortcoming. Thus, it is a possibility that the absence of increase in interposition errors with musical material is related to the use of 6-item sequences but not to the presence of different positional coding scheme between the verbal and musical domains. If this is the case, we should observe the same effect with verbal material as seen in the present experiment.[3] To explore this possibility,

we conducted an online study where participants had to recall sequences of letters in serial order where we manipulated the phonological similarity (similar vs. dissimilar) and the type of grouping (ungrouped vs. grouped).

Experiment 2: forward serial recall of verbal order

This experiment was conducted to determine whether the absence of increase in interposition errors increase in musical grouped sequences observed in Experiments 1 was due to the use of 6-item sequences or due to different positional representation compared with the verbal domain. To this aim, we conducted an online experiment requiring participants to recall visually presented letter lists. The first half of the experiment presented participants with ungrouped sequences and the other half with grouped ones. Moreover, to take into account the fact that with musical material performance can be negatively impacted by tonal proximity (Williamson et al., 2010), half of trials presented sequences composed of phonologically similar (e.g., D–G–C–T–P–V) letters and the other half presented dissimilar letters (e.g., R–L–K–M–F–S).

Method

Sampling plan. This experiment was conducted during the COVID-19 outbreak in Spring 2020. Due to organisational constraints, we had to allow all participants to take part in the study to ensure the validation of their course credits. Thus, no specific criteria were applied to prior participants who took part in this study as well. Consequently, the sampling plan consisted in letting as many students as possible to take part in the study. The URL of the experiment was shared with the participants via a forum used in one of their first-year psychology courses. Exclusion criteria was applied only once the data collection period ended.

Participants. The experiment was approved by the ethics committee of the Faculty of Psychology and Sciences of Education of the University of Geneva. A total of 101 first-year psychology students from the University of Geneva participated in this online experiment in exchange for partial course credit. After the exclusion of 15 participants who met the exclusion criteria, the final sample was composed of 86 participants (gender: 66 females, 19 males and 1 other; age in years: $M = 22.27$, $SD = 6.03$).

Exclusion criteria. We excluded participants with any learning or neurological disorder as well as those not fluent in French.

Stimuli. The stimuli consisted of 160, 6-letter sequences. Half of the sequences were composed of phonologically similar letters, drawn randomly without the replacement

six letters from the pool B, C, D, G, P, T, and V. The other half was composed of phonologically dissimilar letters, drawn without the replacement six letters from the pool X, H, J, L, K, Q, and S. When generating the sequences, we ensured that the same letter did not occur at the same serial position in consecutive trials and that all the sequences were unique. Finally, as in the previous experiments, a new set of 180 sequences was generated for each participant.

Procedure. Each trial started with a countdown from 3 to 1. The countdown was displayed in blue on a white background with a sans-serif font and a font size of 30. Each digit was presented in the centre of the screen for a duration of 500 ms, followed by a blank screen of 100 ms. Immediately after the countdown, the six letters were presented sequentially in the centre of the screen. Letters were displayed in black on a white background with a sans-serif font and using a font size of 40. Each letter was presented for a duration of 500 ms, followed by a blank screen lasting for 100 ms. In grouped trials, an additional pause of 500 ms was added between the third and fourth items.

Directly after the presentation of the last item, a response field represented by an array of six horizontal lines displayed from left to right was shown on the screen. Participants were required to recall the sequence by entering the letters in their order of presentation using the keyboard of their computer. The response field was automatically populated with participants' answers without any possibility to correct their response. Only letters could be entered in the response field. Once the participant typed six letters, a message inviting them to start the next trial was shown on the screen.

The experiment was separated into two blocks. In the first block, participants were presented with ungrouped and grouped sequences in the second block. The order of the trials presenting phonologically similar and dissimilar letters was random. Each block started with four trials, presenting two phonologically similar and dissimilar letter sequences. During the training session, participants had feedback regarding the accuracy of their response, but not during the experimental trials.

The task was programmed with lab.js, a free and open-source online study builder (Henninger et al., 2019). Then, the experiment was exported to and hosted on a protected server of the University of Geneva. The management of online data collection was performed with JATOS, an open-source and free online studies manager (Lange et al., 2015).

Hypotheses

The data obtained so far show that temporal grouping does not lead to increased interpositions when using musical

material, suggesting that different positional codes are required to represent verbal and musical serial order in STM. As there is no evidence in the verbal STM literature that grouping sequences in two groups of three items leads to an increase in interposition errors, an alternative interpretation would be to consider that this effect is specific to the use of longer grouping structures (e.g., 3×3 structure). If the latter is true and that similar positional codes underlie serial order representation in verbal and musical STM, we should observe the same pattern of data with verbal material as observed in Experiment 1 with musical material. In other words, we should observe a recall advantage, a scalloped serial position curve, and response latency with peaks at the beginning of groups for grouped sequences, but no increase in interposition errors. If the former interpretation is true, we should observe the same pattern with an additional increase in interposition errors for grouped sequences.

Regarding the effect of phonological similarity, no predictions were made before running the experiment, except for the expectation that recall accuracy should be worse for phonologically similar sequences. The manipulation of phonological similarity was implemented in this experiment only to take into account the fact there is an inherent effect of pitch proximity when using tone sequences, which is considered as a musical proxy of phonological similarity (Williamson et al., 2010).

Results

As in the previous experiment, the analyses were performed with JASP (JASP Team, 2018), using the same default values for priors and applying the same analysis plan. For each type of analysis (i.e., serial position curves, transposition gradients, and response latencies), data from trials presenting phonologically dissimilar and similar letters were analysed separately.

Serial position curves. We computed the proportion of correct recall as a function of serial position and temporal grouping across all the dissimilar trials for each participant. We then performed a 2×6 repeated-measures BANOVA with serial position (1–6) and grouping condition (grouped vs. ungrouped) factors (see top-left of Figure 4). The results revealed that the best model was the model with the two main effects, preferred over the second best, the full model, by a factor of 4.44 (see "Serial position curves" rows in Table 3). This was confirmed by an analysis of effect that provided decisive evidence for the two main effect (Grouping: $BF_{Inclusion}=1.43e^{14}$; Position: $BF_{Inclusion}=1.43e^{14}$), but anecdotal evidence against the presence of an interaction ($BF_{Inclusion}=0.90$).

The same analysis was performed with data from trials presenting phonologically similar letters, revealing that the best model was the full model and was preferred over the

second best model by a factor of 1.67 (see top-right of Figure 4). Given the ambiguous evidence for preferring the best model over the second best model (see "Serial position curves" rows in Table 4), we performed an analysis of effect. The results yielded decisive evidence in favour of the two main effects (Grouping: $BF_{Inclusion}=2.70e^{11}$; Position: $BF_{Inclusion}=6.67e^{13}$) and moderate evidence in favour of the existence of an interaction ($BF_{Inclusion}=6.70$).

Transposition gradients. Note that for the analysis of transposition errors we removed the participants that produced no error in at least one of the four experimental conditions, leading to a sample of 77 participants.

For each participant, we computed the proportion of errors as a function of absolute distance displacement and temporal grouping across all the dissimilar among all the errors. Then, we analysed the data with a $2 \times 2 \times 5$ repeated-measures BANOVA with absolute transposition distance (1–5) and grouping condition (grouped vs. ungrouped) as factors (see middle-left of Figure 4). The results provided strong evidence in favour of the best model containing only the effect of distance, being preferred over the second best model with the two main effects by a factor of 12.92 (see "Transposition gradients" rows in Table 3).

The same analysis has been reproduced with data from trials with phonologically similar letters (see middle-right of Figure 4). This provided strong evidence that the best model is the full model that was preferred over the second best model containing only the effect of distance by a factor of 1.02 (see "Transposition gradients" rows in Table 4). As the results were ambiguous, we performed an analysis of effects that revealed decisive and moderate evidence supporting the presence of an effect of distance ($BF_{Inclusion}=\infty$) and an interaction between distance and grouping ($BF_{Inclusion}=3.78$), respectively. Given the moderate support for the interaction, we analysed the rate of adjacent transpositions and interposition errors with directed Bayesian paired samples *t*-test (adjacent errors: $H_1 = $ ungrouped $>$ grouped; interpositions: $H_1 = $ ungrouped $<$ grouped), as in the previous experiment. We obtained strong evidence against both an increase in interposition errors ($BF_{01}=12.21$) and a decrease in adjacent transposition ($BF_{01}=25.02$) in grouped trials.

Then, as in the previous experiment, we analysed the rate of within-group and between-group transposition errors, distinguishing for the latest between interposition errors, group-boundary transpositions, and other between-group transpositions (all comparisons involved undirected Bayesian paired samples *t*-test with default prior). As shown in Table 5, there is strong evidence that temporal grouping in dissimilar trials induced an increase of within-group transposition but a decrease of transpositions involving items at the group boundary. At the same time, there was moderate evidence
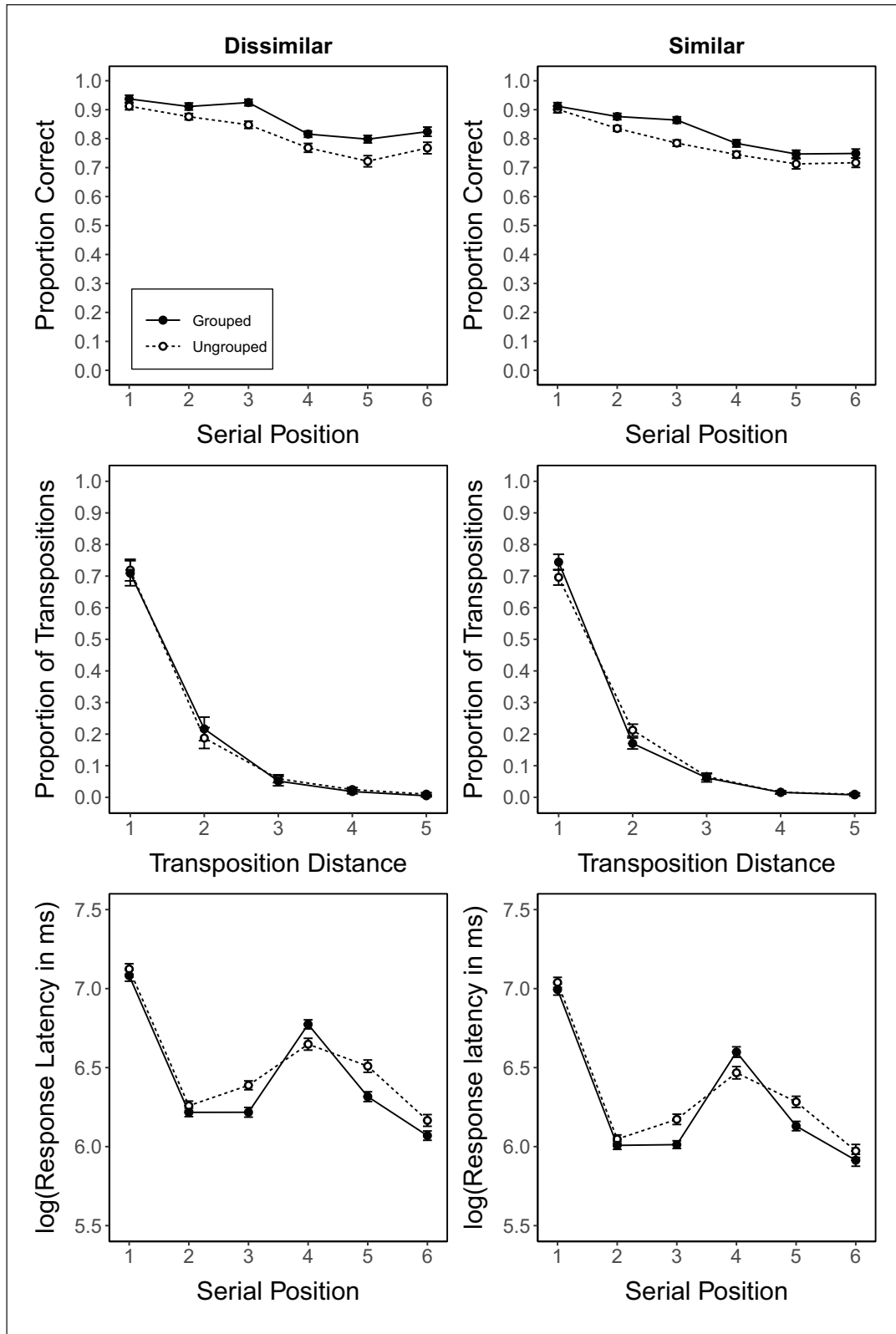
Figure 4. Top panels: serial position curve; middle panels: transposition gradients: bottom panels: response latencies. Left and right parts of the figure depict data from phonologically dissimilar and similar trials, respectively (Experiment 2). Error bars represent confidence interval computed on data corrected for between-subject variability (Morey, 2008), following Baguley (2012, formula 8) recommendations.

Table 3. Results of the Bayesian repeated-measures analyses of variances for the serial position curve, transposition gradients, and response latencies for phonologically dissimilar sequences from Experiment 2.

Analysis	Models	P(M)	P(M | data)	BF $_M$	BF $_{10}$	Error %
Serial position	Null model (incl. subject)	0.20	$1.16e^{-88}$	$4.63e^{-88}$	1.00	
	Condition + Position	0.20	0.82	17.76	$7.05e^{87}$	1.23
	Condition + Position + Condition × Position	0.20	0.18	0.90	$1.59e^{87}$	1.26
	Position	0.20	$1.99e^{-17}$	$7.98e^{-17}$	$1.72e^{71}$	1.49
	Condition	0.20	$9.95e^{-78}$	$3.98e^{-77}$	$8.59e^{10}$	6.18
Transposition gradients	Null model (incl. subject)	0.20	$4.45e^{-234}$	$1.78e^{-233}$	1.00	
	Distance	0.20	0.93	51.09	$2.08e^{233}$	1.94
	Condition + Distance	0.20	0.07	0.31	$1.61e^{232}$	1.27
	Condition + Distance + Condition × Distance	0.20	$8.12e^{-4}$	$3.25e^{-3}$	$1.82e^{230}$	1.32
	Condition	0.20	$3.70e^{-235}$	$1.48e^{-234}$	0.08	6.25
Response latencies	Null model (incl. subject)	0.20	$3.70e^{-256}$	$1.48e^{-255}$	1.00	
	Condition + Position + Condition × Position	0.20	1.00	$1.17e^9$	$2.70e^{255}$	1.26
	Condition + Position	0.20	$3.41e^{-9}$	$1.36e^{-8}$	$9.22e^{246}$	1.24
	Position	0.20	$2.42e^{-13}$	$9.67e^{-13}$	$6.53e^{242}$	1.54
	Condition	0.20	$9.88e^{-256}$	$3.95e^{-255}$	2.67	6.17

BF: Bayes factor.
All models include subject and for each analysis models are compared with the null model; Condition: temporal grouping effect; Position: serial position effect; Distance: transposition distance effect.

Table 4. Results of the Bayesian repeated-measures analyses of variances for the serial position curve, transposition gradients, and response latencies for phonologically similar sequences from Experiment 2.

Analysis	Models	P(M)	P(M | data)	BF $_M$	BF $_{10}$	Error %
Serial position	Null model (incl. subject)	0.20	$2.61e^{-109}$	$1.05e^{-108}$	1.00	
	Condition + Position + Condition × Position	0.20	0.63	6.70	$2.40e^{108}$	1.26
	Condition + Position	0.20	0.37	2.39	$1.43e^{108}$	1.24
	Position	0.20	$2.46e^{-12}$	$9.84e^{-12}$	$9.41e^{96}$	1.51
	Condition	0.20	$4.72e^{-103}$	$1.89e^{-102}$	$1.81e^6$	6.20
Transposition gradients	Null model (incl. subject)	0.20	0.00	0.00	1.00	
	Condition + Distance + Condition × Distance	0.20	0.49	3.78	∞	1.30
	Distance	0.20	0.48	3.65	∞	1.94
	Condition + Distance	0.20	0.04	0.15	∞	1.27
	Condition	0.20	0.00	0.00	0.08	6.25
Response latencies	Null model (incl. subject)	0.20	$3.97e^{-280}$	$1.59e^{-279}$	1.00	
	Condition + Position + Condition × Position	0.20	1.00	$8.54e^6$	$2.52e^{279}$	1.26
	Condition + Position	0.20	$4.64e^{-7}$	$1.86e^{-6}$	$1.17e^{273}$	1.25
	Position	0.20	$4.67e^{-9}$	$1.87e^{-8}$	$1.18e^{271}$	1.54
	Condition	0.20	$1.88e^{-280}$	$7.52e^{-280}$	0.47	6.17

BF: Bayes factor.
All models include subject and for each analysis models are compared with the null model; Condition: temporal grouping effect; Position: serial position effect; Distance: transposition distance effect.

supporting an absence of difference between the rates of interposition errors and other between-group transpositions. Regarding similar trials, the results reported in Table 6 show the exact same pattern as for dissimilar trials, except that there was strong evidence for a difference in the rate of other between-group transpositions.

Table 5. Proportions in dissimilar trials of within- and between-group transposition errors, as a function of grouping condition, for dissimilar trials from Experiment 2.

Grouping type	Within groups	Between groups		
		Boundary	Interpositions	Others
Ungrouped	.68	.13	.06	.13
	(.23)	(.17)	(.08)	(.15)
Grouped	.80	.05	.05	.10
	(.21)	(.09)	(.09)	(.15)
Pairwise-comparisons (ungrouped vs. grouped)	$BF_{10} = 29.38$	$BF_{10} = 34.82$	$BF_{01} = 6.46$	$BF_{01} = 3.63$

BF: Bayes factor.
Values in parentheses are standard deviations. Pairwise-comparisons were conducted with undirected Bayesian paired samples t-tests.

Table 6. Proportions in similar trials of within- and between-group transposition errors, as a function of grouping condition, for similar trials from Experiment 2.

Grouping type	Within groups	Between groups		
		Boundary	Interpositions	Others
Ungrouped	.67	.13	.07	.13
	(.15)	(.10)	(.06)	(.10)
Grouped	.79	.07	.06	.08
	(.19)	(.13)	(.08)	(.08)
Pairwise-comparisons (ungrouped vs. grouped)	$BF_{10} = 147.54$	$BF_{10} = 46.95$	$BF_{01} = 7.53$	$BF_{10} = 63.81$

BF: Bayes factor.
Values in parentheses are standard deviations. Pairwise-comparisons were conducted with undirected Bayesian paired samples t-tests.

Response latencies. For each participant, we determined the mean response latency for correct recall in dissimilar trials as a function of temporal grouping and serial position. The data were next analysed via a 2×6 repeated-measures BANOVA with serial position (1–6) and grouping condition (grouped vs. ungrouped) factors (see bottom left of Figure 4). The results yielded decisive evidence in favour of the full model containing the two main effects and their interaction, this model being preferred over the second best model by a factor of $2.93e^8$ (see Table 3). The same analysis has been performed with similar trials, leading to the same outcome (see bottom-right of Figure 4); the full model being the best model and preferred over the second best by a factor of $2.16e^6$ (see Table 4).

Discussion

In Experiment 2, we observed that regardless of the phonological similarity of the material, grouped sequences were better recalled and characterised by a scalloped serial position curve compared with ungrouped sequences. In addition, the typical pattern of response latencies with a latency peak for the first item of the second group was found. However, in line with the results reported in Experiment 1 with musical material, no increase in interposition errors was observed in grouped sequences for both phonologically similar and dissimilar trials. At the

same time, it should be noted that the performance can be seen as ceiling and that, in such a context, it is difficult to exclude the possibility that the absence of an increase in interposition errors in the grouped sequences is simply due to the fact that the overall number of errors was too low. To determine whether the lack of increase in interpositions is due to ceiling or is specific to the 2×3 grouping structure used in Experiment 2, we conducted an additional experiment replicating the procedure used in Experiment 2 but with an end-of-list distractor aimed at reducing recall performance while keeping the same sequence structure.[4]

Experiment 3: serial recall of verbal order with end-of-list distractor task

The goal of this experiment was to test whether the absence of an increase in interpositions in grouped sequences in Experiment 2 was due to the very low number of errors induced by a ceiling effect or specific to the use of lists of 6 items grouped by three. The procedure was the same as in Experiment 2, except that the presentation of each list was followed by a parity judgement task asking participants to judge whether numbers presented on the screen were even or odd. The purpose of this distracting task was to reduce the precision of the recall—and therefore increase the number of ordering errors—while keeping the same grouping structure as in Experiments 1 and 2.

Method

Sampling plan. Due to the COVID-19 pandemic situation, the experiment was conducted entirely online. As with Experiment 2, the sampling design was to let as many students and non-students from our participant pool take part in the study as possible.

Participants. The experiment was approved by the ethics committee of the Faculty of Psychology of UniDistance Suisse. Participants were recruited through the UniDistance Suisse participant pool, which is composed mainly of German-speaking psychology students and German-speaking non-students interested in participating in experiments. Students received partial course credit for their participation and non-students participated in the experiment on a voluntary basis. A total of 79 participants completed the online experiment. After excluding 14 participants who met the exclusion criteria, the final sample consisted of 55 participants (gender: 47 females and 8 males; age in years: $M = 35.83$, $SD = 9.43$).

Exclusion criteria. We excluded participants with any learning or neurological disorder as well as those who were not fluent in German. Participants were also excluded from the analysis based on their performance in the end-of-list distracting task, to ensure that they were actively performing the task. Therefore, any participant with less than 60% accuracy in the end-of-list distraction task was excluded from the analysis.

Stimuli. The stimuli were the same as in Experiment 2, but with two notable exceptions. First, due to the addition of a distraction task at the end of the list, the duration of a trial was increased compared with Experiment 2. Therefore, in order to keep the task to a similar duration as in Experiment 2, the total number of lists presented to the participant was 102 (25% phonologically similar and ungrouped, 25% phonologically similar and grouped, 25% phonologically dissimilar and ungrouped, and 25% phonologically dissimilar and grouped). Second, because the participants were German speakers, the phonologically dissimilar letters consisted of V, Y, X, Z, J, and Q, and the phonologically similar letters consisted of B, C, D, G, P, and T.

Procedure. The procedure was exactly the same as in Experiment 2, except for the addition of the end-of-list distractor. After the last item was presented, a blank screen was presented for 1,000 ms, followed by eight digits presented in the centre of the screen (700 ms on and 200 ms off). Participants were instructed to press the *S* key as quickly as possible when the digit presented was even, and to press *L* when the digit presented on the screen was odd. They were informed that they could press the keys during the presentation of the numbers as well as during the blank screen after each number was presented. The numbers were randomly selected with replacement.

After the end-list distractor, the recall procedure proceeded as described in Experiment 2. During the training session, participants received feedback after each trial regarding the number of letters correctly recalled and the number of correct parity judgements. No feedback was given during the experimental trials.

The task was programmed with lab.js, a free and open-source online study builder (Henninger et al., 2019), and implemented on a protected server with PHP. Participants accessed the experiment with a custom URL.

Hypotheses

The data from Experiments 1 and 2 support the view that temporal grouping has similar effects on the musical and verbal STM. It is noteworthy that the observed pattern in both domains indicates that for lists of 6 items grouped into threes, there is no increase in interposition errors, contrary to what would be predicted from serial order models that best account for temporal grouping effects in STM (see, for example, Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998). At the same time, the presence of a ceiling effect in recall accuracy in Experiment 2 limits this interpretation for the verbal domain. By adding an end-of-list distractor, this experiment aims to confirm the data from Experiment 2, namely that verbal lists of 6 items grouped into threes do not lead to an increase in interposition errors, as also observed in Experiment 1 with musical material.

In other words, this experiment aimed to test that verbal and musical STM are supported by common ordering mechanisms. The experiment also aimed to verify that the observation of increased interposition errors in recall of grouped lists is characteristic of longer sequences and/or sequences with more groups (e.g., a 3 × 3 grouping structure). If this hypothesis is correct, we would expect to observe the usual temporal grouping effects, except for the increase in interposition errors. As in Experiment 2, there was no specific prediction regarding the phonological similarity effect and its interaction with other factors, except that recall accuracy should be worse for phonologically similar sequences. As a reminder, this manipulation was introduced to have a closer comparison with musical material for which there is an inherent tonal proximity effect (Williamson et al., 2010).

Results

As in the previous experiment, the data were analysed using JASP (version 0.14, JASP Team, 2018) with the same default values for priors and applying the same analysis plan. For each analysis (i.e., serial position curves, transposition gradients, and response latencies), data from

trials presenting phonologically dissimilar and similar letters have been analysed separately.

Serial position curves. We calculated for each participant the proportion of correct recalls as a function of serial position and temporal grouping first for phonologically dissimilar trials. The data were then submitted to a 2×6 repeated-measures BANOVA with serial position (1–6) and grouping condition (grouped vs. ungrouped) factors (see top-left of Figure 5). The results revealed that the best model was the one with the two main effects only, preferred over the second best model (full model) by a factor of 32.76. This result represents strong evidence in favour of an effect of grouping on recall accuracy and on serial position, but no interaction between the two factors (see "Serial position curves" rows in Table 7).

The same analysis was performed with data from trials with phonologically similar letters, leading to the same pattern of data as for phonologically dissimilar letters, with the best model being the model with the two main effects, which was preferred to the full model by a factor of 68.68 (see top-right of Figure 5 and "Serial position curves" rows in Table 8).

Transposition gradients. Prior to statistical analysis of transposition errors, participants who produced no order errors in at least one of the four experimental conditions were removed. After the removal of these participants, the transposition error analysis was finally conducted on a sample of 51 participants.

We calculated for each participant the proportion of errors, as a function of absolute distance shift and temporal grouping, among all order errors in the phonologically similar condition. We then analysed the data with a 2×5 repeated-measures BANOVA with absolute transposition distance (1–5) and grouping condition (grouped vs. ungrouped) as factors (see middle-left of Figure 5). The results provided strong evidence in favour of the model containing only the distance effect as the best model, which was preferred to the second best model with both main effects by a factor of 10.56 (see "Transposition gradients" rows in Table 7).

The same analysis was repeated on data from trials with phonologically similar letters, leading to similar results to those obtained with phonologically dissimilar letters (see middle-right of Figure 5). The results provided strong evidence that the best model was the one with only a main effect of distance, preferred to the second best model with both main effects by a factor of 10.56 (see "Transposition gradients" rows in Table 8).

As in previous experiments, we also analysed the rate of within-group versus between-group transposition errors. For the latter, we distinguished between interposition errors, transpositions at groups boundary, and other between-group transpositions (all comparisons involved undirected Bayesian paired samples *t*-test with default

prior as provided in JASP). As shown in Table 9 the phonologically dissimilar lists showed overall moderate evidence for an absence of difference between the two grouping conditions with respect to different types of transposition errors. Regarding phonologically similar lists (see Table 10), we obtained decisive evidence of a decrease in transpositions at the groups boundary and moderate evidence for an absence of increase in interposition errors in grouped sequences.

Response latencies. For each participant, we averaged the log of response latency in milliseconds for each correct recall in dissimilar trials as a function of temporal grouping and serial position. The data were then analysed via a 2×6 repeated-measures BANOVA with serial position (1–6) and grouping condition (grouped vs. ungrouped) factors (see bottom-left of Figure 5). The results provided decisive evidence in favour of the full model containing the two main effects and their interaction, this model being preferred to the second best model with the two main effects by a factor of 19.47 (see "Response latencies" rows in Table 7). Response latencies for phonologically similar lists have been analysed in the same way, yielding anecdotal evidence ($BF_{10} = 1.65$) in favour of the full model (best model) when compared with the second best model containing only an effect serial position (see "Response latencies" rows in Table 8). Given the ambiguous evidence regarding the presence of an effect of interaction between serial position and grouping on response latency, we conducted an analysis of effect that provided moderate evidence for the presence of such an interaction ($BF_{Inclusion} = 3.93$, see also Figure 5).

Discussion

By introducing an end-of-list distracting task, Experiment 3 aimed to test whether the absence of an increase in interposition errors in the recall of 6-letter grouped lists in Experiment 2 was due to a ceiling in recall or to the specific 2×3 grouping structure used. The experiment also sought to determine whether the lack of increase in interposition errors in the 2×3 grouped sequences observed in Experiment 1 was specific to the musical domain or whether it is a more general feature of STM that extends to the verbal domain as well.

The end-of-list distractor has the expected effect of reducing recall accuracy relative to Experiment 2, especially for phonologically similar lists that are of particular interest for comparison with the music domain. We replicated the usual pattern of temporal grouping effects, but again observed an absence of increase in interposition errors. These results are in line with those reported in Experiments 1 and 2, suggesting that the absence of an increase in interposition errors in the recall of 6-letter lists grouped into two groups of three items was not due to a ceiling effect but might be related to the 2×3 grouping structure used in the experiment. This therefore supports
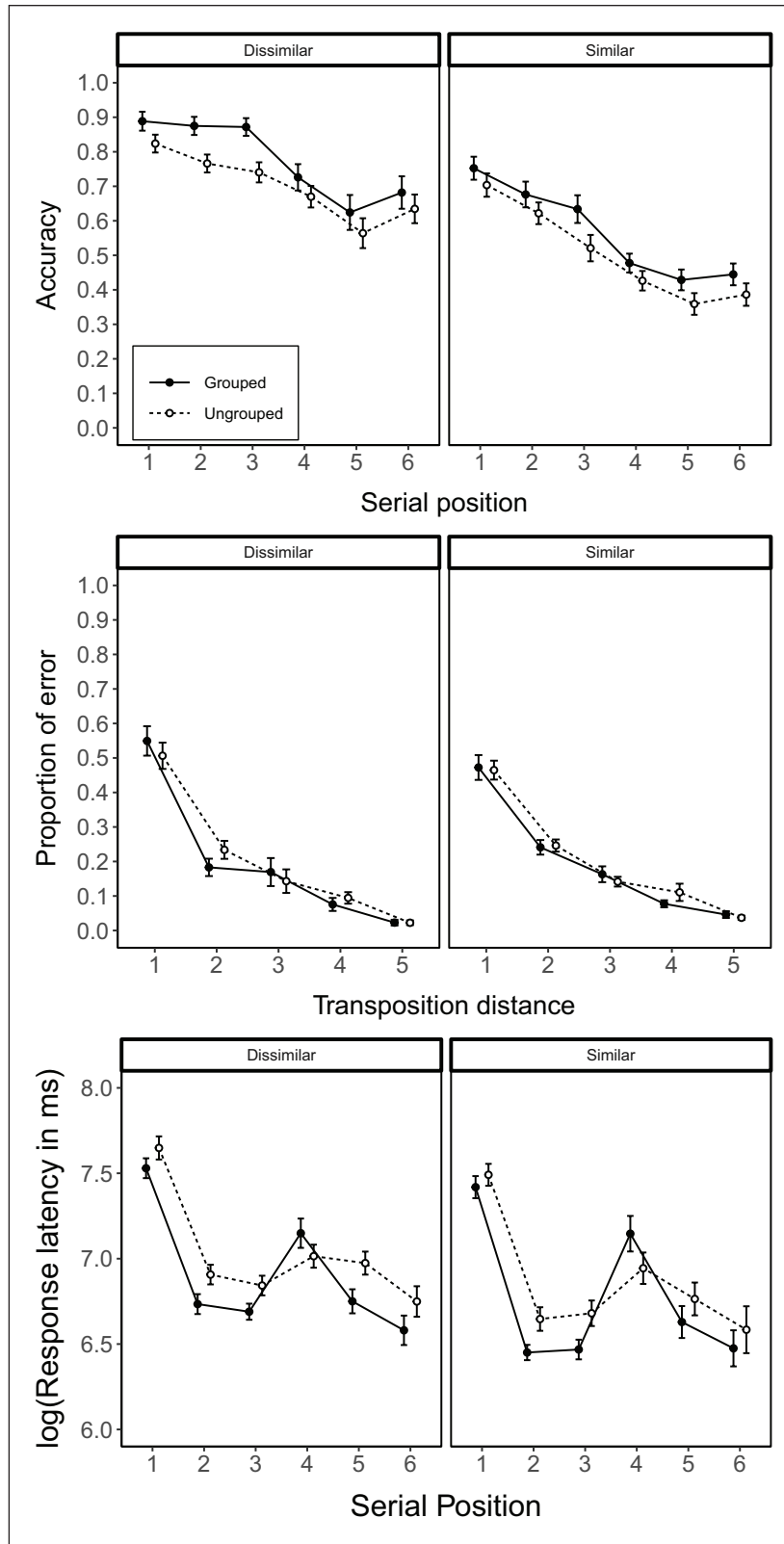
Figure 5. Top panels: serial position curve; middle panels: transposition gradients: bottom panels: response latencies. Left and right parts of the figure depict data from phonologically dissimilar and similar trials, respectively (Experiment 3). Error bars represent confidence interval computed on data corrected for between-subject variability (Morey, 2008), following Baguley (2012, formula 8) recommendations.

Table 7. Results of the Bayesian repeated-measures analyses of variances for the serial position curve, transposition gradients, and response latencies for phonologically dissimilar sequences from Experiment 3.

Analysis	Models	P(M)	P(M | data)	BF$_M$	BF$_{10}$	Error %
Serial position	Null model (incl. subject)	0.20	8.76e^{-48}	3.50e^{-47}	1.00	
	Grouping + Position	0.20	0.97	131.05	1.11e^{+47}	1.23
	Grouping + Position + Grouping $\times$ Position	0.20	0.03	0.12	3.38e^{+45}	1.24
	Position	0.20	6.31e^{-12}	2.52e^{-11}	7.20e^{+35}	1.42
	Grouping	0.20	6.17e^{-40}	2.47e^{-39}	7.04e^{+7}	6.14
Transposition gradients	Null model (incl. subject)	0.20	6.72e^{-111}	2.69e^{-110}	1.00	
	Distance	0.20	0.90	37.66	7.44e^{-111}	1.90
	Grouping + Distance	0.20	0.09	0.37	7.85e^{-110}	1.27
	Grouping + Distance + Grouping $\times$ Distance	0.20	0.01	0.04	6.47e^{-109}	1.31
	Grouping	0.20	6.83e^{-112}	2.73e^{-111}	10.23	6.26
Response latencies	Null model (incl. subject)	0.20	9.27e^{-83}	3.71e^{-82}	1.00	
	Grouping + Position + Grouping $\times$ Position	0.20	0.95	77.88	1.03e^{+82}	1.23
	Grouping + Position	0.20	0.05	0.21	5.27e^{+80}	1.23
	Position	0.20	1.18e^{-5}	4.73e^{-5}	1.28e^{+77}	1.48
	Grouping	0.20	2.13e^{-81}	8.51e^{-81}	22.95	6.14

BF: Bayes factor.
All models include subject and for each analysis models are compared with the null model; Condition: temporal grouping effect; Position: serial position effect; Distance: transposition distance effect.

Table 8. Results of the Bayesian repeated-measures analyses of variances for the serial position curve, transposition gradients, and response latencies for phonologically similar sequences from Experiment 3.

Analysis	Models	P(M)	P(M | data)	BF$_M$	BF$_{10}$	Error %
Serial position	Null model (incl. subject)	0.20	1.69e^{-81}	6.74e^{-81}	1.00	
	Grouping + Position	0.20	0.99	274.68	5.85e^{+80}	1.23
	Grouping + Position + Grouping $\times$ Position	0.20	0.01	0.06	8.51e^{+78}	1.26
	Position	0.20	2.51e^{-6}	1.01e^{-5}	1.49e^{+75}	1.49
	Grouping	0.20	5.55e^{-79}	2.22e^{-78}	329.14	6.17
Transposition gradients	Null model (incl. subject)	0.20	1.14e^{-140}	4.57e^{-140}	1.00	
	Distance	0.20	0.91	40.36	7.96e^{+139}	1.91
	Grouping + Distance	0.20	0.09	0.38	7.54e^{+138}	1.27
	Grouping + Distance + Grouping $\times$ Distance	0.20	4.02e^{-3}	0.02	3.52e^{+137}	1.31
	Grouping	0.20	1.17e^{-141}	4.64e^{-141}	0.10	6.26
Response latencies	Null model (incl. subject)	0.20	4.75e^{-64}	1.90e^{-63}	1.00	
	Grouping + Position + Grouping $\times$ Position	0.20	0.50	3.93	1.04e^{+63}	1.24
	Position	0.20	0.30	1.72	6.33e^{+62}	1.47
	Grouping + Position	0.20	0.20	1.02	4.29e^{+62}	1.24
	Grouping	0.20	1.45e^{-64}	5.80e^{-64}	0.31	6.17

BF: Bayes factor.
All models include subject and for each analysis models are compared with the null model; Condition: temporal grouping effect; Position: serial position effect; Distance: transposition distance effect.

that musical and verbal STM are characterised by similar temporal grouping effects—suggesting the presence of similar ordering mechanisms in both domains—while also indicating the presence of boundary conditions for observing increased interposition errors in the recall of grouped sequences from STM.

Table 9. Proportions in dissimilar trials of within- and between-group transposition errors, as a function of grouping condition, for dissimilar trials from Experiment 3.

Grouping type	Within groups	Between groups		
		Boundary	Interpositions	Others
Ungrouped	.54	.08	.15	.23
	(.17)	(.08)	(.15)	(.13)
Grouped	.58	.07	.18	.18
	(.22)	(.15)	(.18)	(.13)
Pairwise-comparisons (ungrouped vs. grouped)	BF01 = 4.33	BF01 = 5.41	BF01 = 2.83	BF01 = 1.43

BF: Bayes factor.
Values in parentheses are standard deviations. Pairwise-comparisons were conducted with undirected Bayesian paired samples t-tests.

Table 10. Proportions in similar trials of within- and between-group transposition errors, as a function of grouping condition, for dissimilar trials from Experiment 3.

Grouping type	Within groups	Between groups		
		Boundary	Interpositions	Others
Ungrouped	.51	.09	.14	.26
	(.12)	(.05)	(.06)	(.10)
Grouped	.57	.05	.16	.22
	(.18)	(.04)	(.10)	(.09)
Pairwise-comparisons (ungrouped vs. grouped)	BF10 = 1.22	BF10 = 229.75	BF01 = 3.36	BF10 = 2.52

BF: Bayes factor.
Values in parentheses are standard deviations. Pairwise-comparisons were conducted with undirected Bayesian paired samples t-tests.

General discussion

The goal of the present series of experiments was to determine whether the temporal grouping effects predicted by positional theories of serial order in verbal STM (see, for example, Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998) can be extended to the musical domain (Gorin et al., 2018a). In a first experiment, non-musicians were required to reconstruct the serial order of 6-tone sequences in a forward manner. The results showed that grouped sequences were overall better recalled than ungrouped sequences and that the former were characterised by a scalloped-shape recall curve reflecting the grouping structure used in the experiment. Response latencies adopted a classical inverted U-shape with longer latency for the first item in the list, as well as for the first item in the groups in grouped sequences. We did not observe an increase in interposition errors in the recall of temporally grouped musical sequences, but we reported a small decrease in adjacent transposition errors in grouped sequences, reflecting a decrease of transpositions involving items at group boundaries. Since interposition errors in 6-item grouped sequences are not well documented in the verbal STM literature, we conducted an online experiment requiring participants to serially recall grouped and ungrouped 6-letter sequences (Experiment 2) to compare with the observations from the musical domain. The pattern observed was similar to that observed in Experiment 1

but the conclusions were limited by the presence of ceiling effect at recall. In a last online experiment (Experiment 3), we asked participants to performed a task similar to that in Experiment 2 while introducing an end-of-list distractor to reduce ceiling effect. Even in the absence of ceiling effect we reproduced the same pattern of data as observed in Experiments 1 and 2, supporting the view that it is a general phenomenon that grouping 6-item sequences into groups of three is characterised by benchmark grouping effects but without an increase in interposition errors.

The experiments reported here provide additional evidence supporting the claim that temporal grouping effects observed in the verbal domain of STM could be extended to the musical domain as well (Gorin et al., 2018b). First, we obtained clear evidence from all experiments that presenting participants with 6-item verbal and musical sequences grouped by three lead to a recall advantage compared with the recall of the same, but ungrouped sequences. This replicates the recall advantage for grouped sequences observed with verbal (Farrell & Lewandowsky, 2004; Frankish, 1985; Hartley et al., 2016; Hitch et al., 1996; Ng & Maybery, 2002, 2005; Ryan, 1969b) and non-verbal materials (Hurlstone, 2019; Hurlstone & Hitch, 2015, 2018; Parmentier et al., 2004).

Second, in all three experiments, the serial position curve for grouped sequences was characterised by a scalloped appearance reflecting the 2 × 3 grouping structure used in the present study. It is noteworthy that while the

recall of grouped sequences showed a scalloped serial position curve, the interaction between serial position and grouping that characterises the scalloped shape was less strong than usually observed with longer grouped sequences (see, for example, Hartley et al., 2016; Ryan, 1969a). Indeed, the scalloping in our study was mainly limited to that of the first group, and this was similarly for musical and (phonologically similar) verbal sequences. This pattern is nonetheless in line with previous studies on temporal grouping with non-verbal sequences of similar length and grouping structure (Hurlstone & Hitch, 2018; Parmentier et al., 2004).

Third, the use of a forward reconstruction-of-order procedure with musical material in Experiment 1 allowed to demonstrate that the recall of musical material from STM is characterised by the same inverted U-shaped profile, but with a long latency for the first output position. In addition, grouped musical sequences showed an additional latency peak for the first item of each temporal groups. Although the latencies were from different timescales, the same pattern has been reproduced with 6-item verbal sequences in Experiments 2 and 3. This corroborates previous findings in the verbal (Farrell & Lewandowsky, 2004; Maybery et al., 2002; Parmentier & Maybery, 2008) and non-verbal (Hurlstone & Hitch, 2015, 2018; Parmentier et al., 2004) domains of STM regarding the profile of response latencies and the influence of temporal grouping on these latencies in STM tasks.

Finally, across all verbal and musical STM experiments, we observed that temporal grouping had none or only a limited influence on the pattern of transpositions. More importantly, we did not observe any increase in interposition errors. While this is in contradiction with the usually reported effect of temporal grouping on the pattern of transposition errors in the verbal domain (Henson, 1996, 1999; Ng & Maybery, 2002, 2005; Ryan, 1969b), we reported the same absence of effect of temporal grouping on interposition errors for the serial recall of 6-letter sequences in Experiments 2 and 3. Importantly, while data from Experiment 2 limit any interpretation of transposition patterns because of the presence of ceiling effect at recall, the comparison of data from Experiment 1 and phonologically similar verbal sequences from Experiment 3 (mimicking the pitch proximity inherent to musical sequences) clearly supports the view that, similar to the verbal and musical domains, grouping 6-item sequences by groups of three does not increase the proportion of interposition errors compared with ungrouped sequences.

Implication for theories of serial order STM

The observation of key grouping effects in forward reconstruction of tone sequences, as well as the reproduction of the same pattern of data in verbal and musical tasks observed in the current set of experiments is in favour of the view that representing serial order in musical and verbal STM could be supported by similar mechanisms. In the verbal domain, temporal grouping effects are well accommodated by models assuming that serial order is represented based on positional markers coding items or groups for their position in the sequence and within the groups (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). Consequently, the category of models assuming a hierarchical representation of serial order based on positional markers (Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998; Hurlstone, 2019) represents a good candidate to account for the effects reported in the musical and verbal STM tasks described in the present study, and suggests that serial order representation across these two domains is general.

At the same time, the absence of increase in interposition errors in grouped sequences is challenging for STM models assuming a hierarchical representation of serial order (Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016; Henson, 1998; Lewandowsky & Farrell, 2008). The ability of these models to account for grouping effects (Frankish, 1985, 1989; Hartley et al., 2016; Henson, 1996; Hitch et al., 1996; Maybery et al., 2002; Ng & Maybery, 2002, 2005; Ryan, 1969a, 1969b) relies on the hierarchical representation of positional information. However, the consequence of using hierarchical representation of serial order is that any model implementing that mechanism should predict an increase in interposition errors in grouped sequences, even with shorter sequences. As it is not clear from previous research whether the absence of increased interpositions is typical of the recall of 6-item sequences grouped with a 2×3 structure (see Farrell, 2008; Hitch et al., 1996; Maybery et al., 2002; Parmentier & Maybery, 2008), it is a possibility that this specific grouping structure represents a particular case. In some positional models (e.g., Brown et al., 2000; Henson, 1998), terminal positions are represented with greater distinctiveness. Thus, the positional codes of the two groups in a 2-group structure are more distinctive compared with, for instance, the positional codes between the second and third groups in a 3-group structure. It is then a possibility that a 2×3 grouping structure represents a special case in which there is no group at terminal positions, which then prevents the occurrence of interposition errors due to the increased distinctiveness between the groups. Further modelling work would be required to explore this account.

The analysis of interposition errors in grouped sequences is useful to better understand the mechanisms representing serial order in STM and to study the nature of these mechanisms across different domains. Hurlstone (2019) showed that the recall of visuospatial and verbal sequences grouped with a 3×3 structure are characterised by different patterns of transposition errors. To explain the absence of increase in interposition errors in the visuospatial domain, the authors suggested that positional information may be represented differently for visuospatial information. In the present study, the absence of increase

of interpositions does not seem to be linked to the STM domain, but rather appears as specific to the 2 × 3 grouping structure used. Consequently, contrary to the comparison between the visuospatial and verbal domains for which the same grouping structure leads to different patterns of transposition errors and suggests the existence of different ordering codes (Hurlstone, 2019, see also Soemer & Saito, 2016 for similar claim), the comparison in the present study rather supports the existence of similar mechanisms for ordering musical and verbal information while highlighting the fact that the observation of increased interposition errors is dependent on the type of grouping pattern.

Although the pattern of temporal grouping effects is similar across the musical and verbal domains in this study, it does not disprove evidence for domain-specificity of serial order in STM (Hurlstone, 2019; Logie et al., 2016; Saito et al., 2008; Soemer & Saito, 2016). Indeed, the results are also compatible with the view that domain-specific but functionally similar mechanisms for the retention of serial order exists across different domains (Logie et al., 2016). Further studies are then required to distinguish more precisely between the domain-general *versus* domain-specific theories of serial order in the verbal and musical domains of STM. Investigating the effect of cross-modal interference of order between musical and verbal domains in dual-task setting may be of great interest to tackle that question (Depoorter & Vandierendonck, 2009; Vandierendonck, 2016).

Methodological advances in studying musical STM for serial order

This series of experiments extends previous work on the development of a tool to study serial order phenomena in musical STM (Gorin et al., 2018a, 2018b). To address the question of the domain-generality of serial order mechanisms in STM, it is critical to use memory tasks having the same ordering requirements across domains. Gorin et al. (2018a) showed that using the same task as in Experiments 1A and 1B, recall of tone sequences in non-musicians was characterised by errors patterns and sequence length effects similar to those reported in verbal STM tasks. They also reported that the presence of serial position effects was characterised by smaller primacy and recency effects compared with what is usually reported with verbal tasks. In Experiments 1A and 1B, we reproduced the observation of serial position effects characterised by primacy and recency in ungrouped tone sequences, as well as typical transposition gradients, as observed with verbal material (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). In Experiment 2, participants had to reconstruct the sequence in forward serial order, making the task closer to the typical procedure used in verbal serial recall tasks. We also used a larger number of tones, instead of always using the same six tones, to reduce intertrial interference. This new procedure led to a clear improvement in recall accuracy

compared with Experiments 1A and 1B as well as with more pronounced serial position effects. We also replicated the typical transposition gradients and were able to analyse response latencies, the latter being characterised by a shape similar to what is usually reported in the verbal domain (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008). The presence of the same pattern of movement errors and forward recall serial position effects as reported with verbal material (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008) in our musical STM task in Experiment 2 supports the reliability of this task to study serial order phenomena in the musical domain, and opens the way to systematic comparison between order phenomena observed in the musical and other domains.

At the same time, it is important to note that the procedure developed in Experiment 2 still has some critical differences compared with verbal reconstruction tasks. For the latter, participants are asked to reconstruct the serial order of sequences of items (e.g., words, letters, and digits) that are represented at recall. The same principle was implemented in our musical task. However, even though the tones were organised from the lowest (left) to the highest (right) at recall to simplify the procedure, the participants had to search for the correct tones by clicking on the different items. Compared with verbal reconstruction tasks for which there is a direct access to the item at recall, this procedure inevitably created more interference. This could partially explain the poorer performance in the musical domain, in addition to the fact that participants were non-musicians and required more time to complete a trial. In turn, this would explain, at least in part, the presence of overall longer response latencies compared with the verbal domain.

Future directions

Future research should focus on adapting verbal serial reconstruction tasks to match the procedure used in Experiment 1. We could imagine an experiment in which participants would be asked to perform a serial order reconstruction task, as described in Experiment 1, with either tones or auditory consonants. This represents a more direct comparison between the two domains as the two tasks would be the exact same, except for the stimuli, allowing then to draw a better conclusion regarding the generality of serial order phenomena in the verbal and musical domains of STM.

It is important to note that the order phenomena characterising verbal STM tasks (Hurlstone et al., 2014; Lewandowsky & Farrell, 2008) are usually reported when testing a population of adults that are highly familiar with the memoranda (e.g., letters, digits, and words). In other words, one could consider that verbal order phenomena in STM reflects the behaviour of verbal experts in maintaining the order of verbal information (for language-based accounts of serial order processing in verbal STM, see

Acheson & MacDonald, 2009; Majerus, 2013; Schwering & MacDonald, 2020). Consequently, comparing the effect of grouping in musical STM with non-musicians to the same effect in the verbal domain with verbal experts may represent a sub-optimal comparison. A more optimal strategy to assess the domain-general hypothesis of serial order would be to explore the effect of grouping on the reconstruction of 9-tone sequences (e.g., 3-item group structure) in musicians. Using a melodic dictation recall method, Deutsch (1980) showed a positive effect of temporal grouping on recall accuracy of 12-tone sequences in musicians, as well as scalloped serial position curves. In addition, it has been shown that the recall from long-term memory of melodies played with a piano is characterised by interposition-like errors in sequences with a strong metrical structure (Mathias et al., 2015). Assuming the feasibility of asking musicians to reconstruct 9-tone sequences using the procedure described in Experiment 1, and considering the data from Deutsch (1980) and Mathias et al. (2015), further studies exploring grouping effects in musical STM in musicians are required to provide a more stringent test of the domain-generality hypothesis of positional markers in STM.

The observation of an absence of increase in interposition errors in recalling 2×3 grouped sequences, consistently with both musical and verbal material (even in absence of ceiling effect), supports the potential existence of boundary conditions to observe an effect of temporal grouping on transposition errors in STM. While addressing that matter is out of the scope of the current paper, that observation places new constraints on models of serial order. In addition, studying more systematically the factors (e.g., sequence length, group sizes, number of groups), and their interaction, driving the increase in interposition errors in serial recall may help in shedding new light on our understanding of serial order representation in STM. Recent work by Kowialiewski et al. (2021) has shown that sequences of six words grouped into pairs are characterised by an increase in interposition errors compared with the same ungrouped sequences. This indicates that the observation of an increase in interposition errors in grouped sequences seems to depend more on the number of groups in the sequence, rather than on the length of the sequence.

Conclusion

We observed benchmark temporal grouping effects in serial order reconstruction tasks with tone sequences, except for the typical effect of grouping on interposition errors. This pattern was replicated with the serial recall of verbal sequences comparable to the musical material used in the first experiments. The results overall support the view that positional markers described in verbal models of STM to represent serial order (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998) could be extended to the musical domain as well. Further research is nonetheless required to determine whether direct support for positional markers can be witnessed with longer musical sequences in musicians.

ORCID iD

Simon Gorin https://orcid.org/0000-0001-7462-0446

Data accessibility statement

All relevant data and materials (i.e., raw data, processing scripts, statistical analyses output files, and the experimental tasks) are openly accessible via time-stamped registration of the project on the Open Science Framework (OSF) (https://osf.io/e5dg3). The research reported in this paper has been initiated by a series of two preregistered experiments not reported here (the time-stamped preregistration of these experiments is available on the OSF: https://osf.io/kmy7j; the conclusion from these experiments is also available on the OSF: https://osf.io/c8k95/). Experiment 1 has been preregistered prior to data collection (the time-stamped preregistration of the experiment is available on the OSF: https://osf.io/t4rm3). Experiments 2 and 3 were not preregistered.

Notes

1. The research presented in this paper is an extension of the study of temporal grouping effects in musical STM initiated by Gorin et al. (2018a). To this aim, we first conducted two experiments to study the effect of temporal grouping on the unconstrained reconstruction of order of 6-tone sequences. These two studies were preregistered and a discussion of their results can be found on the OSF (see the Open Practice Statement) and in version 2 of the preprint of this paper posted on PsyArXiv (https://psyarxiv.com/mfynu). Since these two experiments used the same unconstrained reconstruction of order procedure as in Gorin et al. (2018a), it

has been decided to not report these two experiments in this paper that focus on forward serial recall. Moreover, there is evidence that unconstrained and forward reconstruction of order can behave differently (Lewandowsky et al., 2008).

2. Note that Farrell and Lewandowsky (2004) did an analysis of the transposition errors with 6-digit sequences and reported an absence of effect of temporal grouping on the transpositions profile but, as described by the authors, this is explained by the ceiling effect in recall accuracy.

3. An alternative would have been to perform the same experiment as Experiment 1 but with longer sequences—that is, 3 × 3 grouped sequences. However, we chose to work with verbal material to avoid floor effect on the recall accuracy if long sequences were used. Although recruiting a population of participants with musical expertise could have been the solution to use longer tone sequences, the present study focused on STM in non-musical participants. The choice was therefore made to use verbal material instead of long tone sequences.

4. The author is indebted to an anonymous reviewer for this suggestion.

References

Acheson, D. J., & MacDonald, M. C. (2009). Verbal working memory and language production: Common approaches to the serial ordering of verbal information. *Psychological Bulletin*, *135*(1), 50–68. https://doi.org/10.1037/a0014411

Audacity Team. (2017). *Audacity (version 2.2.1)* [Computer software]. https://www.audacityteam.org

Baguley, T. (2012). Calculating and graphing within-subject confidence intervals for Anova. *Behavior Research Methods*, *44*(1), 158–175. https://doi.org/10.3758/s13428-011-0123-7

Berger, J. O., & Berry, D. A. (1988). Statistical analysis and the illusion of objectivity. *American Scientist*, *76*, 159–165.

Berz, W. L. (1995). Working memory in music: A theoretical model. *Music Perception: An Interdisciplinary Journal*, *12*(3), 353–364. https://doi.org/10.2307/40286188

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127–181. https://doi.org/10.1037/0033-295X.107.1.127

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581. https://doi.org/10.1037/0033-295X.106.3.551

Depoorter, A., & Vandierendonck, A. (2009). Evidence for modality-independent order coding in working memory. *Quarterly Journal of Experimental Psychology*, *62*(3), 531–549. https://doi.org/10.1080/17470210801995002

Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception & Psychophysics*, *28*(5), 381–389. https://doi.org/10.3758/BF03204881

Deutsch, D. (2013a). Grouping mechanisms in music. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 183–248). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00006-7.

Deutsch, D. (2013b). The processing of pitch combinations. In D. Deutsch (Ed.), *The psychology of music* (3rd ed., pp. 249–325). Elsevier. https://doi.org/10.1016/B978-0-12-381460-9.00007-9.

Dienes, Z. (2016). How Bayes factors change scientific practice. *Journal of Mathematical Psychology*, *72*, 78–89. https://doi.org/10.1016/j.jmp.2015.10.003

Dowling, W. J. (1978). Scale and contour: Two components of a theory of memory for melodies. *Psychological Review*, *85*(4), 341–354. https://doi.org/10.1037/0033-295X.85.4.341

Dowling, W. J., & Tillmann, B. (2014). Memory improvement while hearing music. *Music Perception: An Interdisciplinary Journal*, *32*(1), 11–32. https://doi.org/10.1525/mp.2014.32.1.11

Farrell, S. (2008). Multiple roles for time in short-term memory: Evidence from serial recall of order and timing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(1), 128–145. https://doi.org/10.1037/0278-7393.34.1.128

Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin & Review*, *9*(1), 59–79. https://doi.org/10.3758/BF03196257

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*(1), 115–135. https://doi.org/10.1016/j.jml.2004.03.007

Frankish, C. R. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200–209. https://doi.org/10.1016/0749-596X(85)90024-5

Frankish, C. R. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 469–479. https://doi.org/10.1037/0278-7393.15.3.469

Gorin, S., Kowialiewski, B., & Majerus, S. (2016). Domain-generality of timing-based serial order processes in short-term memory: New insights from musical and verbal domains. *PLOS ONE*, *11*(12), Article e0168699. https://doi.org/10.1371/journal.pone.0168699

Gorin, S., Mengal, P., & Majerus, S. (2018a). A comparison of serial order short-term memory effects across verbal and musical domains. *Memory & Cognition*, *46*(3), 464–481. https://doi.org/10.3758/s13421-017-0778-0

Gorin, S., Mengal, P., & Majerus, S. (2018b). Temporal grouping effects in musical short-term memory. *Memory*, *26*(6), 831–843. https://doi.org/10.1080/09658211.2017.1414848

Hartley, T., Hurlstone, M. J., & Hitch, G. J. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135–178. https://doi.org/10.1016/j.cogpsych.2016.05.001

Henninger, F., Shevchenko, Y., Mertens, U. K., Kieslich, P. J., & Hilbig, B. E. (2019). *Lab.js: A free, open, online study builder*. 10.31234/osf.io/fqr49

Henson, R. N. A. (1996). *Short-term memory for serial order* [Unpublished doctoral thesis, Cambridge University, Cambridge]. http://www.mrc-cbu.cam.ac.uk/people/rik.henson/personal/thesis/

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, *36*(2), 73–137. https://doi.org/10.1006/cogp.1998.0685

Henson, R. N. A. (1999). Positional information in short-term memory: Relative or absolute? *Memory & Cognition*, *27*(5), 915–927. https://doi.org/10.3758/BF03198544

Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal grouping effects in immediate recall: A working memory analysis. *Quarterly Journal of Experimental Psychology Section A*, *49*(1), 116–139. https://doi.org/10.1080/713755609

Hurlstone, M. J. (2019). Functional similarities and differences between the coding of positional information in verbal and spatial short-term order memory. *Memory*, *27*, 147–162. https://doi.org/10.1080/09658211.2018.1495235

Hurlstone, M. J., & Hitch, G. J. (2015). How is the serial order of a spatial sequence represented? insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*(2), 295–324. https://doi.org/10.1037/a0038223

Hurlstone, M. J., & Hitch, G. J. (2018). How is the serial order of a visual sequence represented? insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(2), 167–192. https://doi.org/10.1037/xlm0000440

Hurlstone, M. J., Hitch, G. J., & Baddeley, A. D. (2014). Memory for serial order across domains: An overview of the literature and directions for future research. *Psychological Bulletin*, *140*(2), 339–373. https://doi.org/10.1037/a0034221

JASP Team. (2018). *JASP (version 0.10.0)[computer software]*. https://jasp-stats.org/

Jones, D. M., Farrand, P., Stuart, G., & Morris, N. (1995). Functional equivalence of verbal and spatial information in serial short-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*(4), 1008–1018. https://doi.org/10.1037/0278-7393.21.4.1008

Kieras, D. E., Meyer, D. E., Mueller, S., & Seymour, T. (1999). Insights into working memory from the perspective of the epic architecture for modeling skilled perceptual-motor and cognitive human performance. In A. Miyake & P. Shah (Eds.), *Models of working memory* (pp. 183–223). Cambridge University Press. https://doi.org/10.1017/CBO9781139174909.009.

Kowialiewski, B., Gorin, S., & Majerus, S. (2021). Semantic knowledge constrains the processing of serial order information in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0001031

Krumhansl, C. L. (1990). Tonal hierarchies and rare intervals in music cognition. *Music Perception: An Interdisciplinary Journal*, *7*(3), 309–324. https://doi.org/10.2307/40285467

Lange, K., Kühn, S., & Filevich, E. (2015). "just another tool for online studies" (jatos): An easy solution for setup and management of web servers supporting online studies. *PLOS ONE*, *10*(6), Article e0130834. https://doi.org/10.1371/journal.pone.0130834

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. In H. R. Brian (Ed.), *The psychology of learning and motivation* (*Vol. 49*, pp. 1–48). Academic Press. https://doi.org/10.1016/S0079-7421(08)00001-7.

Lewandowsky, S., Nimmo, L. M., & Brown, G. D. A. (2008). When temporal isolation benefits memory for serial order. *Journal of Memory and Language*, *58*(2), 415–428. https://doi.org/10.1016/j.jml.2006.11.003

Logie, R. H., Saito, S., Morita, A., Varma, S., & Norris, D. (2016). Recalling visual serial order for verbal sequences. *Memory & Cognition*, *44*(4), 590–607. https://doi.org/10.3758/s13421-015-0580-9

Majerus, S. (2013). Language repetition and short-term memory: An integrative framework. *Frontiers in Human Neuroscience*, 7, Article 357. https://doi.org/10.3389/fnhum.2013.00357

Mathias, B., Pfordresher, P. Q., & Palmer, C. (2015). Context and meter enhance long-range planning in music performance. *Frontiers in Human Neuroscience*, 8, Article 1040. https://doi.org/10.3389/fnhum.2014.01040

Maybery, M. T., Parmentier, F. B. R., & Jones, D. M. (2002). Grouping of list items reflected in the timing of recall: Implications for models of serial verbal memory. *Journal of Memory and Language*, *47*(3), 360–385. https://doi.org/10.1016/S0749-596X(02)00014-1

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061

Musacchia, G., Large, E. W., & Schroeder, C. E. (2014). Thalamocortical mechanisms for integrating musical tone and rhythm. *Hearing Research*, *308*, 50–59. https://doi.org/10.1016/j.heares.2013.09.017

Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *Quarterly Journal of Experimental Psychology Section A*, *55*(2), 391–424. https://doi.org/10.1080/02724980143000343

Ng, H. L. H., & Maybery, M. T. (2005). Grouping in short-term memory: Do oscillators code the positions of items? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*(1), 175–181. https://doi.org/10.1037/0278-7393.31.1.175

Ockelford, A. (2007). A music module in working memory? evidence from the performance of a prodigious musical savant. *Musicae Scientiae*, *11*(2 Suppl.), 5–36. https://doi.org/10.1177/10298649070110S202

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*(4), 761–781. https://doi.org/10.1037/0033-295X.105.4.761-781

Palmer, C., & Pfordresher, P. Q. (2003). Incremental planning in sequence production. *Psychological Review*, *110*(4), 683–712. https://doi.org/10.1037/0033-295X.110.4.683

Parmentier, F. B. R., & Maybery, M. T. (2008). Equivalent effects of grouping by time, voice, and location on response timing in verbal serial memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(6), 1349–1355. https://doi.org/10.1037/a0013258

Parmentier, F. B. R., Maybery, M. T., & Jones, D. M. (2004). Temporal grouping in auditory spatial serial memory. *Psychonomic Bulletin & Review*, *11*(3), 501–507. https://doi.org/10.3758/BF03196602

Peretz, I., & Coltheart, M. (2003). Modularity of music processing. *Nature Neuroscience*, *6*(7), 688–691. https://doi.org/10.1038/nn1083

Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, accuracy, and serial order in sequence production. *Cognitive Science*, *31*(1), 63–98. https://doi.org/10.1080/03640210709336985

Rouder, J. N. (2014). Optional stopping: No problem for Bayesians. *Psychonomic Bulletin & Review*, *21*(2), 301–308. https://doi.org/10.3758/s13423-014-0595-4

Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*(2), 137–147. https://doi.org/10.1080/14640746908400206

Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, *21*(2), 148–155. https://doi.org/10.1080/14640746908400207

Saito, S., Logie, R. H., Morita, A., & Law, A. (2008). Visual and phonological similarity effects in verbal immediate serial recall: A test with kanji materials. *Journal of Memory and Language*, *59*(1), 1–17. https://doi.org/10.1016/j.jml.2008.01.004

Schwering, S. C., & MacDonald, M. C. (2020). Verbal working memory as emergent from language comprehension and production. *Frontiers in Human Neuroscience*, 14, Article 68. https://doi.org/10.3389/fnhum.2020.00068

Soemer, A., & Saito, S. (2016). Domain-specific processing in short-term serial order memory. *Journal of Memory and Language*, *88*, 1–17. https://doi.org/10.1016/j.jml.2015.12.003

Vandierendonck, A. (2016). Modality independence of order coding in working memory: Evidence from cross-modal order interference at recall. *Quarterly Journal of Experimental Psychology*, *69*(1), 161–179. https://doi.org/10.1080/17470218.2015.1032987

Wagenmakers, E.-J., Beek, T. F., Rotteveel, M., Gierholz, A., Matzke, D., Steingroever, H., & Pinto, Y. (2015). Turning the hands of time again: A purely confirmatory replication study and a Bayesian analysis. *Frontiers in Psychology, 6*, Article 494. https://doi.org/10.3389/fpsyg.2015.00494

Wagenmakers, E.-J., Love, J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., . . .Morey, R. D. (2018). Bayesian inference for psychology. part ii: Example applications with JASP. *Psychonomic Bulletin & Review*, *25*(1), 58–76. https://doi.org/10.3758/s13423-017-1323-7

Williamson, V. J., Baddeley, A. D., & Hitch, G. J. (2010). Musicians' and nonmusicians' short-term memory for verbal and musical sequences: Comparing phonological similarity and pitch proximity. *Memory & Cognition*, *38*(2), 163–175. https://doi.org/10.3758/MC.38.2.163

Zatorre, R. J., Belin, P., & Penhune, V. B. (2002). Structure and function of auditory cortex: Music and speech. *Trends in Cognitive Sciences*, *6*(1), 37–46. https://doi.org/10.1016/S1364-6613(00)01816-7