

## **Temporal grouping effects in musical short-term memory**

Simon Gorin<sup>a,b</sup>, Pierre Mengal<sup>a</sup> and Steve Majerus<sup>a,b</sup>

*<sup>a</sup>Psychology and Neuroscience of Cognition Research Unit (PsyNCog), University of Liège, Liège, Belgium*

*<sup>b</sup>Fund for Scientific Research–FNRS, Brussels, Belgium*

Correspondence concerning this article should be addressed to Simon Gorin, who is now at the Department of Cognitive Development, Université de Genève, 40 Boulevard du Pont d'Arve, 1205 Genève, Switzerland. E-mail: gorinsimon@gmail.com

All relevant data are available through the Open Science Framework

(<https://osf.io/tdhkv/>).

**This is an Accepted Manuscript of an article published by Taylor & Francis in  
Memory on [date of publication], available online:**

<http://www.tandfonline.com/10.1080/09658211.2017.1414848>.

# Temporal grouping effects in musical short-term memory

Recent theoretical accounts of verbal and visuo-spatial short-term memory (STM) have proposed the existence of domain-general mechanisms for the maintenance of serial order information. These accounts are based on the observation of similar behavioural effects across several modalities, such as temporal grouping effects. Across two experiments, the present study aimed at extending these findings, by exploring a STM modality that has received little interest so far, STM for musical information. Given its inherent rhythmic, temporal and serial organisation, the musical domain is of interest for investigating serial order STM processes such as temporal grouping. In Experiment 1, the data did not allow to determine the presence or the absence of temporal grouping effects. In Experiment 2, we observed that temporal grouping of tone sequences during encoding improves short-term recognition for serially presented probe tones. Furthermore, the serial position curves included micro-primacy and micro-recency effects, which are the hallmark characteristic of temporal grouping. Our results suggest that the encoding of serial order information in musical STM may be supported by temporal positional coding mechanisms similar to those reported in the verbal domain.

Keywords: serial order, working memory, music and language, domain-general, rhythm

## Introduction

In his influential paper, Lashley (1951) considered the capacity of the brain to process serial order information as one of the most complex types of human behaviour. This ability is involved in a wide range of human activities, such as speech perception (Grossberg, 2003) and production (Dell, 1986; Dell, Burger, & Svec, 1997; MacKay, 1970) and musical performance (Mathias, Pfordresher, & Palmer, 2015; Palmer & Pfordresher, 2003; Pfordresher, Palmer, & Jungers, 2007). Ironically, while Lashley (1951) illustrated the *problem of serial order in behaviour* with the case of serial order constraints that musicians have to deal with during musical performance, only little attention has been paid to serial order processing in the musical domain (but see

Pfordresher et al., 2007). Serial ordering capacities have been extensively studied in the verbal domain, particularly in the context of verbal short-term memory (STM) tasks. In the verbal STM domain, a variable that has been observed to have a major impact on serial order processing is the effect of temporal grouping. The manipulation of the temporal grouping of memoranda has been shown to lead to generally improved recall accuracy relative to ungrouped memoranda and to a specific shape of the serial position curve characterised by intra-group primacy and recency effects (Farrell & Lewandowsky, 2004; Hartley, Hurlstone, & Hitch, 2016; Henson, 1996, 1999; Hurlstone & Hitch, 2015; Ng & Maybery, 2002, 2005; Parmentier, Andrés, Elford, & Jones, 2006; Parmentier, Maybery, & Jones, 2004; Ryan, 1969a, 1969b). The present study aimed at furthering our understanding of serial order STM for auditory material, by investigating the impact of temporal grouping manipulations on the maintenance of musical sequence information.

In the verbal STM domain, temporal grouping effects occur when groups of stimuli are characterised by short between-stimulus temporal intervals and are separated by larger between-group temporal intervals. Two main effects can be observed. First, there is a temporal grouping advantage characterised by better recall accuracy for temporally grouped sequences relative to ungrouped sequences (Farrell & Lewandowsky, 2004; Frankish, 1985, 1989; Hartley et al., 2016; Henson, 1999; Maybery, Parmentier, & Jones, 2002; Ng & Maybery, 2002, 2005; Ryan, 1969a, 1969b). In addition to a recall advantage, temporal grouping also modifies the shape of the serial position curve that characterises recall performance. In the absence of temporal grouping manipulations, the serial position curve is bow-shaped with better recall accuracy for items at the beginning and the end of the list. In temporal grouping conditions, the serial position curve is characterised by the appearance of multiple,

within-group micro-primacy and micro-recency effects (Frankish, 1989; Hartley et al., 2016; Hitch, Burgess, Towse, & Culpin, 1996; Ng & Maybery, 2002; Ryan, 1969a, 1969b). Third, temporal grouping manipulations also have a critical influence on the pattern of transposition errors during recall. In verbal STM tasks, serial ordering errors are usually constrained by a locality principle (Henson, 1996), with transposition errors being more frequent for adjacent serial position exchanges than for more distant serial position exchanges. For temporally grouped memoranda, an increase of more distant, between-group transpositions is observed, with migrating items keeping their initial within-group position (Farrell & Lelièvre, 2009; Farrell & Lewandowsky, 2004; Hartley et al., 2016; Henson, 1999; Ng & Maybery, 2002, 2005; Ryan, 1969a, 1969b); these errors are also known as interposition errors (Henson, 1996).

The pattern of effects induced by temporal grouping is of importance for theoretical models of serial order STM. While many models of STM acknowledge a separation between item representations and serial order representations (see, e.g., Brown, Preece, & Hulme, 2000; Burgess & Hitch, 2006; C. L. Lee & Estes, 1981; Majerus, 2013; Martin, Lesch, & Bartha, 1999; Oberauer, Lewandowsky, Farrell, Jarrold, & Greaves, 2012), the nature of the mechanisms underlying serial order representations remains an open question. A major account of serial order coding mechanisms is represented by context-based models (Brown et al., 2000; Burgess & Hitch, 2006; Farrell & Lewandowsky, 2002; Hartley et al., 2016; Henson, 1998; Lewandowsky & Farrell, 2008; Page & Norris, 1998). In these models, the serial order of items is represented by associating the items to the different states of a contextual signal varying during the course of list presentation. For some authors (Farrell & Lewandowsky, 2002; Page & Norris, 1998), these contextual signals are unidimensional, ordinal markers associating successive items with decreasing levels of

activation during encoding, from the start to the end of a sequence. Others advocate a positional theory (see Brown et al., 2000; Burgess & Hitch, 1999, 2006; Henson, 1998; Lewandowsky & Farrell, 2008), in which serial order is represented through associations between item representations and independent positional markers.

Temporal grouping effects are a critical phenomenon against which the validity of proposed models of serial order STM can be tested. Indeed, context-based models relying on unidimensional ordinal signals have difficulties to accommodate temporal grouping effects and particularly the occurrence of interposition errors—that is, between-group transposition errors where items keep the same relative position as in their original group (Henson, 1996). At the same time, context-based models relying on positional markers suggest that order information can be encoded by signals with different levels of resolution, with one signal level tracking the position of item occurrences inside groups and another level encoding position of items or groups at the whole-list level (Brown et al., 2000; Burgess & Hitch, 1999, 2006; Hartley et al., 2016; Henson, 1998; Lewandowsky & Farrell, 2008). Multi-oscillator components have been proposed to track the temporal structure of a sequence at different time scales (Hartley et al., 2016).

It has also been suggested that the same type of encoding processes may support serial order coding across several STM modalities. In a recent review, Hurlstone, Hitch, and Baddeley (2014) highlighted the existence of benchmark serial ordering phenomena in STM for verbal, visual and spatial information, suggesting that similar ordering mechanisms operate for different modalities. Regarding temporal grouping, several studies demonstrated a memory advantage for temporally grouped sequences in verbal and spatial STM tasks (Hurlstone & Hitch, 2015; Parmentier et al., 2006; Parmentier et al., 2004), suggesting that temporal positional information contributes similarly to the

encoding of serial order in the two domains. Moreover, in a recent paper Vandierendonck (2016) showed cross-modal interference between serial recall of verbal and visuo-spatial information, suggesting that the encoding of serial order information in STM is modality independent.

At the same time, evidence in favour of cross-modal serial order coding mechanisms remains scarce, because few studies have explored serial order coding mechanism beyond the verbal and visuo-spatial modalities. In the present study, we further examined the hypothesis of domain-general serial order coding processes by studying temporal grouping effects in musical STM. Williamson, Baddeley, and Hitch (2010) proposed that musical and verbal STM systems involve different domain-specific representational stores (see also Deutsch, 1970; Gorin & Majerus, submitted) while potentially sharing similar sequential refreshing mechanisms. The involvement of similar sequential mechanisms in verbal and musical STM has received further evidence in a recent study by Gorin, Kowialiewski, and Majerus (2016) showing that in verbal and musical STM tasks the maintenance of serial order information, but not item information, is similarly impacted by a temporally organised interfering task. These results were interpreted as reflecting the involvement of a similar timing-based contextual signal to represent order information in both domains.

One study observed effects similar to temporal grouping in musical STM (Deutsch, 1980). In her study, Deutsch investigated the influence of structures—either tonal or temporal—on recall of tone sequences. She required trained musicians to recall sequences of auditorily presented tones via a musical notation method. The author observed a substantial positive effect of temporal segmentation on serial recall accuracy. Deutsch also analysed inter-item dependencies—that is the probability that a correct item is directly followed by a second correct item—and observed that these

dependencies were lower for adjacent items forming a group boundary than for adjacent items inside a group. This was considered by the author as evidence that groups are retained and lost independently.

Nonetheless, there is evidence for interposition-like effects in musical production tasks. In an experiment that required adult pianists to perform musical sequences retrieved from memory, the initiation of a stronger metrical context led participants to transpose more frequently items at positions sharing the same metrical signature as the original position (Mathias et al., 2015). This is similar to the increase of interposition errors witnessed under temporal grouping conditions in verbal STM tasks (Hartley et al., 2016; Henson, 1996, 1999; Ng & Maybery, 2002, 2005; Ryan, 1969a). Mathias et al. (2015) proposed a multidimensional coding model of serial order in musical production: one dimension tracks the position of the items in the sequence and another one tracks their metrical signature. This is similar to the time-based, multi-oscillator models of serial order coding proposed in the verbal domain and mentioned earlier (e.g., Brown et al., 2000; Burgess & Hitch, 1999; Hartley et al., 2016).

Currently, the evidence that temporal grouping can increase performance in musical STM remains however indirect and has only been tested in musical experts. The aim of the present study was to provide direct evidence for temporal grouping effects on memory performance in musical STM tasks, using a hybrid recall-recognition musical STM task that can be performed by participants that have no advanced musical expertise. The reason for including participants with no advanced musical expertise was to allow for a better test of the generality of the effects under investigation. By conducting the same study in musical experts, results may be valid only for this population characterised by specialised and overlearned skills for processing and maintaining musical stimuli (Schulze, Mueller, & Koelsch, 2011; Schulze, Zysset,

Mueller, Friederici, & Koelsch, 2011). We therefore conducted a set of two experiments where groups of participants with no advanced musical expertise completed STM tasks requiring the maintenance of serial order information under conditions of temporal grouping or no grouping and by adapting a task that has been recently developed to study serial order STM in non-musician participants (Gorin et al., 2016).

## **Experiment 1**

When studying musical STM in populations with no advanced musical expertise, researchers often rely on list recognition paradigms due to the generally poor ability of musically untrained participants to accurately reproduce musical sequences (see Müllensiefen & Wiggins, 2011). At the same time, temporal grouping effects reported for verbal and visuo-spatial stimuli stem from experiments relying on a serial recall procedure, as these recall procedures are highly informative about the manner participants organise information during recall and allow for an analysis of transposition errors. For this reason, we used a mixed recall-recognition procedure, by taking advantage of a paradigm recently developed by Gorin et al. (2016). This paradigm requires participants to covertly recall the sequence of a previously presented melody, like in serial recall, and to make a serial order judgment about an isolated probe tone presented at a given serial position, like in recognition paradigms. We expected to observe, in the temporal grouping condition, an increase of recognition performance and the appearance of multiple micro-primacy and micro-recency effects as a function of the temporal groupings that were induced. We also added mismatching trial involving interposition-like displacements to assess whether interposition effects also characterise STM for grouped tone sequences.

## ***Method***

### *Participants*

Thirty-four participants took part in the experiment on a voluntary basis. Valid data were obtained for 31 participants; 2 participants had to be discarded due to technical problems during task administration and one participant had misinterpreted the task instruction as evidenced by very low task performance (recognition accuracy: .37). The final sample of 31 participants (20 women) had a mean age of 32.7 years ( $SD = 10.5$ ), with a low level of musical expertise ( $M = 1.3$  years of musical instrument or singing practice,  $SD = 3.3$ , range: 0-14 years), and a high educational level (all participants had completed their secondary school studies and 24 participants had graduated from university). All participants were asked about their hearing status; no participant reported having absolute pitch or hearing impairment, except for one participant who had mild tinnitus.

For this study, we recruited participants with no musical experience ( $n = 23$ ) or minimal musical experience ( $n = 8$ ). The participants with minimal musical experience had been engaged in musical activities for less than three years at the time of the study, and they exerted these activities less than three times per week (e.g., if a participant had started playing the piano four years before the experiment and was still playing at the moment of the study, the participant was not retained for the experiment). For participants who had practiced music in the past and were not playing anymore at the moment of the experiment, they had to have stopped musical practice for a number of years equal or higher to the years of musical practice, at the time of this experiment (e.g., if a participant had played the guitar for six years but had stopped four years before the moment of the experiment, this participant was not retained for the study). This ensured that only non-musicians or musicians with minimal current musical experience were

included in the study sample. Finally, participants completed a written informed consent before starting the experimental session; the experiment had been approved by the local ethics committee.

### *Materials*

The musical stimuli used in the present experiment were 14 tones with pitch ranging from 65 Hertz (C<sub>2</sub>) to 247 Hertz (B<sub>3</sub>) and following the steps of a diatonic C major scale. Anvil Studio 2001 (version 2011.09.06) was used to generate 300 ms MIDI tone stimuli with a piano timbre (with a rise and fall period of 10 ms) that were converted in .wav format. We also used a beat sound serving as a metronome during the presentation of the tone sequences. The beat sound (cross-stick drum timbre, 25 ms) was generated with Guitar Pro (version 6) and converted to a .wav file.

### *Design*

The experiment consisted in the presentation of 144 six-tone sequences spread into 2 blocks of 72 trials. One block contained ungrouped tone sequences presented at a regular pace while the other block was composed of sequences containing two three-tone temporal groups. All the tones composing a sequence were of different pitch, in order to avoid tone repetition effects (Deutsch, 1972, 1975). Also, in order to conform to the musical structures familiar to our participants, all the sequences followed Western-style musical patterns. This was confirmed by high Pearson correlations between the tone distribution of the sequences and the C major scale profile (Krumhansl & Kessler, 1982). This was done via an algorithm developed by Krumhansl and Schmuckler (cited in Krumhansl, 1990) and which correlates the distribution of pitch class occurrences of a given musical excerpt with the 12 major and the 12 minor key profiles of the Western musical system. The highest correlation is known as the

maximum key-profile correlation (MKC) and indicates the key that is most likely represented by a musical excerpt. We ensured that the absolute size of the intervals forming the sequences could not exceed five semitones.

The set of 144 unique sequences had a mean absolute interval size of 2.79 semitones ( $SD = .24$ ) and a mean MKC of .83 ( $SD = .02$ ) with the C major scale profile. These values were the same for the ungrouped and grouped sequence sets: ungrouped sequences (mean absolute interval size = 2.81 semitones,  $SD = .20$ ; mean MKC with the C major scale profile = .83,  $SD = .02$ ) and grouped sequences (mean absolute interval size = 2.78 semitones,  $SD = .28$ ; mean MKC with the C major scale profile = .83,  $SD = .02$ ).

### *Procedure*

The tone sequences were presented to participants at a comfortable sound level through headphones connected to a portable workstation. The sequence lasted for 2100 ms, corresponding to the delay between the occurrence of the first tone and the end of the last tone. For the ungrouped condition, the tones were presented with a regular inter-onset-interval (IOI) of 360 ms. For the grouped condition, the tones within the groups were presented with a regular IOI of 300 ms and we induced temporal grouping by adding a silent period of 300 ms between the third and the fourth tone (see Figure 1 for a graphical representation of the task design).

Each trial began with the presentation of a red circle appearing on the centre of the screen for 1100 ms. Each tone occurrence inside a sequence was played synchronously with the beat sound serving as a metronome (see Figure 1 for more details). The presentation of the target sequences was followed by a 3 maintenance phase. Next, a blue circle appeared on the centre of the screen for 1100 ms, indicating the beginning of the recognition phase. During the recognition phase, participants heard

again the sequence of beat sounds which had the same temporal organisation as in the target sequence and a probe tone was played at one of the six serial positions; the probe was always a tone that had been presented in the target sequence. The participants were required to covertly recall the target sequence in time with the beat sequence and to make a same/different judgment relative to the position of the probe tone in the target sequence by pressing one of the six response buttons representing different levels of confidence (1 = *very sure same*, 2 = *sure same*, 3 = *same*, 4 = *different*, 5 = *sure different*, 6 = *very sure different*), as used in previous experiments on musical STM (e.g., Dowling & Tillmann, 2014; Tillmann et al., 2013). For more details about our motivation to use levels of confidence, see the *Data and statistical analyses* section.

For each block condition, three types of trials probing equally the six serial positions were presented. For one third of trials, the matching probe tone was played at the correct serial position relative to its position in the target sequence. For another third of trials, the probe tone was displaced to an adjacent serial position (adjacent transposition) relative to its initial position in the target sequence. The final third of trials included mismatching probe tones displaced by three serial positions (distant transposition). Note that by including mismatching trials with tones moved by three serial positions relative to their initial position, we generated interposition movements (Henson, 1999), i.e. between-group transpositions of items preserving their initial within-group positions. This choice of using two types of mismatching trials was motivated by evidence from the verbal STM domain demonstrating that temporal grouping decreases the incidence of adjacent transposition errors but increases the occurrence of interposition errors (e.g., Hartley et al., 2016; Henson, 1996, 1999; Ng & Maybery, 2002, 2005). The probe tone and the tones of the target sequence preceding and following the probed position could not differ by more than five semitones.

The two blocks were presented in a fixed order with the ungrouped condition being presented first, in line with previous studies (Farrell & Lewandowsky, 2004; Henson, 1999). This was done in order to avoid that participants confronted first with the grouped condition subsequently implemented grouping strategies for the ungrouped condition. The presentation of the trials inside the blocks was randomised for each participant. Each block started with two practice trials. The experiment was programmed and presented using Opensesame software (version 3.0.1, Mathot, Schreij, & Theeuwes, 2012).

Participants completed a demographic questionnaire initially and were asked at the end of the experiment to report the strategies they had used during the task by filling out a strategy-related questionnaire. Participants could indicate one of the following strategies: 1) passive listening of the target melodies without any kind of rehearsal, 2) auditory rehearsal of the melodies, 3) rehearsal of the melodies based on verbal relabelling of the tone names, 4) use of visual imagery (i.e., visual representation of the shape of the melody where tone sequences are represented as visual curves, the curves going up and down as a function of the pitch changes of the melody), 5) use of motor/gestural codes to rehearse the melody (i.e., up and down finger or head movements as a function of the pitch changes of the melody), 6) use of a grouping strategy (e.g., creating subgroups of notes to rehearse the melody), 7) no strategy. The participants could report more than one strategy.

#### *Data and statistical analysis*

The responses for the different types of trials were analysed using the receiver operating characteristic (ROC) and transformed into areas under the ROC curve (Swets, 1973). Each curve was derived from plotting the responses to 24 matching probes (same trials) *versus* 24 mismatching probes (different trials) according to their associated level of

confidence. We next determined the area under the ROC curve and used it as a measure of memory recognition performance, with .50 representing the chance threshold.

The choice of using ROC measures was motivated by the fact that in musical recognition tasks, participants with no advanced musical expertise may express some uncertainty in their response and may yet respond above chance level (see Dowling, Kwak, & Andrews, 1995). Also, contrary to other measures of sensitivity such as the  $d'$  score, the area under the ROC curve is not affected by response bias such as the tendency to respond more or less frequently “same” or “different” (see Dowling et al., 1995; Verde, Macmillan, & Rotello, 2006). ROC curves therefore deal more efficiently with response bias especially when there is some degree of uncertainty associated with the responses. This type of measure, taking explicitly into account the uncertainty that may be associated with the responses, is very frequently used in the musical STM literature (e.g., Dewar, Cuddy, & Mewhort, 1977; Dowling & Tillmann, 2014; Halpern, Bartlett, & Dowling, 1995; Tillmann et al., 2013).

In addition to ROC measures we analysed the rates of correct recognitions for same and different trials using specific analyses (see, e.g., Dowling, Magner, & Tillmann, 2016; Dowling, Tillman, & Ayers, 2002; Dowling & Tillmann, 2014; Halpern et al., 1995, for other studies using the same type of analysis strategy). This was motivated by previous studies on musical STM showing that experimental manipulations can lead to changes in either same or different correct recognition rates while leaving ROC scores unaffected (see, e.g., Dowling et al., 2002; Dowling & Tillmann, 2014).

All the statistical analyses were based on a Bayesian approach (Rouder, Morey, Speckman, & Province, 2012; Rouder, Speckman, Sun, Morey, & Iverson, 2009), conducted with the open source software JASP (version 0.8.0.0, JASP Team, 2016).

This choice was motivated by criticisms relative to fundamental problems when using  $p$ -values from frequentist statistical methods to make statistical inference, as reported in Wagenmakers (2007), but see also Wagenmakers, Lee, Lodewyckx, and Iverson (2008) for a more detailed report relating to the problems and advantages of using frequentist and Bayesian inferences, respectively. Some critical advantages of using Bayesian inferences are that they allow quantifying statistical evidence, are able to provide evidence in favour of the null hypothesis and allow applying a model comparison and selection method indicating which model predicts the data best. We used Bayes factor (BF) when reporting results. The BF is the resulting statistic of a model comparison analysis that provides the extent to which, after looking at the data, the relative odds between two models has changed; the BF can therefore be interpreted as a measure of statistical evidence (Morey, 2015). The reporting of  $BF_{01}$  and its associated value indicates evidence in favour of the null hypothesis relative to the alternative hypothesis given the data. Inversely,  $BF_{10}$  and its associated value indicate evidence in favour of the alternative hypothesis relative to the null hypothesis given the data. Finally, we used the classification proposed by M. D. Lee and Wagenmakers (2014) to interpret the strength of evidence associated with the BFs reported, where a BF lesser than three was considered as anecdotal evidence for the model under investigation, between 3 and 10 as moderate evidence, between 10 and 30 as strong evidence, between 30 and 100 as very strong evidence and higher than 100 as decisive evidence.

## ***Results***

### *Recognition performance*

The first analysis was conducted on areas under the ROC curve with a  $2 \times 2$  Bayesian repeated measures ANOVA containing a two-level temporal grouping factor

(ungrouped *versus* grouped) and a two-level transposition distance factor (adjacent *versus* distant). The results showed that (see Figure 2), compared to the null model containing only the subject variable as nuisance factor, the model with the highest BF was the model with both grouping and transposition distance effects ( $BF_{10} = 3.05$ ), followed by the model with only the effect of grouping ( $BF_{10} = 2.36$ ). The direct comparison of these two models provided only anecdotal evidence in favour of the model including the two main effects ( $BF = 1.29$ ). Given the insensitivity of the data to distinguish the two models, we further looked at the analysis of specific effects associated with each variable. This analysis is based on a model averaging method where the amount of evidence for a specific effect is determined by averaging evidences across all the models containing the effect of interest (relative to the null model). Evidence for the inclusion of the temporal grouping effect, the transposition distance effect and the interaction between the two factors remained very weak with  $BF_{\text{Inclusion}}$  values of 2.00, 0.97 and 0.41, respectively.

We next analysed the effect of temporal grouping as a function of the serial position that was probed, separately for same and different trials. In order to obtain reliable response estimates per serial position, the six possible response types ranging from ‘sure same’ to ‘sure different’ were aggregated to a binary (same/different) judgment. In other words, responses with the options ‘very surely the same’, ‘surely the same’ or ‘the same’ were aggregated into a unique ‘same’ response category, while responses with the ‘very surely different’, ‘surely different’ or ‘different’ response options were aggregated into a unique ‘different’ response category. This allowed us to determine response accuracy for same and different trials. A first Bayesian repeated measures ANOVA assessed the effects of grouping (2 levels) and serial position (6 levels) on recognition accuracy scores for same trials. We observed that the model

explaining the data best was the model containing only the effect of position ( $BF_{10} = 8.16E+9$ ), followed by the model containing the two main effects of position and of grouping ( $BF_{10} = 4.38E+9$ ) (see Figure 3). The direct comparison between these two models provided only anecdotal evidence in favour of a model containing only the position factor ( $BF = 1.86$ ). Since the analysis did not allow distinguishing clearly between the two models, we conducted an analysis of specific effects. These results definitively support the presence of an effect of position ( $BF_{Inclusion} = 7.12E+9$ ), but the evidence in favour of an effect of grouping remained very low ( $BF_{Inclusion} = 0.55$ ).

The same analysis was conducted on response accuracy for different trials involving adjacent transpositions (see Figure 4). This analysis revealed that the null model containing only the participant factor was favoured over all other models (evidence against the presence of a grouping effect:  $BF_{01} = 7.29$ ; evidence against a position effect:  $BF_{01} = 62.64$ ; evidence against both effects:  $BF_{01} = 445.19$ ; evidence against the full model:  $BF_{01} = 4.05$ ).

Finally, recognition accuracy scores for different trials involving distant transpositions were subjected to the same  $2 \times 6$  Bayesian repeated measures ANOVA (see Figure 5) leading to identical results. The null model was favoured over all alternative models (evidence against the grouping effect:  $BF_{01} = 6.68$ ; evidence against the position effect:  $BF_{01} = 4.05$ ; evidence against the grouping and position effects:  $BF_{01} = 26.43$ ; evidence against the full model:  $BF_{01} = 15.13$ ).

### *Analysis of strategies*

Table 1 displays for each condition the distribution of the strategies reported by the participants. Most of the participants reported to use visual-based or motor-based mental imagery strategies (ungrouped: 56% of the participants; grouped: 51% of the participants). The second most frequently used strategy was auditory rehearsal

(ungrouped: 26% of the participants; grouped: 19% of the participants), followed by subjective grouping strategies (ungrouped: 7% of the participants; grouped: 19% of the participants). As one can see, very few participants used explicit grouping strategies, this even for the grouped condition.

### ***Discussion***

The results of Experiment 1 did not provide robust evidence for the presence of grouping effects in musical STM. The analysis of ROC scores showed that, compared to the null model, models including the grouping effect were accompanied by BF values ranging between 2 and 3. When looking at Figure 3 depicting recognition accuracy for same trials as a function of serial position, we observed better recognition performance for position 3, providing decisive evidence for a mini-recency effect in the first group ( $BF_{10} = 274.92$ ). However, for other serial positions, evidence rather favoured the null model ( $BF_{01}$  ranging from 3.36 to 5.17).

Overall, although some elements of the results are in favour of a grouping effect, the evidence is inconsistent. However, a number of aspects of the experimental design of the task used in Experiment 1 may have influenced the results. The grouped condition was always presented in the second part of the experiment, which might have reduced the benefit of potential temporal grouping effects, with participants starting to show mental fatigue in the latter half of the session when the grouped lists were administered, the experiment lasting more than 45 minutes. An analysis of strategies also indicated that participants rarely used grouping strategies to complete the task, but instead showed a tendency to form visual- or motor-based representations of up/down pitch variations of the memory sequence.

Second, participants reported difficulties with the six-level response procedure which was likely to require STM resources by itself, given the need to temporarily

activate six different response options and to relate these options to the representations held in STM. Third, results from Experiment 1 showed that recognition rates for different trials are very close to .50 (ungrouped adjacent transpositions: .49; ungrouped distant transpositions: .51; grouped adjacent transpositions: .47; grouped distant transpositions: .49), which amounts to chance level recognition/rejection rates. This aspect of the results suggests that participants could not reliably reject different trials. In order to increase the sensitivity of the task, while keeping task length at a reasonable level and avoid fatigue effects, we decided to retain only one type of negative trials for Experiment 2. We kept the most informative negative trials, i.e. those involving adjacent tone displacements only, which are known to yield the highest amount of errors and thus require the most precise memory representation of serial order information.

Also, by removing distant different trials in Experiment 2 we could not assess the occurrence of interposition errors anymore in the grouping condition. It should be noted here that the phenomenon of increased interposition errors under grouping conditions has been observed mainly in experiments using recall procedures and is a relatively rare phenomenon. For example, Hartley et al. (2016) recently showed that in serial recall tasks for digit lists, the proportion of responses corresponding to interposition errors was 11% and 14% for ungrouped and grouped lists, respectively. It would therefore be very difficult to reliably track interposition errors and their small increase in grouping conditions by using a recognition paradigm for which the different trials would need to precisely and reliably predict when and where a participant would make interposition errors for the stimuli held in memory. At the same time, a very large number of interposition-like mismatching trials would be required to optimally detect possible interposition memory errors in an experimental setting involving recognition,

leading to a drastic increase in task duration. A recognition procedure is thus not optimally suited for probing the occurrence of interposition errors. Importantly, the removal of distant different trials in Experiment 2 does not alter the principal aim of our study, which consists in showing that temporal grouping will lead to an advantage in overall recognition performance and in specific recognition patterns as a function of the serial position of the stimuli being tested.

## **Experiment 2**

The second experiment aimed at addressing the difficulties identified for Experiment 1 and to assess evidence for an effect of temporal grouping in musical STM in a more robust manner by focusing only on the overall effect of temporal grouping on memory performance. We simplified the response scale, by proposing four instead of six response choices. We also counterbalanced the order of presentation of the blocks between participants. Finally, we diminished the length of the experiment by reducing the number of mismatching trials and by keeping only mismatching trials with adjacent tone displacements.

## ***Method***

### *Participants*

Thirty-two participants took part in Experiment 2 on a voluntary basis and were selected following the same criteria as used in Experiment 1. One participant was excluded from the sample due to poor testing condition (i.e. disturbance during the testing session).

The final sample was composed of 31 participants with a mean age of 28.4 years (SD = 8.8, seven women) and a low level of musical experience, as reflected by the average number of years of instrumental or singing experience reported by the participants (M =

0.7 years, SD = 2.7, range: 0-11 years). Nine participants had graduated from university; the other participants had achieved their secondary studies. All participants reported to have satisfactory hearing abilities; two participants reported to experience episodes of mild tinnitus. Finally, all participants received and completed a written informed consent before the beginning of the testing session and the experiment had been approved by the local ethics committee.

### *Materials*

The material used in Experiment 2 was the same as the one used in Experiment 1, with the exception that there were no mismatching trials with distant displacements.

### *Design*

Since we discarded distant mismatching trials, Experiment 2 consisted in 96 trials composed of six-tone melodic sequences that were presented into two separated blocks of 48 trials.

### *Procedure*

The procedure was the same as in Experiment 1, but, critically, we simplified the response scale by limiting the response options to four (1 = *sure same*, 2 = *probably same*, 3 = *probably different*, 4 = *sure different*). Furthermore, the order of presentation of the blocks was counterbalanced across participants.

### *Data and statistical analysis*

Two area scores, each based on 48 data points (24 same trials and 24 different trials), were computed. One area score reflected discrimination between matching and mismatching trials in the grouped condition and the other one in the ungrouped

condition.

## **Results**

### *Recognition performance*

As in Experiment 1, the first analysis focused on the areas under the ROC curve (see Figure 6). Bayesian paired samples *t*-test comparing the area scores between the temporally grouped ( $M = .60$ ,  $SD = .08$ ) and ungrouped ( $M = .58$ ,  $SD = .09$ ) conditions did not reveal any evidence for a grouping effect ( $BF_{10} = 0.44$ ).

When considering recognition performance separately for same and different trials, a different picture of results emerged (see Figure 7). A  $2 \times 6$  Bayesian repeated measures ANOVA on recognition accuracy scores for same trials (after reducing the four-choice responses to a binary response score) revealed that the model with the highest BF was the full model ( $BF_{10} = 5.17E+16$ ), followed by the model with the two main effects without the interaction ( $BF_{10} = 2.28E+16$ ). The direct comparison of these two models showed that the model with the main effects and their interaction was favoured over the other model by a factor of 2.27, which represents anecdotal evidence in favour of the full model. An analysis of specific effects showed decisive evidence for the effect of serial position ( $BF_{Inclusion} = +\infty$ ), while the effect of temporal grouping and the interaction were associated with strong ( $BF_{Inclusion} = 23.18$ ) and moderate ( $BF_{Inclusion} = 8.28$ ) evidence, respectively.

The interaction was explored with Bayesian paired samples *t*-tests. As shown in Figure 7, the analysis provided strong evidence in favour of an effect of temporal grouping for serial position 2 ( $BF_{10} = 24.35$ ) and 3 ( $BF_{10} = 22.96$ ). For all the remaining positions, the analysis provided evidence in favour of the *absence* of a temporal grouping effect (position 1:  $BF_{01} = 4.54$ ; position 4:  $BF_{01} = 2.75$ , position 5:  $BF_{01} =$

5.10, position 6:  $BF_{01} = 4.77$ ). The increase of recognition accuracy for the third position representing the last item of the first group indicates the presence of a micro-recency effect; in Figure 7, the shape of the serial position curve also suggests a micro-primacy effect given the slightly higher recognition accuracy for position 4 in the grouped *versus* ungrouped conditions, and corresponding to the first item of the second group. This particular scalloped shape of the serial position curve is typical of temporal grouping effects (e.g., Hitch et al., 1996).

When we conducted the same analysis on recognition accuracy scores for different trials, we obtained the same null results as observed in Experiment 1 (see Figure 8). The null model was favoured over all alternative models (evidence against a grouping effect:  $BF_{01} = 8.93$ ; evidence against a serial position effect:  $BF_{01} = 36.37$ ; evidence against the two main effects:  $BF_{01} = 314.62$ ; evidence against the full model:  $BF_{01} = 475.66$ ).

#### *Analysis of strategies*

The pattern of strategies used during the task appears to be fairly similar to the pattern observed in Experiment 1. Table 2 displays for each grouping condition the distribution of the strategies reported by the participant. As in Experiment 1 participants mainly relied on visual-based or motor-based mental imagery of up/down pitch variations of the memory sequence in both grouping conditions (ungrouped: 60% of the participants; grouped: 57% of the participants), followed by rehearsal strategies (ungrouped: 30% of the participants; grouped: 33% of the participants), while the use of subjective, explicit grouping strategies remained very low (3% of the participants for both ungrouped and grouped conditions).

## ***Discussion***

Experiment 2 addressed some of the possible methodological weaknesses we had identified for Experiment 1. By using a simpler response scale, by reducing experiment duration and by counterbalancing the order of the grouping conditions, we observed in Experiment 2 an effect of grouping, although it was limited to recognition accuracy for some trials. Furthermore, the serial position curve took a scalloped shape, indicating micro-recency and micro-primacy effects for grouped tone sequences.

## **General discussion**

The two experiments reported here investigated how serial order information is represented in musical STM, by focusing on the study of temporal grouping effects. We had participants with no advanced musical expertise perform a serial order recognition task in which we manipulated the temporal grouping pattern of the tone sequences. In Experiment 1, the results obtained were partially in favour of a temporal grouping advantage on musical STM when using ROC sensitivity scores. An analysis of the grouping effect on recognition accuracy per serial position revealed better recognition accuracy for position 3 and indicated a mini-recency effect in the first group. However, there was no evidence for an advantage of grouping on recognition accuracy for the other serial positions. The results of Experiment 1 could also have been biased by the complexity of the 6-level response scale used for determining the ROC scores and by the long task duration. Experiment 2, using a simpler response scale and a shorter task duration, provided evidence for an effect of grouping, with a higher detection rate of items occurring in correct serial position in the grouped condition, and the appearance of micro-recency and micro-primacy effects for grouped tone sequences.

This study provides partial evidence in favour of temporal grouping effects in STM for musical stimuli in participants with no advanced musical experience. The results mirror those previously reported by Deutsch (1980) in musical experts. Our results are also in line with key temporal grouping effects reported in other STM modalities, and more particularly with the appearance of micro-recency and micro-primacy effects for grouped stimulus sequences (Farrell & Lewandowsky, 2004; Hartley et al., 2016; Henson, 1996, 1999; Hurlstone & Hitch, 2015; Ng & Maybery, 2002, 2005; Parmentier et al., 2006; Parmentier et al., 2004; Ryan, 1969a, 1969b). We observed that for ungrouped sequences the serial position curve was characterised by start-of-list primacy and end-of-list recency effects, while for temporally grouped sequences there were additional mini-recency and mini-primacy effects, for the first and second group, respectively.

At the same time, it is important to note that our results are strongly dependent upon the measures that are used as evidence in favour of a grouping effect was observed only for the detection rate of matching probe stimuli. What could be the reasons for these task-specific effects observed in Experiment 2? It is important to keep in mind that the ROC measures integrate information about both same and different trials. Hence, given the absence of temporal grouping effects for different trials, the absence of temporal grouping effects on ROC measures is not surprising. It is precisely the discrepancy of results for same and different trials that needs further consideration. First, it should be noted that this type of discrepancy is not uncommon in the musical STM literature (Dowling et al., 2001, see Experiment 3; Dowling & Tillmann, 2014, see Experiment 4). For example, in a same/different musical melody recognition experiment, Dowling and Tillmann (2014) compared recognition after short and long delays and observed that, in comparison to the short delay, the long delay led to

diminished recognition accuracy for same trials but increased accuracy for different trials, with no effect on ROC scores. Dowling and Tillmann (2014) interpreted these results as a shift in terms of the response criterion in the long delay condition, with participants responding less frequently “same”. However, the results of the present study do not seem to reflect such a shift in response criterion given that the positive effect of grouping observed for same trial recognition accuracy was not accompanied by a negative effect on different trials; rather, different trials did not seem to be sensitive at all to the manipulation of temporal grouping.

Another reason for the specific effect on same trials could be related to factors influencing the strength of the memory trace. The strength of memory traces has been studied mostly in episodic memory experiments, in which, during the recognition phase, previously learnt items and novel items are presented for old-new recognition judgment. Rotello and Macmillan (2008) suggested that, in the absence of feedback about recognition accuracy, memory strength influences overall recognition accuracy while letting the decision criterion unaffected. This is illustrated in a study by Verde and Rotello (2007) comparing recognition performance for strong-old, weak-old and novel items; strong items were those presented more frequently or for a longer period of time during the learning phase and were supposed to lead to stronger memory traces. Verde and Rotello (2007) showed that, for recognition tests containing either strong old/new or weak old/new items, recognition was better for strong old items than for weak old items, while the rejection of novel items did not differ between the two recognition tests. By transposing these results to the present study, we can compare the old/novel distinction of the Verde and Rotello (2007) procedure to the same/different distinction in our STM task. The specific effect of grouping observed for same trials in the present study could reflect the fact that grouping increases the memory strength, and hence

leads to a higher detection rate of matching (old) item-to-position associations, while leaving rejection rates for mismatching (novel) trials unaffected. Temporal grouping may strengthen item-to-position associations, via binding items to different serial position dimensions at the same time, one keeping track of the groupings, and another one for the whole list, as suggested by Burgess and Hitch (2006) and Hurlstone et al. (2014).

Another aspect that needs to be considered is that serial order STM is most typically assessed using recall rather than recognition procedures as recall procedures provide full information about recall accuracy for all serial positions on each single trial, and hence represent a more sensitive measure of serial order STM than recognition procedures. At the same time, in a comparison between tasks requiring either serial recall or serial order recognition, it should be noted that Oberauer (2003) found similar serial position effects when comparing recall and recognition STM procedures, and this for a recognition procedure where serial position effects were established based on same trials only. The author proposed that similar mechanisms underlie serial position effects such as primacy and recency in recall and recognition tasks (see also Cowan, Saults, Elliott, & Moreno, 2002). His results suggest that the assessment of serial order STM performance based on same trials in a recognition paradigm represents a valid alternative to full recall STM procedures.

At a more general and theoretical level, the presence in our study of temporal grouping effects similar to those witnessed in other serial order STM modalities indicates that the theoretical models of serial order STM developed in the verbal domain may also apply to the musical domain. Hurlstone et al. (2014) proposed that context-based models of verbal STM for serial order relying on multidimensional positional coding mechanisms provide the most valid account of temporal grouping effects so far

(e.g., Brown et al., 2000; Burgess & Hitch, 1999; Henson, 1998; Lewandowsky & Farrell, 2008), as opposed to models representing serial order at only a single dimension, as in the primacy model of Page and Norris (1998). Indeed, by using multiple levels to represent serial order information, context-based models representing serial order through item-to-position associations at multidimensional levels are able to reliably account for temporal grouping effects (i.e., the appearance of micro-recency and micro-primacy effects as well as of interposition errors) but also for the main primacy and recency effects of the entire memory list. Given the temporal grouping effects observed for similar trials in Experiment 2, it could be considered that similar multidimensional position marking mechanisms are involved in musical STM for serial order. This view is also in line with a model of musical production where serial order is coded according to the position of tones in the sequence but also according to their metrical hierarchical level (Mathias et al., 2015).

At the same time, we need to acknowledge some limitations of our study. While our data are in line with models relying on positional markers to represent serial order in STM, the paradigm used in the present study was not designed to specifically address the question of the nature of serial ordering errors—such as the occurrence of interposition errors—as already mentioned in the discussion of Experiment 1. Therefore, the results that we obtained urge us to remain cautious regarding the positional nature of serial order representations in musical STM. Indeed, in the verbal domain of STM, positional theories of serial order are based on the fact that temporal grouping increases recall accuracy, modifies the serial position curve and leads to specific types of transposition errors such as interpositions (Henson, 1996). The increase of interposition errors in temporal grouping conditions has been considered as strong empirical support for positional theories of serial order in verbal STM. To the

opposite, in the visuo-spatial domain of STM, temporal grouping does not increase interposition errors, and may even lead to a decrease of these errors (Hurlstone & Hitch, 2015; Parmentier et al., 2006). According to Hurlstone and Hitch (2015, in press), the presence of temporal grouping effects but not interposition errors in spatial and visual STM tasks may be accommodated by positional mechanisms coding serial order information differently than in the verbal domain. Indeed, the authors proposed that, as in the verbal domain, groups in the visuo-spatial modality are coded depending on their position in the sequence. However, items in the visuo-spatial modality are coded as a function of their position in the whole sequence, while items in the verbal domain may be coded depending on their position inside groups. Further studies are necessary to determine the extent to which the effect of temporal grouping in musical STM is limited to an overall memory advantage—as observed in the visuo-spatial modality—or whether it is also characterised by interposition errors as in the verbal modality. Interposition errors have indeed been observed in musical production tasks although these studies did not directly investigate musical STM (see Mathias et al., 2015).

As our results provide only partial evidence in favour of positional coding mechanisms for serial order in musical STM, additional studies are required to further investigate the presence of temporal grouping effects and their impact on recall errors in the musical domain. Recognition procedures are not the best suited to study and compare serial order errors as by definition, recognition procedures need to ‘guess’ the serial position errors that a participant would make, and need a very large number of trials to probe all the different error types that are theoretically possible. For probing serial ordering errors in musical STM, future studies need to use recall procedures which are more sensitive and efficient for studying this question as any type of serial ordering error can occur on any trial. However, it should be noted here that it may be

difficult to achieve reliable musical recall output in non-musician participants not trained to produce singing responses (Pfordresher & Brown, 2007; Pfordresher, Brown, Meier, Belyk, & Liotti, 2010) when using a serial recall procedure. Therefore, when using singing responses, it could be difficult to distinguish errors due to the participants' lack of efficient sensori-motor mapping skills for musical output from memory-related errors, such as serial order transpositions.

To conclude, this study provides partial evidence for the presence of temporal grouping effects in a musical STM task in participants with no advanced musical expertise. The results point in favour of the involvement of positional serial order mechanisms, as also witnessed in other STM modalities. At the same time, future studies need to address the precise nature of positional representations in musical STM.

#### Acknowledgement

This work was supported by the Fund for Scientific Research–FNRS under grant FRESH to the Simon Gorin; Belgian Science Policy Office (BELSPO) under grant PAI-IUAPP7/11 to Simon Gorin and Steve Majerus; University of Liège under grant ARC12/17/01REST to the Steve Majerus. The authors report no conflicts of interest; the funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

#### References

- Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, *107*(1), 127-181. doi: 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. doi: 10.1037/0033-295x.106.3.551
- Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*(4), 627-652. doi: 10.1016/j.jml.2006.08.005
- Cowan, N., Saults, J. S., Elliott, E. M., & Moreno, M. V. (2002). Deconfounding Serial Recall. *Journal of Memory and Language*, *46*(1), 153-177. doi: 10.1006/jmla.2001.2805
- Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Review*, *93*(3), 283-321. doi: 10.1037/0033-295X.93.3.283
- Dell, G. S., Burger, L. K., & Svec, W. R. (1997). Language Production and Serial Order: A Functional Analysis and a Model. *Psychological Review*, *104*(1), 123-147. doi: 10.1037/0033-295X.104.1.123

- Deutsch, D. (1970). Tones and numbers: specificity of interference in immediate memory. *Science*, *168*(3939), 1604-1605. doi: 10.1126/science.168.3939.1604
- Deutsch, D. (1972). Effect of repetition of standard and of comparison tones on recognition memory for pitch. *Journal of Experimental Psychology*, *93*(1), 156-162. doi: 10.1037/h0032496
- Deutsch, D. (1975). Facilitation by repetition in recognition memory for tonal pitch. *Memory and Cognition*, *3*(3), 263-266. doi: 10.3758/BF03212909
- Deutsch, D. (1980). The processing of structured and unstructured tonal sequences. *Perception and Psychophysics*, *28*(5), 381-389. doi: 10.3758/BF03204881
- Dewar, K. M., Cuddy, L. L., & Mewhort, D. J. (1977). Recognition memory for single tones with and without context. *Journal of Experimental Psychology: Human Learning and Memory*, *3*(1), 60-67. doi: 10.1037/0278-7393.3.1.60
- Dowling, W. J., Kwak, S., & Andrews, M. W. (1995). The time course of recognition of novel melodies. *Perception and Psychophysics*, *57*(2), 136-149. doi: 10.3758/BF03206500
- Dowling, W. J., Magner, H., & Tillmann, B. (2016). Memory improvement with wide-awake listeners and with nonclassical guitar music. *Psychomusicology: Music, Mind, and Brain*, *26*(1), 26-34. doi: 10.1037/pmu0000106
- Dowling, W. J., Tillman, B., & Ayers, D. F. (2002). Memory and the Experience of Hearing Music. *Music Perception: An Interdisciplinary Journal*, *19*(2), 249-276. doi: 10.1525/mp.2001.19.2.249
- Dowling, W. J., & Tillmann, B. (2014). Memory Improvement While Hearing Music. *Music Perception: An Interdisciplinary Journal*, *32*(1), 11-32. doi: 10.1525/mp.2014.32.1.11
- Farrell, S., & Lelièvre, A. (2009). End anchoring in short-term order memory. *Journal of Memory and Language*, *60*(1), 209-227. doi: 10.1016/j.jml.2008.09.004
- Farrell, S., & Lewandowsky, S. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin Review*, *9*(1), 59-79. doi: 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. doi: 10.1016/j.jml.2004.03.007
- Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200-209. doi: 10.1016/0749-596x(85)90024-5
- Frankish, C. (1989). Perceptual organization and precategorical acoustic storage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(3), 469-479. doi: 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), e0168699. doi: 10.1371/journal.pone.0168699
- Gorin, S., & Majerus, S. (submitted). Contribution of Musical Transitional Knowledge to Musical Short-Term Memory
- Grossberg, S. (2003). Resonant neural dynamics of speech perception. *Journal of Phonetics*, *31*(3-4), 423-445. doi: 10.1016/S0095-4470(03)00051-2
- Halpern, A. R., Bartlett, J. C., & Dowling, W. J. (1995). Aging and experience in the recognition of musical transpositions. *Psychology and Aging*, *10*(3), 325-342. doi: 10.1037/0882-7974.10.3.325

- Hartley, T., Hurlstone, M. J., & Hitch, G. (2016). Effects of rhythm on memory for spoken sequences: A model and tests of its stimulus-driven mechanism. *Cognitive Psychology*, *87*, 135-178. doi: 10.1016/j.cogpsych.2016.05.001
- Henson, R. N. A. (1996). *Short-term memory for serial order*. (Unpublished doctoral thesis).
- Henson, R. N. A. (1998). Short-Term Memory for Serial Order: The Start-End Model. *Cognitive Psychology*, *36*(2), 73-137. doi: 10.1006/cogp.1998.0685
- Henson, R. N. A. (1999). Positional information in short-term memory: relative or absolute? *Memory and Cognition*, *27*(5), 915-927. doi: 10.3758/BF03198544
- Hitch, G. J., Burgess, N., Towse, J. N., & Culpin, V. (1996). Temporal Grouping Effects in Immediate Recall: A Working Memory Analysis. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *49*(1), 116-139. doi: 10.1080/713755609
- 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. doi: 10.1037/a0038223
- Hurlstone, M. J., & Hitch, G. J. (in press). How is the serial order of a visual sequence represented? Insights from transposition latencies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*
- 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. doi: 10.1037/a0034221
- JASP Team. (2016). JASP (Version 0.8.0.0) [Computer software].
- Krumhansl, C. L. (1990). *Cognitive foundations of musical pitch*. New-York, NY: Oxford University Press.
- Krumhansl, C. L., & Kessler, E. J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review*, *89*(4), 334-368. doi: 10.1037/0033-295x.89.4.334
- Lashley, K. S. (1951). The problem of serial order in behavior. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior: The Hixon symposium*. New-York; NY: John Wiley.
- Lee, C. L., & Estes, W. K. (1981). Item and order information in short-term memory: Evidence for multilevel perturbation processes. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(3), 149-169. doi: 10.1037/0278-7393.7.3.149
- Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.
- Lewandowsky, S., & Farrell, S. (2008). Short-Term Memory: New Data and a Model. *The Psychology of Learning and Motivation*, *49*, 1-48. doi: 10.1016/S0079-7421(08)00001-7
- MacKay, D. S. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, *8*, 323-350
- Majerus, S. (2013). Language repetition and short-term memory: an integrative framework. *Frontiers in Human Neuroscience*, *7*, 357. doi: 10.3389/fnhum.2013.00357
- Martin, R. C., Lesch, M. F., & Bartha, M. C. (1999). Independence of Input and Output Phonology in Word Processing and Short-Term Memory. *Journal of Memory and Language*, *41*(1), 3-29. doi: 10.1006/jmla.1999.2637

- Mathias, B., Pfordresher, P. Q., & Palmer, C. (2015). Context and meter enhance long-range planning in music performance. *Frontiers in Human Neuroscience*, 8, 1040. doi: 10.3389/fnhum.2014.01040
- Mathot, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: an open-source, graphical experiment builder for the social sciences. *Behavioral Research Methods*, 44(2), 314-324. doi: 10.3758/s13428-011-0168-7
- 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. doi: 10.1016/S0749-596X(02)00014-1
- Morey, R. D. (2015). *Workshop: Bayesian analysis with BayesFactor*. Katholieke Universiteit Leuven, Belgium.
- Müllensiefen, D., & Wiggins, G. A. (2011). Sloboda and Parker's recall paradigm for melodic memory: a new, computational perspective. In I. Deliège & J. W. Davidson (Eds.), *Music and the Mind: Essays in Honour of John Sloboda* (pp. 161-186). New-York, NY: Oxford University Press.
- Ng, H. L. H., & Maybery, M. T. (2002). Grouping in short-term verbal memory: Is position coded temporally? *The Quarterly Journal of Experimental Psychology Section A*, 55(2), 391-424. doi: 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, & Cognition*, 31(1), 175-181. doi: 10.1037/0278-7393.31.1.175
- Oberauer, K. (2003). Understanding serial position curves in short-term recognition and recall. *Journal of Memory and Language*, 49(4), 469-483. doi: 10.1016/S0749-596X(03)00080-9
- Oberauer, K., Lewandowsky, S., Farrell, S., Jarrold, C., & Greaves, M. (2012). Modeling working memory: An interference model of complex span. *Psychonomic Bulletin & Review*, 19(5), 779-819. doi: 10.3758/s13423-012-0272-4
- Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105(4), 761-781. doi: 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. doi: 10.1037/0033-295X.110.4.683
- Parmentier, F. B. R., Andrés, P., Elford, G., & Jones, D. M. (2006). Organization of visuo-spatial serial memory: interaction of temporal order with spatial and temporal grouping. *Psychological Research*, 70, 200-217. doi: 10.1007/s00426-004-0212-7
- 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. doi: 10.3758/BF03196602
- Pfordresher, P. Q., & Brown, S. (2007). Poor-Pitch Singing in the Absence of "Tone Deafness". *Music Perception: An Interdisciplinary Journal*, 25(2), 95-115. doi: 10.1525/mp.2007.25.2.95
- Pfordresher, P. Q., Brown, S., Meier, K. M., Belyk, M., & Liotti, M. (2010). Imprecise singing is widespread. *The Journal of the Acoustical Society of America*, 128(4), 2182-2190. doi: 10.1121/1.3478782
- Pfordresher, P. Q., Palmer, C., & Jungers, M. K. (2007). Speed, Accuracy, and Serial Order in Sequence Production. *Cognitive Science*, 31(1), 63-98. doi: 10.1080/03640210709336985

- Rotello, C. M., & Macmillan, N. A. (2008). Response Bias in Recognition Memory. In A. S. Benjamin & B. H. Ross (Eds.), *Skill and Strategy in Memory Use* (Vol. 48, pp. 61-94): Academic Press.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, *56*(5), 356-374. doi: 10.1016/j.jmp.2012.08.001
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin and Review*, *16*(2), 225-237. doi: 10.3758/PBR.16.2.225
- Ryan, J. (1969a). Grouping and short-term memory: Different means and patterns of grouping. *Quarterly Journal of Experimental Psychology*, *21*(2), 137-147. doi: 10.1080/14640746908400206
- Ryan, J. (1969b). Temporal grouping, rehearsal and short-term memory. *Quarterly Journal of Experimental Psychology*, *21*(2), 148-155. doi: 10.1080/14640746908400207
- Schulze, K., Mueller, K., & Koelsch, S. (2011). Neural correlates of strategy use during auditory working memory in musicians and non-musicians. *European Journal of Neuroscience*, *33*(1), 189-196. doi: 10.1111/j.1460-9568.2010.07470.x
- Schulze, K., Zysset, S., Mueller, K., Friederici, A. D., & Koelsch, S. (2011). Neuroarchitecture of verbal and tonal working memory in nonmusicians and musicians. *Human Brain Mapping*, *32*(5), 771-783. doi: 10.1002/hbm.21060
- Swets, J. A. (1973). The Relative Operating Characteristic in Psychology. *Science*, *182*(4116), 990-1000. doi: 10.1126/science.182.4116.990
- Tillmann, B., Dowling, W. J., Lalitte, P., Monlin, P., Schulze, K., Poulin-Charronnat, B., . . . Bigand, E. (2013). Influence of Expressive Versus Mechanical Musical Performance on Short-term Memory for Musical Excerpts. *Music Perception: An Interdisciplinary Journal*, *30*(4), 419-425. doi: 10.1525/mp.2013.30.4.419
- Vandierendonck, A. (2016). Modality independence of order coding in working memory: Evidence from cross-modal order interference at recall. *The Quarterly Journal of Experimental Psychology*, *69*(1), 161-179. doi: 10.1080/17470218.2015.1032987
- Verde, M. F., Macmillan, N. A., & Rotello, C. M. (2006). Measures of sensitivity based on a single hit rate and false alarm rate: the accuracy, precision, and robustness of  $d'$ ,  $A_z$ , and  $A'$ . *Perception & Psychophysics*, *68*(4), 643-654. doi: 10.3758/bf03208765
- Verde, M. F., & Rotello, C. M. (2007). Memory strength and the decision process in recognition memory. *Memory & Cognition*, *35*(2), 254-262. doi: 10.3758/BF03193446
- Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin and Review*, *14*(5), 779-804. doi: 10.3758/BF03194105
- Wagenmakers, E.-J., Lee, M., Lodewyckx, T., & Iverson, G. J. (2008). Bayesian Versus Frequentist Inference. In H. Hoijtink, I. Klugkist & P. A. Boelen (Eds.), *Bayesian Evaluation of Informative Hypotheses* (pp. 181-207). New-York, NY: Springer.
- 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 and Cognition*, *38*(2), 163-175. doi: 10.3758/MC.38.2.163



Table 1. Distribution of the strategies reported by the participants in Experiment 1 as a function of temporal grouping conditions.

<b>Strategy type</b>	<b>Proportion of reports</b>	
	<b>Ungrouped</b>	<b>Grouped</b>
Passive	.04	.07
Auditory rehearsal	.26	.19
Auditory rehearsal of tone names	.07	.04
Mental imagery of the melody shape (visual-based representation)	.15	.07
Mental imagery of the melody shape (motor-based representation)	.41	.44
Grouping	.07	.19
No strategy used	.00	.00

Table 2. Distribution of the strategies reported by the participants in Experiment 2 as a function of temporal grouping conditions.

Strategy type	Proportion of reports	
	Ungrouped	Grouped
Passive	.03	.00
Auditory rehearsal	.30	.33
Auditory rehearsal of tone names	.03	.07
Mental imagery of the melody shape visual-based representation)	.27	.27
Mental imagery of the melody shape (motor-based representation)	.33	.30
Grouping	.03	.03
No strategy used	.00	.00

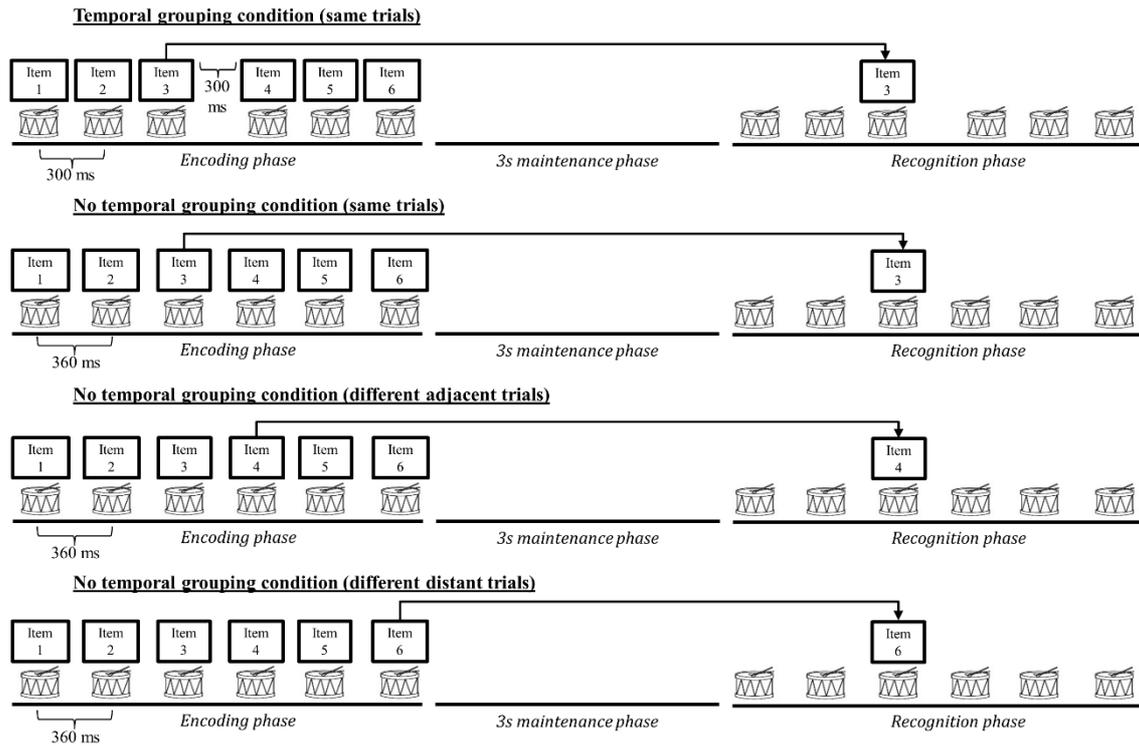


Figure 1. Schematic representation of task design. The two first examples represent same trials for grouped and ungrouped sequences. The two last examples represent different adjacent and different distant trials (from top to bottom) for ungrouped sequences. Note that for different trials the same task setup was also applied to grouped sequences (see Methods for timing details).

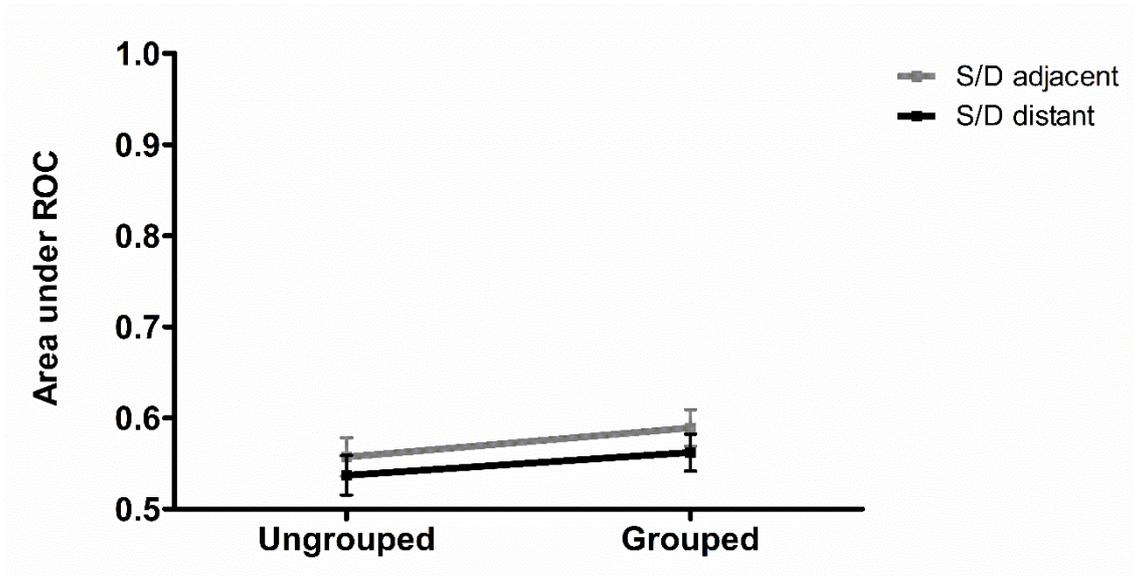


Figure 2. Means and standard errors for area under the curve for ROC analyses in Experiment 1. The scores indicate discrimination levels between similar and different adjacent trials (S/D adjacent) and between similar and different distant trials (S/D distant), as a function of temporal grouping conditions.

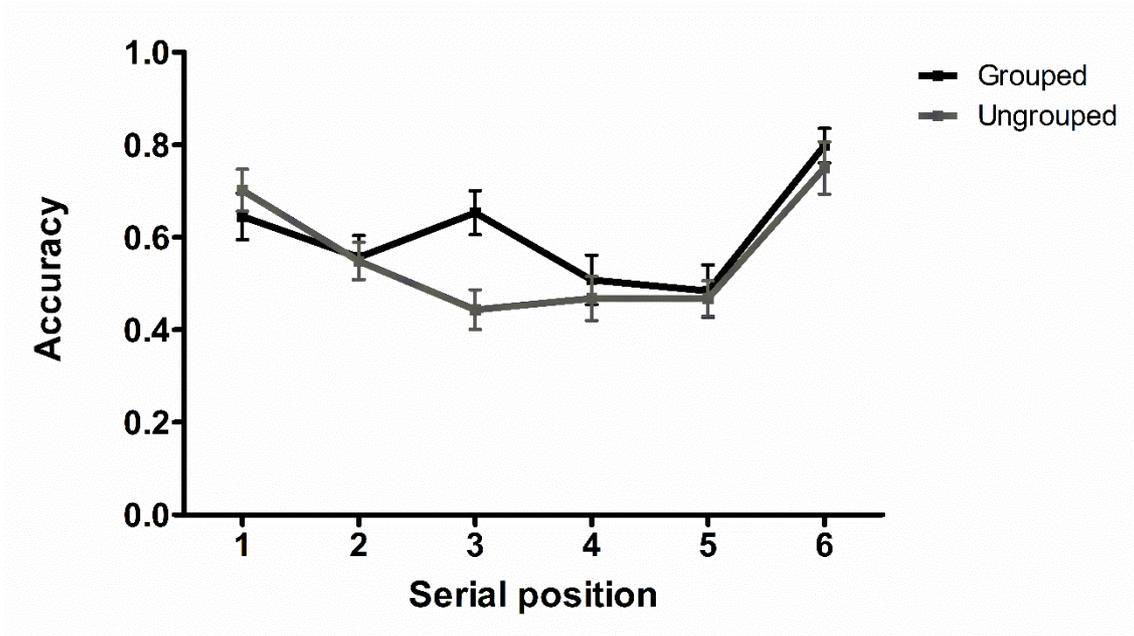


Figure 3. Means and standard errors for the proportion of correct detections of matching probe trials in Experiment 1, as a function of serial position and temporal grouping conditions.

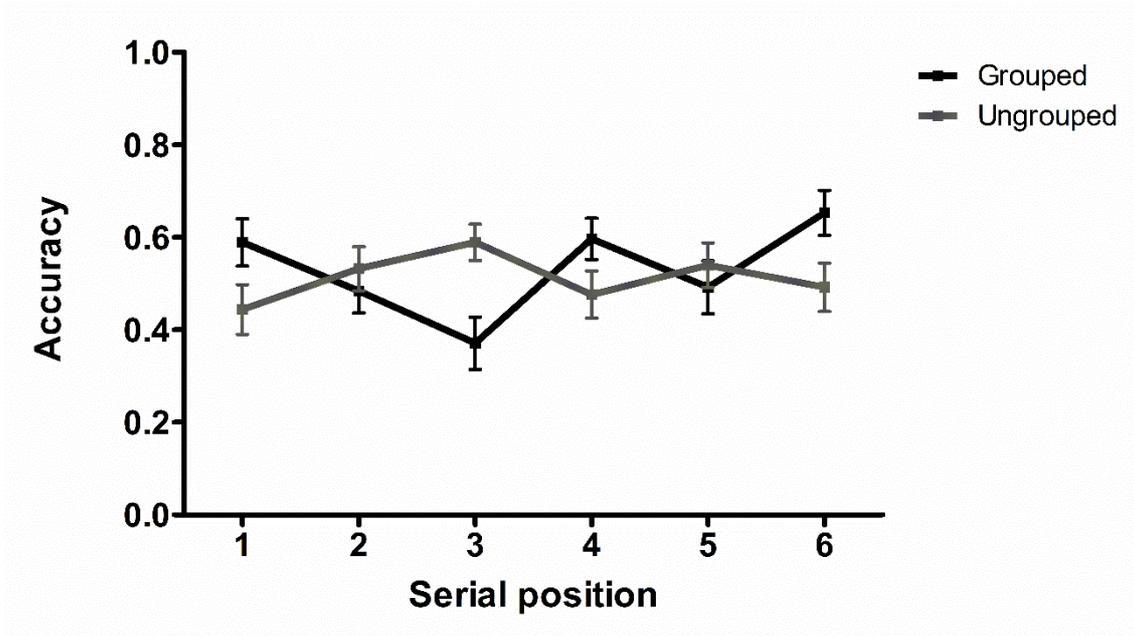


Figure 4. Means and standard errors for the proportion of correct rejections of non-matching adjacent probe trials in Experiment 1, as a function of serial position and temporal grouping conditions.

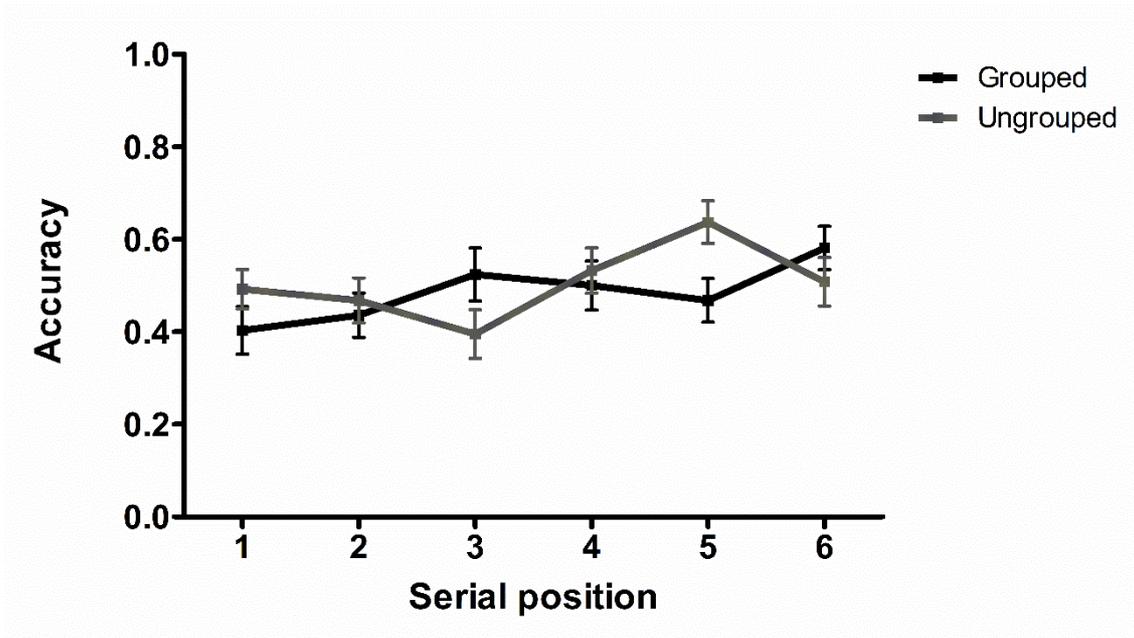


Figure 5. Means and standard errors for the proportion of correct rejections of non-matching distant probe trials in Experiment 1, as a function of serial position and temporal grouping conditions.

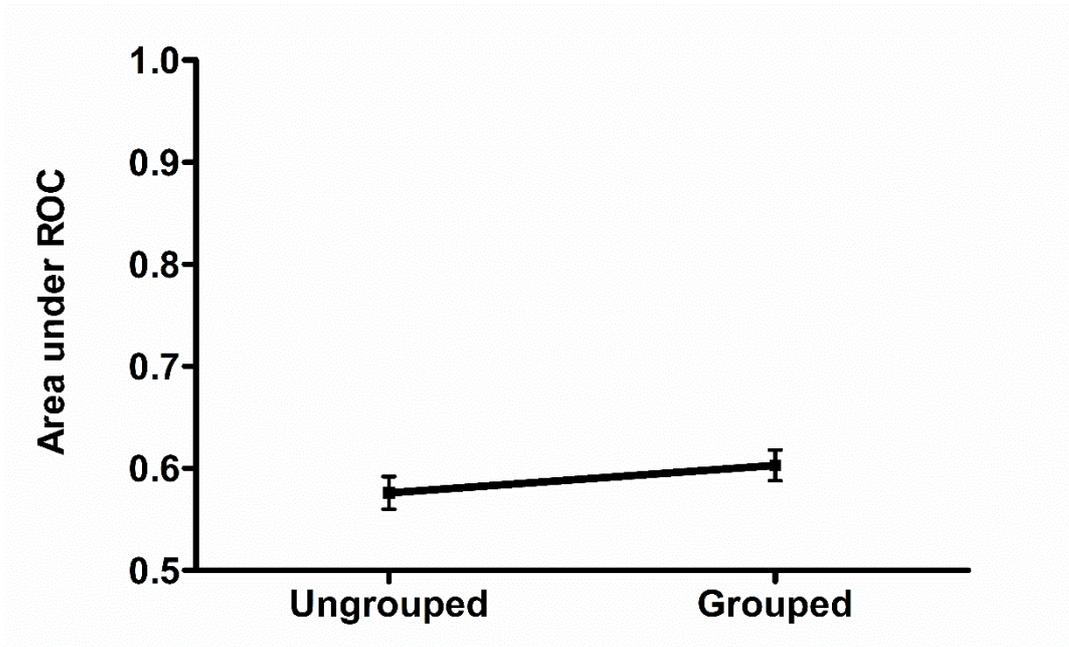


Figure 6. Means and standard errors for area under the curve for ROC analyses in Experiment 2, as a function of temporal grouping conditions.

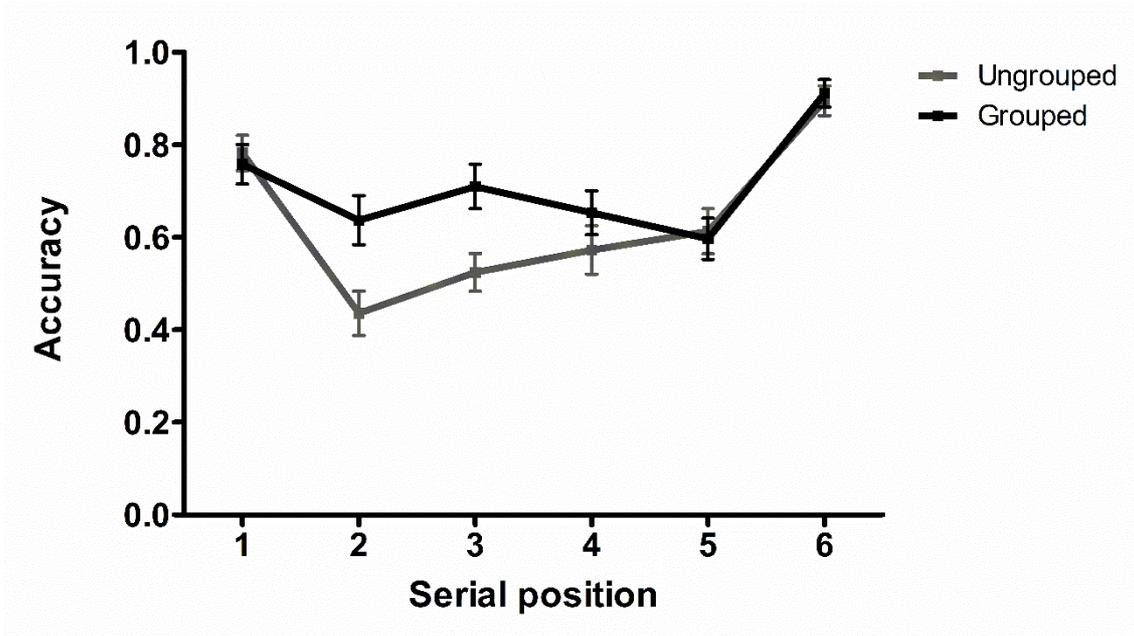


Figure 7. Means and standard errors for the proportion of correct detections of matching probe trials in Experiment 2, as a function of serial position and temporal grouping conditions.

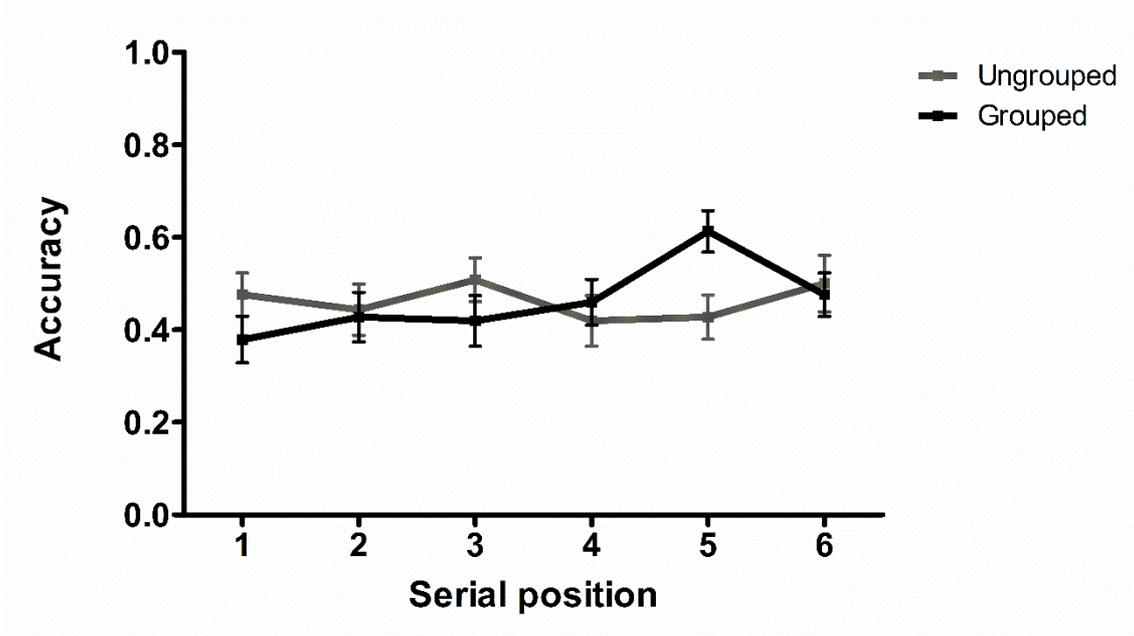


Figure 8. Means and standard errors for the proportion of correct rejections of non-matching probe trials in Experiment 2, as a function of serial position and temporal grouping conditions.