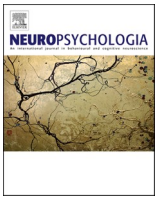
Neuropsychologia 163 (2021) 108080

Contents lists available at ScienceDirect

Neuropsychologia

journal homepage: www.elsevier.com/locate/neuropsychologia



Comparisons between short-term memory systems for verbal and rhythmic stimuli

Joshua D. Hoddinott [a,b], Dirk Schuit [a], Jessica A. Grahn [a,c,*]

[a] Brain and Mind Institute, University of Western Ontario, London, Ontario, Canada
[b] Neuroscience Program, University of Western Ontario, London, Ontario, Canada
[c] Department of Psychology, University of Western Ontario, London, Ontario Canada

ARTICLE INFO

Keywords:
Music
Rhythm
Timing
fMRI
Auditory perception

ABSTRACT

Auditory short-term memory is often conceived of as a unitary capacity, with memory for different auditory materials (such as syllables, pitches, rhythms) posited to rely on similar neural mechanisms. One spontaneous behavior observed in short-term memory studies is 'chunking'. For example, individuals often recount digit sequences in groups, or chunks, of 3–4 digits, and chunking is associated with better performance. Chunking may also operate in musical rhythm, with beats acting as potential chunk boundaries for tones in rhythmic sequences. Similar to chunking, beat-based structure in rhythms also improves performance. Thus, it is possible that beat processing relies on the same mechanisms that underlie chunking of verbal material. The current fMRI study examined whether beat perception is indeed a type of chunking, measuring brain responses to chunked and 'unchunked' letter sequences relative to beat-based and non-beat-based rhythmic sequences. Participants completed a sequence discrimination task, and comparisons between stimulus encoding, maintenance, and discrimination were made for both rhythmic and verbal sequences. Overall, rhythm and verbal short-term memory networks overlapped substantially. When contrasting rhythmic and verbal conditions, rhythms activated basal ganglia, supplementary motor area, and anterior insula more than letter strings did, during both encoding and discrimination. Verbal letter strings activated bilateral auditory cortex more than rhythms did during encoding, and parietal cortex, precuneus, and middle frontal gyri more than rhythms did during discrimination. Importantly, there was a significant interaction in the basal ganglia during encoding: activation for beat-based rhythms was greater than for non-beat-based rhythms, but verbal chunked and unchunked conditions did not differ. The interaction indicates that beat perception is not simply a case of chunking, suggesting a dissociation between beat processing and chunking-based grouping mechanisms.

1. Introduction

One of the most influential models of auditory short-term memory is Baddeley's phonological loop model (Baddeley and Hitch, 1974). This model is developed largely on the basis of studies that use linguistic and verbal material, but may also account for auditory short-term memory processing of other material, such as rhythm. Verbal short-term memory studies indicate that grouping or 'chunking' is a spontaneous behavior that benefits short-term memory performance (Gobet et al., 2001). When individuals verbally recount digit strings (such as in digit span tasks) they recount numbers in groups of 3 or 4, even when no grouping is present at encoding. This chunking in verbal memory may be analogous to beat perception in rhythm, with beats acting as chunk boundaries for groups of time intervals that make up the rhythm. As shown in chunking, behavioral performance, such as reproduction accuracy, is enhanced when temporal sequences have a regular beat (Grahn, 2009, 2012; Grahn and Brett, 2007; Patel et al., 2005). Beat-based structure may therefore enable encoding of rhythmic patterns in chunks (Schaefer et al., 2011), and the same neural mechanisms may underlie chunking-based performance improvements in both verbal short-term memory and beat-based rhythms. The current fMRI study examines this possibility, comparing brain responses to chunked versus unchunked verbal sequences and beat-based versus nonbeat-based rhythmic sequences.

* Corresponding author. Brain and Mind Institute & Department of Psychology, University of Western Ontario, London, Ontario, N6A 5B7, Canada.
E-mail address: jgrahn@uwo.ca (J.A. Grahn).

https://doi.org/10.1016/j.neuropsychologia.2021.108080
Received 19 October 2020; Received in revised form 26 October 2021; Accepted 29 October 2021
Available online 30 October 2021

1.1. Auditory short-term memory for verbal and rhythmic stimuli

Most models of auditory short-term memory posit two components: one involved in representing the to-be-remembered items and the other involved in maintaining those representations (Baddeley and Hitch, 1974; Cowan, 1999). In Baddeley's influential model of short-term memory (Baddeley and Hitch, 1974), the first component is the 'phonological short-term store', a storage buffer for auditory memory traces. The traces decay rapidly, therefore the 'articulatory loop' is required to maintain items in auditory memory through active rehearsal (Baddeley, 2003). Auditory stimuli automatically enter the phonological store, which acts as an 'inner ear', remembering the sounds in the correct order. The articulatory process acts as an 'inner voice', repeating the information to prevent it from decaying. A third component, called a timing signal, has been proposed to mark the serial order of items in the store (Baddeley, 2000; Henson et al., 2003). However, as this component signals the order of items, not the absolute or relative timing between items, it is unclear how the specific temporal intervals in a rhythm might be represented in the model.

Although it is unclear exactly how the model represents temporal aspects of rhythms, previous work suggests that the articulatory loop component is important (Hall and Gathercole, 2011; Saito, 1994, 2001). For example, Saito (1998) asked participants to listen to a short rhythm and reproduce it after a 5-s delay. During encoding and delay participants either engaged in articulatory suppression (silently mouthing the vowels AEIOU), which occupies the articulatory loop, or drawing of squares, which does not. Articulatory suppression interfered with rhythm performance much more than square drawing, suggesting that rhythm encoding and maintenance rely on the articulatory loop.

Similarly, rhythmic tapping can interfere with verbal short-term memory, although the interference depends on task parameters. Tapping a regular (isochronous) rhythm may interfere with short-term memory, depending on whether it is self-paced or externally paced. When tapping is self-paced, participants may simply adjust their rate of tapping to align with the presentation or rehearsal rate of the to-be-remembered verbal stimuli. However, externally paced regular tapping during encoding and maintenance does interfere with verbal short-term memory performance when performed at a different rate from the verbal stimuli (Henson et al., 2003). Moreover, rhythmic tapping of complex (not isochronous) rhythms during encoding and maintenance interferes with verbal short-term memory performance regardless of whether it is self-paced or externally-paced (Saito, 1994). Importantly, this is likely an influence of rhythmic motor output, as previous work reveals the enhancement of verbal memory when an isochronous tone is played (but not tapped along with) during maintenance (Plancher et al., 2018).

Additional evidence for a link between auditory short-term memory and rhythm processing comes from behavioral measures of individual differences. Auditory short-term memory capacity varies across individuals, as assessed by span tasks, such as digit span. To measure digit span, a list of spoken digits is heard then repeated. The number of correctly recalled digits is the capacity of the short-term memory store, and is usually between five and nine items (Miller, 1956). Digit span correlates with the ability to reproduce rhythms, further supporting the idea that the systems for remembering verbal and rhythmic material may overlap (Grahn and Schuit, 2012).

Finally, there is substantial neural evidence to suggest overlap between short-term memory and rhythm processing networks. A review of short-term memory studies (Rottschy et al., 2012) found consistent activation in Broca's area, pre-supplementary motor area (pre-SMA), dorsal and ventrolateral premotor cortex (PMC), inferior frontal gyrus, cerebellum (lobule VI), as well as the intraparietal sulcus (IPS) and lateral prefrontal cortex. Subcortical activations were found in bilateral thalamus and left basal ganglia. Verbal compared to non-verbal tasks were more likely to recruit left Broca's area, whereas non-verbal tasks were more likely to recruit left pre-SMA, SMA, and bilateral dorsal PMC. Articulatory rehearsal, specifically, activates a subset of these short-term memory areas, including Broca's area, SMA and pre-SMA, dorsal and ventrolateral PMC, cerebellum, and anterior insula (Awh et al., 1996; Chen and Desmond, 2005; Fiez et al., 1996; Gruber and von Cramon, 2003; Rauschecker and Scott, 2009). These areas are also commonly activated in studies of rhythm perception and production (Chen et al., 2008a, 2008b; Grahn, 2009; Vuust et al., 2006; Li et al., 2019; Trost et al., 2014; Araneda et al., 2016; Heard and Lee, 2020; Vikene et al., 2019), suggesting reliance on at least partially overlapping neural processes. Overlapping neural activations do not necessarily reflect general memory processes, as rhythm perception studies without a memory component also elicit activation in the SMA, PMC, cerebellum, and insula (Grahn & Rowe, 2009, 2013).

Previous work has specifically compared short-term memory for verbal and musical material, but has generally focused on pitch, not rhythm (Gaab et al., 2003; Hickok et al., 2003). Consistent with the review study above (Rottschy et al., 2012), these studies find overlap between the brain areas involved in short-term memory for verbal and pitch sequences. Activations are observed in parietal cortex (supramarginal gyrus (SMG) and intraparietal sulcus (IPS)), posterior temporal cortex (planum temporale or area SPt), ventrolateral and dorsolateral PMC, Broca's area, and dorsolateral cerebellum (Gaab et al., 2003; Hickok et al., 2003; Zatorre et al., 1994). More recent studies (Koelsch et al., 2009; Schulze and Koelsch, 2012) find similar overlap between verbal and pitch stimuli during rehearsal: vlPMC and dlPMC, the anterior insula, the SMG/IPS, the planum temporale, the IFG, pre-SMA, basal ganglia, and the cerebellum. The authors suggest that vlPMC and Broca's area are part of an active rehearsal component of the articulatory loop, as these areas did not respond to simply subvocalizing without a rehearsal function. Therefore, these areas might be expected to respond during rehearsal of rhythm as well.

1.2. Beat perception and chunking

Both beat perception (in rhythm) and chunking (in verbal sequences) are known to improve short-term memory performance. Beat perception spontaneously arises in the context of auditory rhythm: humans perceive a regular pulse, or beat, that marks equally spaced points in time (Large and Palmer, 2002; Nettl, 2000). Perception of a beat occurs without effort as long as the auditory sequence has a regular temporal structure, such as periodically occurring event onsets in the range of ~300–900 ms (Parncutt, 1994; van Noorden and Moelants, 1999). Several studies confirm that beat perception leads to higher accuracy in rhythm synchronization, discrimination, and reproduction (Grahn, 2012; Grahn and Brett, 2007; Kung et al., 2013; Patel et al., 2005).

Chunking refers to the process of taking individual units of information and grouping them into larger units (chunks). A common example of chunking occurs in telephone numbers, in which individual digits are grouped into 3- and 4-unit chunks. Chunking is a useful method for information reduction. By grouping individual elements into larger blocks, information becomes easier to retain and recall. The grouping of sequential units into a single, larger, unit to facilitate performance is observed in a variety of domains (Ericsson et al., 1980; Gobet, 1998). For example, many chunking studies have been conducted in the motor learning domain, and find that sequences of finger movements are spontaneously chunked (Kennerley et al., 2004; Sakai et al., 2003; Verwey et al., 2009), reducing short-term memory load during ongoing performance (Bo and Seidler, 2009; Ericsson et al., 1980).

1.3. Neural responses during beat perception and chunking

Neural areas that respond to motor sequence chunking overlap with those that respond to beat perception. Beat compared to nonbeat rhythms activate the SMA, left inferior frontal gyrus, and the basal ganglia (Bengtsson et al., 2009; Grahn and Brett, 2007). In cases of basal ganglia activity disruption, behavioral benefits of the beat, such as

improved sequence memory, are reduced (Grahn and Rowe, 2009). Similarly, motor sequence chunking elicits basal ganglia activity, with basal ganglia neurons firing at chunk boundaries during motor sequence production (Graybiel, 2008; Thorn and Graybiel, 2010; Yin and Knowlton, 2006), and basal ganglia disruption reduces motor sequence chunking (Tremblay et al., 2010; Boyd et al., 2009). Because of this overlap, the processes of extracting a beat and chunking a sequence may rely on common neural resources – particularly in the basal ganglia.

Both perceptual chunking and beat perception have been investigated using fMRI. One previous study presented 6-letter visual sequences (Henson et al., 2000), and found that chunked sequences (sequences in which a brief pause was inserted between 3-letter groups) evoked greater activity in right inferior frontal gyrus (BA 47). However, the findings may not readily apply to the auditory modality, as chunking effects are generally smaller for visual compared to auditory sequences (Frankish, 1985, 1989; Hitch et al., 1996). More recently, Kalm et al. (2012) investigated chunking using auditory sequences of 6 or 9 letters. Chunking resulted in reduced activity in auditory areas. For 9-letter strings only, parietal areas were more active for chunked than unchunked strings. When comparing activation differences for chunked and unchunked letter strings to activation differences observed between beat and nonbeat rhythms, not much overlap exists. However, as verbal chunking and beat perception are yet to be compared in the same study, it is difficult to conclude whether neural overlap exists between the two phenomena. Therefore, in the current study we tested whether beat perception and verbal chunking rely on similar neural mechanisms, using a short-term memory paradigm that could be applied to both verbal and rhythmic stimuli. Each trial consisted of stimulus presentation (encoding), a variable-length silent delay period for rehearsal (maintenance), and a second stimulus presentation (discrimination), after which participants indicated whether the second stimulus was the same as or different from the first. Half of the trials involved discriminating rhythmic sequences (comparing two rhythms to determine whether the timing was same or different), and the other half involved discriminating letter sequences (comparing two strings of different letters to determine whether letter order was the same or different). Half of the rhythm trials used beat-based sequences, which induced a perception of a regular beat, and the other half were nonbeat-based sequences, with irregular timing in which no beat perception was possible. Similarly, on half of the letter trials, the timing of the letter presentation created two equal chunks, and on the other half, the timing was irregular, with no equal chunks. Thus, half of the letter trials could be easily encoded as chunks upon stimulus presentation, and the other half could only be chunked if the subject internally recoded the stimulus into chunks. Because of the phenomenological and behavioral similarities between beat perception and chunking, we examined whether neural responses overlapped for the two phenomena when tested in the auditory modality. As both beat perception and chunking involve structuring a sequence into hierarchical groups, and both involve the basal ganglia, we predicted that, if beat perception and chunking relied on the same neural mechanisms, differences in basal ganglia activity between beat and non-beat rhythms would be identical to differences between chunked and unchunked verbal sequences.

2. Methods

2.1. Participants

18 volunteers (4 female; $M_{age} = 28.3$, $SD = 8.65$) participated in the brain imaging study. All participants completed the experiment and received financial compensation for participation. The Cambridge University Psychological Research Ethics Committee provided clearance for the study (CPREC, 2009.17).

2.2. Stimuli

Recordings of spoken letters from a single male speaker were used to construct verbal and non-verbal stimuli. Verbal sequences were constructed using different letters for each interval (fourteen letters were used across the experiment, including 'B, C, D, E, G, I, J, K, L, M, O, P, Q, & U'). Non-verbal sequences were constructed using only one letter for each interval onset (Fig. 1). The letter used for each rhythmic sequence was balanced across beat and non-beat sequences. Each recorded letter had similar pitch ($M = 113.8$ Hz, $SD = 2.8$ Hz), duration ($M = 162$ ms, $SD = 20$ ms), intensity ($M = 75$ db, $SD = 1.9$ db) and peak onset time ($M = 35$ ms, $SD = 17$ ms). Verbal and non-verbal conditions included short (e.g., 4-letter) and long (e.g., 8-letter) sequences to decorrelate the BOLD response between task stages (jittered lengths aid in decorrelation). Sequence length was therefore not included in the final analysis, as stimuli that are systematically different in length cannot be compared. All stimuli were created using GarageBand (Apple, Inc., v4.1.2 (248.7)).

Beat rhythms were constructed using the following six core patterns: 1111, 112, 211, 22, 31, 4, similar to previous work (Grahn and Brett, 2007; Grahn and Rowe, 2009). Short and long rhythms consisted of two or three core patterns (e.g., 1122114), respectively. The shortest interval (i.e., 1) ranged from 220 to 280 ms, in 10 ms steps, creating seven potential tempi. The other intervals in the rhythm were integer multiples of the shortest interval. One final letter was added to the end of each sequence to mark the end of the last interval. None of the six core patterns were repeated within a rhythm. On each trial, one of the seven different tempi was used. The trial-to-trial tempo change prevented carry-over of the beat from one trial to the next trial.

Inter-tone intervals in each beat rhythm were modified to create nonbeat counterpart rhythms. One third of the intervals in a rhythm kept their original length, another third were increased by 1/3 of the length of the 1 unit and the final third of the intervals were decreased by 1/3 of the length of the 1 unit. Thus, the nonbeat rhythms were the same as the beat rhythms in overall duration and number of intervals but had irregular timing and no regular beat (see Fig. 1).

For the verbal sequences, strings of four (short) or eight (long) different, easily intelligible and distinguishable letters were created (e.g., 'Q L D C U M J P'). Half of the sequences had regular timing and half had irregular timing. For the regular sequences, the strings were divided in half to create two groups of letters (two groups of two letters for short sequences, or two groups of four letters for long sequences), ensuring that the sequences were initially encoded as two chunks (Bower and Winzenz, 1969; Dowling, 1973). The letter onsets within a group were separated by 400 ms, and each group was separated by 800 ms. The letter onsets in the irregular strings were separated by unequal time intervals. The four-letter sequences used 233, 533 and 833 ms intervals (in random order). The eight-letter sequences used 257, 307, 357, 457, 557 and 657 ms intervals (again, in random order). Because the 4-letter

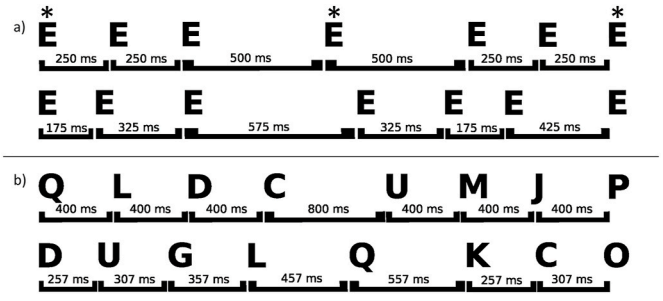


Fig. 1. Schematic examples of the four different stimulus types, constructed with spoken letters. Time intervals between letter onsets are given in milliseconds (ms). (a) A beat rhythm (upper) and nonbeat rhythm (lower); * indicates theoretical beat locations. (b) A chunked verbal sequence (upper) and an unchunked verbal sequence (lower).

sequences were comprised of three intervals, six unique temporal patterns (i.e., the order of intervals) were created for this condition. Because there were six different time intervals for the 8-letter sequences, each sequence had a unique temporal pattern. Importantly, subjects were unaware of which sequences were being presented on each trial, and no participants reported detection of repetitions of timing patterns of any kind upon debriefing. Sample sequences are shown in Fig. 1.

2.3. Design

A 2 × 2 within-subject design was used, with experimental factors *stimulus type* (rhythm, verbal) and *temporal regularity* (beat/nonbeat for rhythm trials, chunked/unchunked for verbal trials) for each stage (encoding, maintenance, discrimination of same rhythms/verbal sequences, and discrimination of different rhythms/verbal sequences).

There were three trial types: full trials, stimulus-response-only trials and null trials. Full trials consisted of encoding, maintenance, and discrimination stages as well as a response. For the participants, each stage was distinguished by a differently coloured display. During the *encoding stage*, the initial stimulus was heard with a blue display. The subsequent *maintenance stage* had a black display and lasted zero, one, or two times the length of the initial stimulus ($M = 2.8$ s, $SD = 1.9$ s). The *discrimination stage* stimulus was heard with a green display. The stimulus was either the same as or different from the initial stimulus. For 'different' stimuli, a different sequence of the same type as the stimulus (beat/chunked or nonbeat/unchunked) was used. During the *response stage*, the screen was red and participants had 2 s to indicate whether the stimuli were same or different with a button press. Stimulus-response-only trials had only the initial stimulus presentation, after which the screen turned red and text instructed the participant to press the left or right button. Null trials consisted of a 9-s blank screen, and were experienced by the participant as simply an extended inter-trial interval. Null trials were used to measure baseline activity throughout the experiment. The variable length of maintenance stages, stimulus-response-only trials, and null trials were necessary to allow the hemodynamic response to the different trial stages to be de-correlated and therefore estimable. The variable length of the maintenance stage decouples the encoding stage from the maintenance stage and also the maintenance stage from the discrimination stage in the design. The stimulus-response-only trials decouple the encoding stage from the discrimination stage and the discrimination stage from the response.

There were 16 blocks in the experiment. One block comprised eight full trials (2 with a zero-length maintenance stage, 4 with 1x-length maintenance, 2 with 2x-length maintenance), two stimulus-response-only trials and two null trials. Each block contained only rhythm or verbal trials. The stimulus type alternated (rhythm, verbal condition) with each block. Each block contained an equal number of regular/irregular trials, balanced for trial types (4 full trials, 1 stimulus-response-only trial per condition). Thus, across the experiment each of the 4 conditions (beat, nonbeat, chunked, and unchunked) were presented across 32 full trials and 8 stimulus-response-only trials. For each subject, each unique sequence was presented as the encoded sequence on 1 trial. Though the order of trials was randomly selected for each subject, trials were identical across subjects: Each encoded sequence (sequence 1 in the trial) appeared on either a 'same' or 'different' trial identically for each subject. For example, every subject was presented with the sequence 'G Q B P O C E U' followed by the same sequence as the target. Thus, the correct response for each trial was consistent across subjects, and only the order of the trials was randomized.

2.4. Procedure

Participants gave written informed consent and practiced one rhythm and one verbal block prior to entering the scanner. Each practice block contained four full trials and one stimulus-response-only trial. All stimuli were unique to the practice session.

2.5. MR scanning specifications

A 3T Siemens Tim Trio MRI scanner was used to collect two runs with 540 echoplanar imaging (EPI) volumes in each. All EPI data had 36 slices, matrix size of 64 × 64, echo time (TE) = 30 ms, repetition time (TR) = 2.19 s, field of view = 19.2 × 19.2 cm, flip angle = 78°, slice thickness 3 mm, interslice distance of 0.75 mm, and in-plane resolution of 3 × 3 mm. High-resolution MPRAGE anatomical images (TR = 2250 ms, TE = 2.99 ms, flip angle = 9°, inversion time = 900 ms, 256 × 256 x 192 isotropic 1 mm voxels) were collected for anatomic localization and coregistration.

2.6. Data pre-processing and analysis

SPM8 was used for data analysis (SPM8; Wellcome Centre for Neuroimaging, London, UK). The first five EPI volumes of each run were discarded to allow for T1 equilibration. Images were sinc-interpolated in time to correct for acquisition time differences within each volume and realigned spatially with respect to the first image of the first run using trilinear interpolation. The coregistered MPRAGE image was segmented and normalized using affine and smoothly nonlinear transformations to the T1 template in Montreal Neurological Institute (MNI) space. The normalization parameters were then applied to the EPIs and all normalized EPI images were spatially smoothed with a Gaussian kernel of full-width half-maximum 8 mm. For each participant, encoding, maintenance, discrimination, and response were modelled separately for each condition. These were modelled using a regressor made from an on-off boxcar convolved with a canonical hemodynamic response function (apart from response, which was modelled with a delta function convolved with the canonical hemodynamic response function). EPI volumes associated with discrete artifacts were included as covariates of no interest (nulling regressors). This included volume displacements >4 mm or spikes of high variance in which scaled volume to volume variance was 4 times greater than the mean variance of the run. Autocorrelations were modelled using an AR (1) process and low-frequency noise was removed with a standard high-pass filter of 128 s.

The contrast images estimated from single participant models were entered into second-level random effects analyses for group inference (Penny and Holmes, 2003). Separate ANOVAs were conducted for encoding, maintenance, same discrimination, and different discrimination stages. Each ANOVA was a 2 × 2 within-subject ANOVA with the factors *temporal regularity* and *stimulus type*. All effects were estimated using t-contrasts. Significance level was $\alpha = 0.05$, using cluster-wise False Discovery Rate correction (Chumbley and Friston, 2009), with a cluster-forming threshold of $p < .001$ uncorrected.

In addition, for each stage (encoding, maintenance, discrimination same, discrimination different) regions of interest for the average of all conditions versus rest were created. Each region was defined as a 10-mm radius sphere around the peak voxel in a region, except putamen which was 5 mm. ANOVAs (2 × 2, as above) were conducted on each region's activity. This enabled us to test more sensitively for differences between conditions, using orthogonally-defined task-relevant regions. The specific regions tested are given in the Supplementary material. Effects are reported here for brain regions not already identified as significant by whole-brain analyses. Significance level was $\alpha = 0.05$, Bonferroni-corrected for number of regions tested at that particular stage. Thus, Encoding: $\alpha = 0.0031$; Maintenance: $\alpha = 0.0045$; Discrimination same: $\alpha = 0.0033$; Discrimination different: $\alpha = 0.0031$.

3. Results

3.1. Behavioral results

Performance accuracy (percentage of correctly discriminated sequences) was compared for each condition. Overall, accuracy was high ($M = 86\%$, $SD = 6.3\%$), as was intended to maximize the number of

trials that could be included in the fMRI analysis. Performance was better in the verbal conditions than in the rhythm conditions ($M_{\text{rhythm beat}} = 83\%$, $SD = 11\%$; $M_{\text{rhythm nonbeat}} = 73\%$, $SD = 12\%$; $M_{\text{verbal chunked}} = 94\%$, $SD = 6\%$; $M_{\text{verbal unchunked}} = 93\%$, $SD = 4\%$). A 2 (stimulus type) x 2 (temporal regularity) repeated measures ANOVA on percent correct scores indicated a significant interaction (F (1, 17) = 22.39, $p < .001$, $\eta_p^2 = 0.57$), driven by a significant difference in performance between beat and nonbeat rhythms (t (1, 17) = 5.13, $p < .001$) but no significant difference between verbal chunked and unchunked conditions (t (1,17) = 1.1, $p = .27$).

3.2. fMRI results: whole brain analyses

Only correct trials were modelled in the fMRI analyses to remove any confounds related to different error rates across conditions.

3.2.1. Encoding

Several brain areas responded during the encoding stage, across all conditions. These include bilateral superior temporal gyri, premotor cortex, SMA, inferior frontal gyrus, parietal cortex, cerebellum, and basal ganglia (see Supplementary Table 1 for coordinates of maxima). Analyses comparing verbal and rhythm conditions (Verbal > Rhythm) showed more activation in bilateral superior and middle temporal gyri for verbal sequences than for rhythms (see Supplementary Table 2), likely because of the greater acoustic variety in verbal sequences compared to rhythms. In the opposite contrast (Rhythm > Verbal), rhythms activated the basal ganglia, SMA, left cerebellum and right inferior frontal gyrus more than verbal sequences (Table 1). A significant interaction between stimulus type and temporal regularity was found in the bilateral basal ganglia (see Tables 1–3). As shown in Fig. 2, the basal ganglia had greater activity for beat than nonbeat rhythms, but no activity differences between chunked and unchunked verbal sequences. Both types of verbal sequences elicited activity levels similar to that of nonbeat rhythms (post hoc Bonferroni corrected pairwise t-tests between conditions were only significant for the rhythm beat condition versus each other condition (all p's < 0.001), the other comparisons were not significant).

3.2.2. Maintenance

Across all conditions, activation was observed in bilateral SMA, premotor cortex, insula, inferior frontal gyri, inferior parietal cortex, and right cerebellum (see Fig. 2 and Supplementary Table 3). There were no other significant main effects or interactions.

3.2.3. Discrimination of same sequences

Across all conditions, activation was observed in bilateral SMA, premotor cortex, insula, inferior frontal gyri, inferior parietal cortex, and superior temporal gyri (see Fig. 2 and Supplementary Table 4). Verbal sequences elicited greater activity in bilateral superior and middle temporal gyri, and left inferior frontal gyrus, than rhythm sequences (see Supplementary Table 5). There were no other significant main effects or interactions.

3.2.4. Discrimination of different sequences

The activation loci for this contrast are reported in Supplementary Tables 6-7 for completeness, but this stage is associated with detection of acoustic changes and response preparation, therefore it is likely to encompass many cognitive processes that are not of interest.

3.3. fMRI analyses: region of interest analyses

Most ROIs that showed significant main effects or interactions in ROI analyses were already identified by whole brain analyses, with the exception of the right putamen, which showed a significant interaction

Table 2
Stereotaxic locations of peak voxels during encoding: Interaction between rhythm and verbal conditions (beat – nonbeat rhythms) – (chunked – unchunked letter sequences).

Brain area	Cluster*	voxel t	pFDR (cluster)	x	y	z
L insula	cluster 2	3.82	<.01	−27	18	−6
R pallidum	cluster 2	3.85	<.01	18	3	−3
L putamen	cluster 1	4.74	<.01	−24	9	3

Cluster volumes: 1 = 121 voxels, 2 = 46 voxels.

Table 3
Stereotaxic locations of peak voxels during encoding: beat – nonbeat rhythm conditions.

Brain area	Cluster*	voxel t	pFDR (cluster)	x	y	z
L putamen	cluster 1	5.87	<.005	−24	3	0
R putamen	cluster 2	5.18	<.005	21	9	6
R putamen	cluster 2	4.39	<.005	24	6	−3

Cluster volumes: 1 = 178 voxels, 2 = 140 voxels.

Table 1
Stereotaxic locations of peak voxels during encoding: rhythm condition – verbal condition.

Brain area	Brodmann area	Cluster*	voxel t	pFDR (cluster)	x	y	z
L inferior frontal gyrus, p. triangularis	BA 47	cluster 3	3.38	<.001	−48	21	0
L inferior frontal gyrus, p. triangularis	BA 45	cluster 3	3.92	<.001	−57	21	3
R inferior frontal gyrus, p. triangularis	BA 45	cluster 6	4.23	<.05	42	30	9
R inferior frontal gyrus, p. triangularis	BA 47	cluster 6	3.65	<.05	36	30	3
L inferior frontal gyrus, p. opercularis	BA 6	cluster 3	4.13	<.001	−48	9	15
R inferior frontal gyrus p. opercularis	BA 44	cluster 1	5.89	<.001	54	12	24
R inferior frontal gyrus p. opercularis	BA 6	cluster 1	5.02	<.001	54	9	15
L insula		cluster 3	4.07	<.001	−36	3	−3
L insula		cluster 3	4.46	<.001	−45	0	6
L insula		cluster 3	4.38	<.001	−39	0	6
L supplementary motor area	BA 6	cluster 4	4.32	<.01	−9	−3	66
R supplementary motor area	BA 6	cluster 4	4.24	<.01	6	3	66
R premotor cortex	BA 6	cluster 1	3.46	<.001	48	3	42
L Rolandic operculum		cluster 3	4.58	<.001	−51	3	12
R middle temporal gyrus	BA 21	cluster 5	4.24	<.05	48	−36	51
L amygdala	BA 36	cluster 3	4.31	<.001	−27	0	−24
L pallidum		cluster 3	4.57	<.001	−21	6	3
L pallidum		cluster 3	4.36	<.001	−15	0	−9
R pallidum		cluster 2	5.70	<.001	18	3	−6
L putamen		cluster 3	4.64	<.001	−18	12	3
R putamen		cluster 2	5.48	<.001	21	3	15

Cluster volumes: 1 = 322 voxels, 2 = 286 voxels, 3 = 328 voxels, 4 = 96 voxels, 5 = 61 voxels, 6 = 56 voxels.

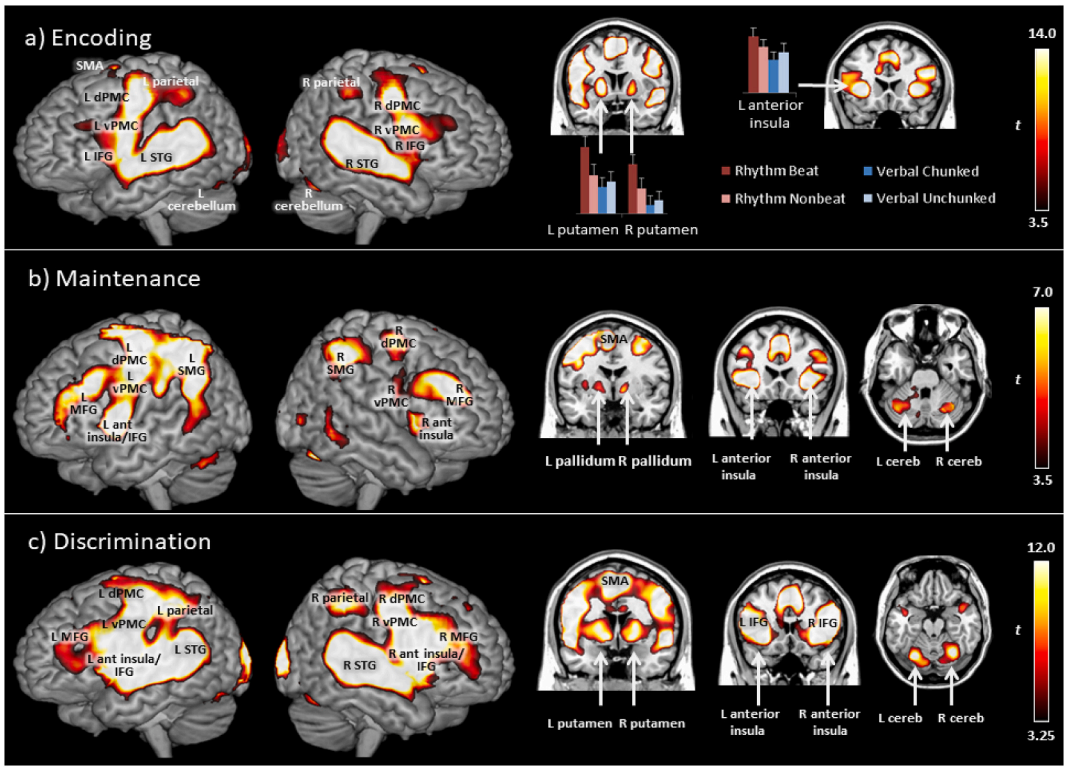


Fig. 2. Significant activation for encoding (a), maintenance (b), and discrimination phases (c). The brain renderings show whole brain contrasts for all conditions (beat rhythms, nonbeat rhythms, chunked letter sequences, and unchunked letter sequences) versus rest. Inset bar graphs show activation levels for each condition in brain regions that showed significant interactions between conditions. (L, left; R, right; cereb, cerebellum; dPMC, dorsal premotor cortex; vPMC, ventral premotor cortex; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; SMG, supramarginal gyrus; STG, superior temporal gyrus; SMA, supplementary motor area).

between temporal regularity and stimulus type (t (1,17) = 4.14, p = .001). As shown in Fig. 2, this stems from greater activation for beat than nonbeat rhythms, but no difference between chunked and unchunked verbal sequences. A full list of regions is shown in Supplementary Table 8).

4. Discussion

The goal of the study was to compare the neural networks involved in short-term memory for verbal and rhythmic stimuli. In general, similar to previous short-term memory studies, rhythm and verbal networks were characterized by many areas of overlap. During encoding, all conditions activated frontoparietal areas, premotor and auditory cortices, and the basal ganglia and cerebellum. Activity in auditory areas was greater for verbal than rhythm stimuli, and activity in motor areas was greater for rhythm than verbal stimuli. During maintenance, activity was observed in similar areas to encoding, apart from reduced activity in auditory cortex. During discrimination, activity was observed in similar areas to encoding, but generally with greater spatial extent. Greater auditory activity was observed for verbal than rhythmic stimuli. For maintenance and discrimination, no interactions between stimulus type and temporal structure were observed. However, for encoding, greater activity in the basal ganglia was observed for beat than nonbeat rhythms, but no difference was found for chunked and unchunked verbal sequences. This interaction supports the idea that beat perception and chunking may not be the same process.

Overall, the activity during the different stages is consistent with previous research that suggests a 'core' short-term memory network that includes inferior frontal areas, parietal areas, pre-SMA, PMC, cerebellum, and left basal ganglia (Rottschy et al., 2012). Particularly for auditory stimuli, fMRI and MEG studies reveal inferior frontal, dorsolateral prefrontal, parietal, and auditory cortices are active during encoding and maintenance of both verbal and tonal sequences (Albouy et al., 2013, 2018; Nolden et al., 2013). Further, individuals with amusia, who experience deficits in memory for tonal sequences have less activity, as well as structural variations, in auditory and inferior frontal areas, further implicating these areas in auditory memory (Albouy et al., 2013). This network appears to be sensitive to stimulus materials; tonal sequences elicit greater activity than verbal sequences (Albouy et al., 2018). Similarly, we find that the short-term memory network differentially activates for rhythm and verbal sequences, and is also sensitive to structural differences (beat structure or chunks). The influence of temporal structure is important for frameworks of auditory short-term memory, such as Baddeley (2000) model, which currently does not take temporal regularity into account.

Activation differences between rhythm and verbal stimuli were mainly observed during the encoding and discrimination stages. In both stages, verbal stimuli compared to rhythmic stimuli elicited greater activity in the auditory cortex. However, an absence of interactions with temporal regularity suggests that observed auditory activation differences are driven by basic stimulus differences between verbal and rhythmic stimuli, not their temporal structure. There is greater acoustic variation in the string of different letters heard during verbal stimuli, than in the repetition of single letters heard during rhythm stimuli. Future work could keep the acoustic variation consistent across stimulus types, perhaps by constructing rhythms with letter variation within the sequence. Here, however, we found it necessary to use single-letter strings for the rhythm condition to help participants focus on the timing of intervals, rather than on the letter elements. This was especially important as the focus between letter elements and temporal intervals switched between imaging blocks, and would be harder to track without stimulus differences to cue participants.

During encoding only, rhythmic stimuli elicited greater activity (relative to verbal stimuli) in the basal ganglia (putamen and pallidum),

SMA, and bilateral inferior frontal cortex. This is broadly consistent with the idea that motor areas may play a greater role in the processing of rhythmic auditory information than in processing auditory identity. Relatedly, a meta-analysis (Rottschy et al., 2012) contrasting verbal (e.g., letters, words) and non-verbal (e.g., figures, objects) stimuli found left Brodmann areas 44 and 45 are more closely associated with verbal stimuli. By their criteria, both the rhythm and letter conditions in the current study would be 'verbal', and indeed both rhythm and letter conditions activate this area. However, we find even greater activity for rhythm than verbal stimuli. Therefore, the focus on the temporal aspects of the stimuli, as opposed to the identities of the letters, appears to recruit inferior frontal cortex to a greater degree, perhaps because of its role in temporal sequencing (Clos et al., 2013; Gelfand and Bookheimer, 2003).

During encoding, an interaction was observed in the basal ganglia bilaterally, with beat rhythms inducing significantly greater activity than nonbeat rhythms, and also greater activity than either chunked or unchunked verbal stimuli, which did not differ from each other. The presence of a significant interaction in the basal ganglia suggests that beat perception is not simply a form of chunking, and that beat perception recruits the basal ganglia specifically. This is consistent with other neuroimaging work that associates the basal ganglia with temporal structure, particularly in discrimination tasks (Geiser et al., 2012; Grahn and Brett, 2007; Teki et al., 2011). The presence of a significant interaction during encoding but not maintenance or discrimination suggests a few possibilities about the role of the basal ganglia in beat perception, although it must also be emphasized that any interpretation of null results must be treated with caution. It may be that the basal ganglia are important for an initial reorganization of rhythmic stimuli that occurs when a beat is perceived—intervals are recoded from a series of separate durations to onsets relative to the beat. The motivation to recode the intervals is high in a discrimination task, when performance depends on an accurate representation, and beat perception can enhance performance. It is also possible, contrary to the conclusions of previous research (Grahn and Rowe, 2013), that the basal ganglia are important for the perception of the beat, rather than maintenance or internal generation of the beat during the maintenance stage. We feel that this interpretation, though possible, is less likely, as subthreshold differences between beat and nonbeat rhythm activity were observed during the maintenance stage.

We did not find any clear differences between verbal chunked and unchunked stimuli. Previous work has reported greater right IFG activation for chunked than unchunked sequences, although the sequences were visual (Henson et al., 2003). Kalm et al. (2012) found differences in parietal cortex activity for chunked compared to unchunked auditory letter sequences, but only when the sequences were nine letters long, not six, and were thus designed to exceed auditory short-term memory span. Our sequences were four and eight letters long, and were not intended to reliably exceed our participants' spans; we wanted to keep overall performance high, as the task was merely intended to induce participants to attend to the stimuli, rather than to serve as a correlate of interest. We assessed behavioral performance with discrimination, not recall, which is usually used to determine span. As performance was high (although not perfect) in both the chunked and unchunked conditions (>90%), the task may not have been taxing enough to elicit the neural activation differences between chunked and unchunked conditions observed when span is exceeded.

An alternative explanation for our null effect between chunked and unchunked sequences is that participants attempted to chunk *all* verbal sequences, regardless of timing. We induced chunking by presenting the chunked verbal sequences in groups of 2 or 4 letters, separated by a silent pause. The unchunked sequences were intended to disrupt chunking by presenting letters with random inter-onset intervals, thus making it difficult to parse the sequence into equal chunks during encoding. However, participants may have internally recoded the unchunked sequences into chunks, allowing improved performance on the task (a follow-up experiment on an independent sample revealed that both chunked and unchunked sequences were susceptible to internal chunking; see supplementary material). Internal chunking of irregular sequences may also rely on neural resources used to hold the chunked sequences in short-term memory. If true, our imaging contrast of chunked > unchunked verbal sequences would not reveal differences between these two conditions. However, this possibility still supports the conclusion that beat perception differs from chunking. If chunking and beat perception were the same process, and if participants used chunking for all verbal sequences, then activity to beat-based rhythms would be similar to verbal sequences. This is not what we found. We found that rhythm, compared to verbal sequences, elicited greater activity in the basal ganglia, but driven specifically by increased activity in the putamen for rhythms with a beat (see Fig. 2a). Thus, beat-related basal ganglia activity differed from chunking-related basal ganglia activity, even if participants utilized chunking on all verbal sequences. Finally, although we observed clear increases in neural activation to beat perception relative to chunking, it is important to note that we only tested a perceptual task. For motor *production*, in particular motor sequencing output, a role for the basal ganglia in chunking has been proposed: The basal ganglia may link sequences of actions together into chunks. Basal ganglia neurons appear to represent chunks by firing preferentially at the beginning and end of action sequences (Graybiel, 2008; Thorn and Graybiel, 2010; Yin and Knowlton, 2006), and disruption of this firing impairs sequence learning (Jin and Costa, 2010). Moreover, individuals with basal ganglia dysfunction, such as in Parkinson's disease (Tremblay et al., 2010) or stroke (Boyd et al., 2009), are less likely to chunk motor sequence outputs. Therefore, the motor sequencing literature suggests that, for production, there may be neural overlap between chunking of movement sequences and beat processing in the basal ganglia. Further research to compare chunking in motor sequences and rhythmic sequence production versus perception may support the idea that beat processing and chunking engage common networks when motor production is required.

Overall, the data suggest that beat perception and verbal chunking do not rely on a common mechanism. Although maintenance and encoding of rhythm and verbal sequences overlap in many brain areas, rhythmic sequences with an underlying beat elicit greater activation in the basal ganglia than non-beat rhythms, and chunked or unchunked verbal sequences. Because of this difference, it is unlikely that beat perception is reduceable to temporal chunking.

Authorship statement

Joshua D. Hoddinott: Writing – Original Draft, Review & Editing, Validation, Visualization. Dirk Schuit: Methodology, Formal Analysis, Investigation, Writing – Original Draft. Jessica A. Grahn: Conceptualization, Methodology, Formal Analysis, Investigation, Resources, Writing – Original Draft, Review & Editing, Visualization, Supervision, Funding Acquisition.

Funding

This work was supported by the Medical Research Council [Cost Code U.1055.01.001.00001.01]; and the NSERC Discovery Grant [RGPIN-2016-05834].

Acknowledgements

Thank you to Matt Davis for assistance and advice about stimulus creation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.neuropsychologia.2021.108080.

References

Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., Tillmann, B., 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. Brain 136 (5), 1639–1661. https://doi.org/10.1093/brain/awt082.

Albouy, P., Peretz, I., Bermudez, P., Zatorre, R.J., Tillmann, B., Caclin, A., 2018. Specialized neural dynamics for verbal and tonal memory: fMRI evidence in congenital amusica. Hum. Brain Mapp. 40 (3), 855–867. https://doi.org/10.1002/hbm.24416.

Araneda, R., Renier, L., Ebner-Karestinos, D., Dricot, L., De Volder, A.G., 2016. Hearing, feeling or seeing a beat recruits a supramodal network in the auditory dorsal stream. Eur. J. Neurosci. 45 (11), 1439–1450. https://doi.org/10.1111/ejn.13349.

Awh, E., Jonides, J., Smith, E.E., Schumacher, E.H., Koeppe, R.A., Katz, S., 1996. Dissociation of storage and rehearsal in verbal working memory: evidence from positron emission tomography. Psychol. Sci. 7 (1), 25–31. https://doi.org/10.1111/j.1467-9280.1996.tb00662.x.

Baddeley, A., 2000. The episodic buffer: a new component of working memory? Trends Cognit. Sci. 4 (11), 417–423. https://doi.org/10.1016/S1364-6613(00)01538-2.

Baddeley, A., 2003. Working memory: looking back and looking forward. Nat. Rev. Neurosci. 4 (10), 829–839. https://doi.org/10.1038/nrn1201nrn1201 ([pii]).

Baddeley, A.D., Hitch, G., 1974. Working memory. In: Bower, G.H. (Ed.), The Psychology of Learning and Motivation: Advances in Research and Theory, vol. 8. Academic Press, New York, pp. 47–89.

Bengtsson, S.L., Ullén, F., Henrik Ehrsson, H., Hashimoto, T., Kito, T., Naito, E., Sadato, N., 2009. Listening to rhythms activates motor and premotor cortices. Cortex 45 (1), 62–71. https://doi.org/10.1016/j.cortex.2008.07.002.

Bo, J., Seidler, R., 2009. Visuospatial working memory capacity predicts the organization of acquired explicit motor sequences. J. Neurophysiol. 101, 3116–3125. https://doi.org/10.1152/jn.00006.2009.

Bower, G.H., Winzenz, D., 1969. Group Structure, coding, and memory for digit series. J. Exp. Psychol. 80 (2), 1–17. https://doi.org/10.1037/h0027249.

Boyd, L.A., Edwards, J.D., Siengsukon, C.S., Vidoni, E.D., Wessel, B.D., Linsdell, M.A., 2009. Motor sequence chunking is impaired by basal ganglia stroke. Neurobiol. Learn. Mem. 92 (1), 35–44 ([pii])10.1016/j.nlm.2009.02.009-S1074-7427(09)00057-4.

Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008a. Listening to musical rhythms recruits motor regions of the brain. Cerebr. Cortex 18 (12), 2844–2854. https://doi.org/10.1093/cercor/bhn042.

Chen, J.L., Penhune, V.B., Zatorre, R.J., 2008b. Moving on time: brain network for auditory-motor synchronization is modulated by rhythm complexity and musical training. J. Cognit. Neurosci. 20 (2), 226–239. https://doi.org/10.1162/jocn.2008.20018.

Chen, S.H., Desmond, J.E., 2005. Temporal dynamics of cerebro-cerebellar network recruitment during cognitive task. Neuropsychologia 43 (9), 10. https://doi.org/10.1016/j.neuropsychologia.2004.12.015.

Chumbley, J.R., Friston, K.J., 2009. False discovery rate revisited:FDR and topological inference using Gaussian random fields. Neuroimage 44 (1), 8. https://doi.org/10.1016/j.neuroimage.2008.05.021.

Clos, M., Amunts, K., Laird, A.R., Fox, P.T., Eickhoff, S.B., 2013. Tackling the multifunctional nature of Broca's region meta-analytically: Co-activation-based parcellation of area 44, 0 Neuroimage 83, 174–188. https://doi.org/10.1016/j.neuroimage.2013.06.041.

Cowan, N., 1999. An Embedded-Processes Model of working memory. In: Miyake, A., Shah, Priti (Eds.), Models of Working Memory: Mechanisms of Active Maintenance and Executive Control. Cambridge University Press, Cambridge, pp. 62–101.

Dowling, W.J., 1973. Rhythmic groups and subjective chunks in memory for melodies. Percept. Psychophys. 14, 37–40. https://doi.org/10.3758/BF03198614.

Ericsson, K.A., Chase, W.G., Faloon, S., 1980. Acquisition of a memory skill. Science 208 (4448), 2. https://doi.org/10.1126/science.7375930.

Fiez, J.A., Raife, E.A., Balota, D.A., Schwarz, J.P., Raichle, M.E., Petersen, S.E., 1996. A positron emission tomography study of the short-term maintenance of verbal information. J. Neurosci. 16 (2), 808–822 doi:JNEUROSCI.16-02-00808.1996.

Frankish, C., 1989. Perceptual organization and precategorical acoustic storage. J. Exp. Psychol. Learn. Mem. Cogn. 15 (3), 469–479. https://doi.org/10.1037/0278-7393.15.3.469.

Frankish, C., 1985. Modality-specific grouping effects in short-term memory. J. Mem. Lang. 24 (2), 200–209. https://doi.org/10.1016/0749-596X(85)90024-5.

Gaab, N., Gaser, C., Zaehle, T., Jaencke, L., Schlaug, G., 2003. Functional anatomy of pitch memory–an fMRI study with sparse temporal sampling. Neuroimage 19, 1417–1426. https://doi.org/10.1016/s1053-8119(03)00224-6.

Geiser, E., Notter, M., Gabrieli, J.D.E., 2012. A corticostriatal neural system enhances auditory perception through temporal context processing. J. Neurosci. 32 (18), 6177–6182. https://doi.org/10.1523/Jneurosci.5153-11.2012.

Gelfand, J.R., Bookheimer, S.Y., 2003. Dissociating neural mechanisms of temporal sequencing and processing phonemes. Neuron 38 (5), 831–842. https://doi.org/10.1016/s0896-6273(03)00285-x.

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

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

Grahn, J.A., 2009. The role of the basal ganglia in beat perception: neuroimaging and neuropsychological investigations. Ann. N. Y. Acad. Sci. 1169, 35–45. https://doi.org/10.1111/j.1749-6632.2009.04553.x.

Grahn, J.A., 2012. See what I hear? Beat perception in auditory and visual rhythms. Exp. Brain Res. 220 (1), 51–61. https://doi.org/10.1007/s00221-012-3114-8.

Grahn, J.A., Brett, M., 2007. Rhythm perception in motor areas of the brain. J. Cognit. Neurosci. 19 (5), 893–906. https://doi.org/10.1162/jocn.2007.19.5.893.

Grahn, J.A., Rowe, J.B., 2009. Feeling the beat: premotor and striatal interactions in musicians and non-musicians during beat processing. J. Neurosci. 29 (23), 7540–7548. https://doi.org/10.1523/JNEUROSCI.2018-08.2009.

Grahn, J.A., Rowe, J.B., 2013. Finding and feeling the musical beat: striatal dissociations between detection and prediction of regularity. Cerebr. Cortex 23 (4), 913–921. https://doi.org/10.1093/cercor/bhs083.

Grahn, J.A., Schuit, D., 2012. Individual differences in rhythmic abilities: behavioural and neuroimaging investigations. Psychomusicology: Music, Mind, & Brain 22, 105–121. https://doi.org/10.1037/a0031188.

Graybiel, A.M., 2008. Habits, rituals, and the evaluative brain. Annu. Rev. Neurosci. 31, 359–387. https://doi.org/10.1146/annurev.neuro.29.051605.112851.

Gruber, O., von Cramon, D.Y., 2003. The functional neuroanatomy of human working memory revisited: evidence from 3-T fMRI studies using classical domain-specific interference tasks. Neuroimage 19 (3), 12. https://doi.org/10.1016/s1053-8119(03)00089-2.

Hall, D., Gathercole, S.E., 2011. Serial recall of rhythms and verbal sequences: impacts of concurrent tasks and irrelevant sound. Q. J. Exp. Psychol. 64 (8), 1580–1592. https://doi.org/10.1080/17470218.2011.564636.

Heard, M., Lee, Y.S., 2020. Shared neural resources of rhythm and syntax: an ALE meta-analysis. Neuropsychologia 137. https://doi.org/10.1016/j.neuropsychologia.2019.107284.

Henson, R., Hartley, T., Burgess, N., Hitch, G., Flude, B., 2003. Selective interference with verbal short-term memory for serial order information: a new paradigm and tests of a timing-signal hypothesis. The Quarterly Journal of Exerimental Psychology 56A (8), 1307–1334. https://doi.org/10.1080/02724980244000747.

Henson, R.N.A., Burgess, N., Frith, C.D., 2000. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. Neuropsychologia 38, 426–440. https://doi.org/10.1016/S0028-3932(99)00098-6.

Hickok, G., Buchsbaum, B., Humphries, C., Muftuler, T., 2003. Auditory-motor interaction revealed by fMRI: speech, music, and working memeory in area spt. J. Cognit. Neurosci. 15, 673–682. https://doi.org/10.1162/089892903322307393.

Hitch, G.J., Burgess, N., Towse, J.N., Culpin, V., 1996. Temporal grouping effects in immediate recall: a working memory analysis. Q. J. Exp. Psychol. 49A, 140–158 doi:/10.1080%2F713755609.

Jin, X., Costa, R.M., 2010. Start/stop signals emerge in nigrostriatal circuits during sequence learning. Nature 466 (7305), 457–462. https://doi.org/10.1038/nature09263.

Kalm, K., Davis, M.H., Norris, D., 2012. Neural Mechanisms underlying the grouping effect in short-term memory. Hum. Brain Mapp. 33 (7), 13. https://doi.org/10.1002/hbm.21308.

Kennerley, S.W., Sakai, K., Rushworth, M.F.S., 2004. Organization of action sequences and the role of the pre-SMA. J. Neurophysiol. 91, 978–993. https://doi.org/10.1152/jn.00651.2003.

Koelsch, S., Schulze, K., Sammler, D., Fritz, T., Muller, K., Gruber, O., 2009. Functional architecture of verbal and tonal working memory: an FMRI study. Hum. Brain Mapp. 30 (3), 859–873. https://doi.org/10.1002/hbm.20550.

Kung, S.J., Chen, J.L., Zatorre, R.J., Penhune, V.B., 2013. Interacting cortical and basal ganglia networks underlying finding and tapping to the musical beat. J. Cognit. Neurosci. 25 (3), 401–420. https://doi.org/10.1162/jocn_a_00325.

Large, E.W., Palmer, C., 2002. Perceiving temporal regularity in music. Cognit. Sci. 26, 1–37. https://doi.org/10.1016/S0364-0213(01)00057-X.

Li, Q., Liu, G., Wei, D., Liu, Y., Yuan, G., Wang, G., 2019. Distinct neuronal entrainment to beat and meter: revealed by simultaneous EEG-fMRI. Neuroimage 194, 128–135. https://doi.org/10.1016/j.neuroimage.2019.03.039.

Miller, G.A., 1956. The magical number seven plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2), 81–97. https://doi.org/10.1037/h0043158.

Nettl, B., 2000. An ethnomusicologist contemplates universals in musical sound and musical culture. In: Wallin, N.L., Merker, B., Brown, S. (Eds.), The Origins of Music. MIT Press, Cambridge, MA, pp. 463–472.

Nolden, S., Grimault, S., Guimond, S., Lefebvre, C., Bermudez, P., Jolicoeur, P., 2013. The retention of simultaneous tones in auditory short-term memory: a magnetoencephalography study. Neuroimage 82, 384–392. https://doi.org/10.1016/j.neuroimage.2013.06.002.

Parncutt, R., 1994. A perceptual model of pulse salience and metrical accent in musical rhythms. Music Perception 11, 409–464. https://doi.org/10.2307/40285633.

Patel, A.D., Iversen, J.R., Chen, Y., Repp, B.H., 2005. The influence of metricality and modality on synchronization with a beat. Exp. Brain Res. 163, 226–238. https://doi.org/10.1007/s00221-004-2159-8.

Penny, W., Holmes, A.P., 2003. Random effects analysis. In: Frackowiack, R.S.J., Friston, K.J., Frith, C.D., Dolan, R., Price, C.J., Ashburner, J., Penny, W., Zeki, S. (Eds.), Human Brain Function II, 2 ed. Elsevier Academic Press, San Diego.

Plancher, G., Lévêque, Y., Fanuel, L., Piquandet, G., Tillmann, B., 2018. Boosting maintenance in working memory with temporal regularities. J. Exp. Psychol. Learn. Mem. Cognit. 44 (5), 812–818.

Rauschecker, J.P., Scott, S.K., 2009. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat. Rev. Neurosci. 12 (6), 718–724. https://doi.org/10.1038/nn.2331.

Rottschy, C., Langner, R., Dogan, I., Reets, K., Laird, A.R., Schulz, J.B., Eickhoff, S.B., 2012. Modelling neural correlates of working memory: a coordinate-based meta-analysis. Neuroimage 60 (1), 16. https://doi.org/10.1016/j.neuroimage.2011.11.050.

Saito, S., 1994. What effect can rhythmic finger tapping have on the phonological similarity effect? Mem. Cognit. 22 (2), 181–187. https://doi.org/10.3758/BF03208889.

Saito, S., 1998. Rhythmic information in working memory: effects of concurrent articulation on reproduction of rhythms. Jpn. Psychol. Res. 40 (1), 10–18. https://doi.org/10.1111/1468-5884.00070.

Saito, S., 2001. The phonological loop and memory for rhythms: an individual differences approach. Memory 9 (4–6), 313–322. https://doi.org/10.1080/09658210143000164.

Sakai, K., Kitaguchi, K., Hikosaka, O., 2003. Chunking during human visuomotor sequence learning. Exp. Brain Res. 152, 229–242. https://doi.org/10.1007/s00221-003-1548-8.

Schaefer, R.S., Vlek, R.J., Desain, P., 2011. Decomposing rhythm processing: electroencephalography of perceived and self-imposed rhythmic patterns. Psychol. Res. 75, 95–106. https://doi.org/10.1007/s00426-010-0293-4.

Schulze, K., Koelsch, S., 2012. Working memory for speech and music. Ann. N. Y. Acad. Sci. 1252, 229–236. https://doi.org/10.1111/j.1749-6632.2012.06447.x.

Teki, S., Grube, M., Kumar, S., Griffiths, T.D., 2011. Distinct neural substrates of duration-based and beat-based auditory timing. J. Neurosci. 31 (10), 3805–3812. https://doi.org/10.1523/JNEUROSCI.5561-10.2011.

Thorn, C.A., Graybiel, A.M., 2010. Pausing to regroup: thalamic gating of cortico-basal ganglia networks. Neuron 67 (2), 175–178 doi:10.1016/j.neuron.2010.07.010 S0896-6273(10)00548-9 [pii].

Tremblay, P.L., Bedard, M.A., Langlois, D., Blanchet, P.J., Lemay, M., Parent, M., 2010. Movement chunking during sequence learning is a dopamine-dependant process: a study conducted in Parkinson's disease. Exp. Brain Res. 205 (3), 375–385. https://doi.org/10.1007/s00221-010-2372-6.

Trost, W., Frühholz, S., Schön, D., Labbé, C., Pichon, S., Grandjean, D., Vuilleumier, P., 2014. Getting the beat: entrainment of brain activity by musical rhythm and pleasantness. Neuroimage 103, 55–64. https://doi.org/10.1016/j.neuroimage.2014.09.009.

van Noorden, L., Moelants, D., 1999. Resonance in the perception of musical pulse. J. N. Music Res. 28, 43–66. https://doi.org/10.1076/jnmr.28.1.43.3122.

Verwey, W.B., Abrahamse, E.L., Jimenez, L., 2009. Segmentation of short key sequences does not spontaneously transfer to other sequences. Hum. Mov. Sci. (28), 348–361. https://doi.org/10.1016/j.humov.2008.10.004.

Vikene, K., Skeie, G.-O., Specht, K., 2019. Abnormal phasic activity in saliency network, motor areas, and basal ganglia in Parkinson's disease during rhythm perception. Hum. Brain Mapp. 40 (3), 916–927. https://doi.org/10.1002/hbm.24421.

Vuust, P., Roepstorff, A., Wallentin, M., Mouridsen, K., Ostergaard, L., 2006. It don't mean a thing. Keeping the rhythm during polyrhythmic tension, activates language areas (BA47). Neuroimage 31 (2), 832–841. https://doi.org/10.1016/j.neuroimage.2005.12.037.

Yin, H.H., Knowlton, B.J., 2006. The role of the basal ganglia in habit formation. Nat. Rev. Neurosci. 7 (6), 464–476. https://doi.org/10.1038/nrn1919.

Zatorre, R.J., Evans, A.C., Meyer, E., 1994. Neural mechanisms underlying melodic perception and memory for pitch. J. Neurosci. 14 (4), 1908–1919. https://doi.org/10.1523/JNEUROSCI.14-04-01908.