

Neural Mechanisms Underlying the Grouping Effect in Short-Term Memory

Kristjan Kalm*, Matthew H. Davis, and Dennis Norris

MRC Cognition and Brain Sciences Unit, Cambridge, United Kingdom

Abstract: Dividing auditory sequence into groups, or imposing rhythmic, tonal, or spatial structure during presentation, improves recall performance. Several competing computational models have been proposed to account for these effects, but little is known about the neural correlates of grouping and hence the representations that encode grouped sequences. The present study used functional magnetic resonance imaging (fMRI) to compare the auditory encoding of grouped and ungrouped lists of sub-span (six letters) and supra-span (nine letters) length in an immediate serial recall (ISR) task. Analysis of activation revealed an extensive premotor and prefrontal network, which was significantly less active when short-term memory (STM) span was exceeded during encoding. Only primary auditory cortex showed an increase in activation when memory span was exceeded. Comparison of activation for grouped and ungrouped lists showed that during the subspan phase bilateral planum temporale showed less activation for grouped stimuli, while during the supra-span phase supramarginal and inferior parietal areas were more active for grouped lists. The magnitude of both temporal and parietal activations predicted enhanced recall of grouped lists. Thus neural signatures of grouping seem to reflect more structured processing in parietal areas instead of reliance on perceptual-auditory processing in temporal regions. *Hum Brain Mapp* 33:1634–1647, 2012. © 2011 Wiley Periodicals, Inc.

Key words: short-term memory; fMRI; grouping; span

INTRODUCTION

The effect of grouping has fascinated psychologists for decades. Almost a century ago, in 1915, Henry Foster Adams at the University of Michigan wanted to know which classical feet of poetic metre were the easiest to remember. He published his findings as “A note on the effect of rhythm on memory” in *Psychological Review* [Adams, 1915] reporting significant differences between rhythms and noting that “it is entirely natural that the three-part rhythms should be the best in this part of the

experiment, for most of us have been trained from our earliest days in the grade schools to group numbers by hundreds, thousands, millions, etc.” Here we explore the neural mechanisms supporting the advantage of grouping for verbal sequences.

It is a commonplace observation that recall is improved when the material to be remembered is in some way organized or grouped. In short-term memory (STM) tasks, several means of grouping have been studied (temporal, rhythmical, and spatial): the most usual finding is an improvement in recall with grouping [Broadbent and Broadbent, 1981; Frankish, 1985; Henson et al., 2000; Hitch et al., 1996; Ryan, 1969; Wickelgren, 1964].

The use of such grouping strategies is apparent in many everyday tasks that involve memory for sequences. For example, people will spontaneously divide arbitrary sequences, such as phone numbers into subgroups and may look for perceptual and semantic cues in longer sequences [e.g. common abbreviations such as BBC, NASA, WYSIWYG; Henson, 2001].

*Correspondence to: Kristjan Kalm, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7E, UK. E-mail: kristjan.kalm@mrc-cbu.cam.ac.uk

Received for publication 26 July 2010; Revised 23 February 2011; Accepted 23 February 2011

DOI: 10.1002/hbm.21308

Published online 7 July 2011 in Wiley Online Library (wileyonlinelibrary.com).

Since the advent of cognitive psychology the grouping effect has been mostly investigated with the immediate serial recall (ISR) or “memory-span” task—a task that remains the dominant empirical tool behind contemporary theories of STM, such as Baddeley and Hitch’s [1974] working memory framework. In the immediate serial recall (ISR) task, participants have to repeat aloud a presented sequence of letters or digits without a delay, i.e. no rehearsal or maintenance is included. When items in such a task are temporally grouped (e.g. inserting a pause after every three items) significant improvement of recall is observed [Hitch et al., 1996; Ryan, 1969; Wickelgren, 1964].

The effect of grouping raises important questions about STM mechanisms: how can minor modifications to the manner in which item sequences are presented produce a dramatic improvement in memory capacity? Like many behavioral phenomena, this apparently straightforward effect masks a rich and complex host of issues. Several computational models have been proposed to address the effect [Botvinick and Plaut, 2006; Brown et al., 2000; Burgess and Hitch, 2006; Henson, 1998; Page and Norris, 1998], but behavioral data has thus far failed to distinguish between these competing accounts and there have been few investigations of the neural basis of temporal grouping in STM.

Here we used functional brain imaging to provide additional insights into the neural systems that support increased STM capacity for grouped lists of letters. By assessing neural correlates of grouping in STM we can obtain evidence to link grouping effects to specific neural representations; particularly whether these are perceptual or at a higher level. Perhaps due to the difficulties involved in presenting spoken materials and recording vocal responses in fMRI, there are no previous studies that have addressed this issue in conditions comparable to those used in standard behavioral studies. One previous study that investigated grouping effects with fMRI used visual presentation of sub-span letter sequences and item-order probe lists rather than verbal recall [Henson et al., 2000]. In that study, grouped lists presented with a brief pause (e.g. RBTMDS vs. RBT-MDS) evoked reduced activity during probe sequences in left dorsolateral premotor cortex (BA6). The use of sub-span stimuli (which resulted in near-ceiling behavioral performance for both grouped and ungrouped conditions) and visual presentation [for which grouping effects are significantly less pronounced, Frankish, 1985, 1989; Hitch et al., 1996] likely reduced the power of Henson et al.’s study to capture grouping effects, and the absence of behavioral effects of grouping on performance make it difficult to assess whether differences in neural activity are related to behavioral effects of grouping.

We therefore adopted the ISR paradigm, in which participants have to verbally recall the stimuli immediately after presentation. The absence of a delay period ensures that rehearsal and maintenance processes are minimized. Importantly, participants need to recall the items in the order they were presented. ISR performance thus reflects

basic demands on STM, when neither maintenance nor manipulation of items (comparison, reorganization etc) is necessary. It is also the easiest way to examine perceptual influences (grouping) on STM for order with one task.

We used this approach to tackle two important questions using fMRI. First, we investigated the interaction between grouping and memory span. Our understanding of behavioral grouping effects is necessarily based on error patterns: participants need to recall lists that exceed the immediate STM span [supra-span lists, usually more than seven items; Miller, 1956] as, by definition, subspan lists (e.g. six items) produce few errors. Traditionally, experiments have used supra-span nine-item lists presented sequentially in three letter triplets (e.g. FNH JQS ZLY). The first three items are perceptually identical in both ungrouped and grouped lists and therefore cannot produce grouping effects. Any effect of grouping on the second triplet of a six-item list cannot be measured behaviorally because performance on six items is almost always perfect. It has also been proposed [Henson et al., 1996] that the need for more short-term storage drives grouping effects, which only becomes evident when subjects have to recall items from supra-span lists. Thus, behavioral grouping effects are confounded with load effects, and it is impossible to tell how grouped stimuli are processed before STM span is exceeded. fMRI, however, allows us to look at neural responses at all stages of encoding. In particular, by using a partial-trial method in which six- and nine-item sequences are randomly intermixed, we can separate neural activity due to the subspan and supra-span portion of nine-item grouped and ungrouped lists.

Additionally, we investigated the role of perceptual processing in the grouping effect. Grouping effects in the verbal domain are dependent on the modality of the stimuli and vary in size according to the grouping cue used. Frankish [1985] showed that recall accuracy is much higher for lists that are segmented by means of extended pauses when presentation is auditory, but not when it is visual. Hitch et al. [1996] observed that with visual presentation, grouping effects are eliminated by concurrent articulatory suppression. Frankish [1989] also demonstrated that grouping effects can be obtained for auditory sequences that are structured by voice or spatial location. Such findings might suggest that grouping effects arise before phonological recoding and might be observed in early auditory encoding areas in the superior temporal gyrus, and elicited by the range of spatiotemporal features of an auditory stimulus that are processed in these regions [Buchsbaum et al., 2005a,b; Griffiths and Warren, 2002]. Conversely, it might be that systems involved in recoding auditory stimuli into higher level representations (e.g. articulatory codes in inferior parietal and prefrontal regions [Buchsbaum, 2008; Hickok and Poeppel, 2007]; operate more effectively for grouped sequences.

These arguments lead to two specific hypotheses that we test in the present study: if grouping effects in verbal STM reflect facilitated perceptual processing we would expect to see evidence that auditory perceptual brain areas

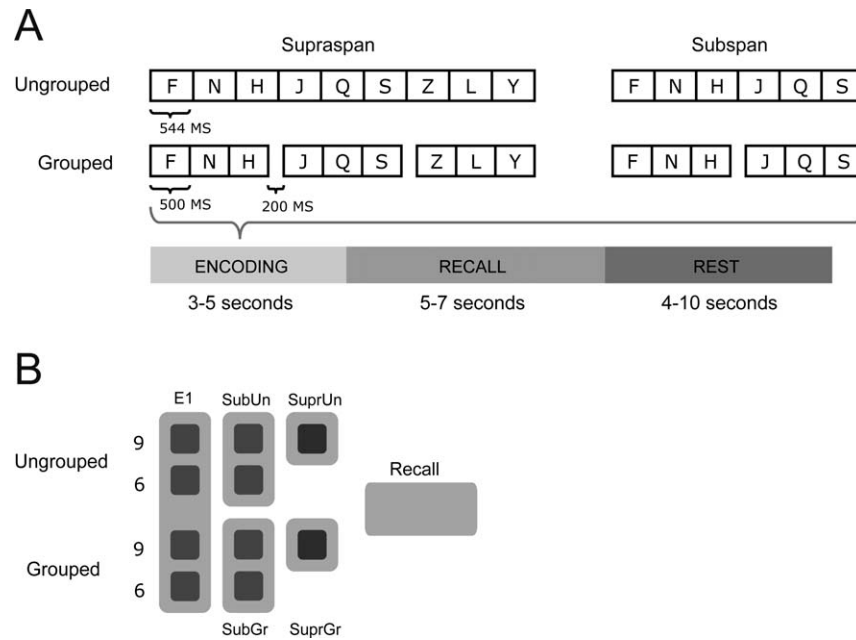


Figure 1.

(A) Structure of trials. (B) Regressors. Each trial was modelled with a single regressor for a retrieval phase and multiple regressors for the encoding phase, depending on the load and grouping conditions.

are differentially activated for grouped and ungrouped sequences. Alternatively, if grouping effects are related to recoding material into higher level chunks, we would expect to see activity in post-perceptual areas of the temporal, parietal and frontal lobes.

MATERIALS AND METHODS

Participants

In total, 23 right-handed volunteers (14 female, 19–34 years old) gave informed, written consent for participation in the study after its nature and possible consequences had been explained to them. Subjects reported no history of psychiatric or neurological disorders and no current use of any psychoactive medications. Three subjects were later excluded from the study because of the excessive motion artefacts in the collected fMRI data. The study was approved by the Cambridge Local Research Ethics Committee (LREC) (Cambridge, UK).

Task

Participants had to recall items from auditorily presented lists of letters in the correct order. Lists were either temporally grouped (grouped trial) or ungrouped (ungrouped trial), and sub-span (six items) or supra-span lists (nine items), which were randomly intermixed so as to dissociate sub- and supra-span encoding effects. During presentation, letters were read out loud by a male voice.

For each trial, participants were visually presented with a cross on a projector screen to indicate the start of the auditory presentation of a novel sequence. Each letter was then presented for 500 ms for the grouped lists and 544 ms for the ungrouped lists, so as to equate the total presentation time for both conditions. This created an imperceptible difference in presentation rates within each group of trials (44 ms per letter). However, previous behavioral studies have shown that the effect of presentation rate becomes significant only when it is slow enough to allow between-items rehearsal [more than a second per item in Brodie and Murdock, 1977 and Jahnke, 1968]. In such light a difference in rate of 44 ms per item is negligible. Given the goal of our study to assess neural activity associated with grouped and ungrouped letter triplets, we felt it preferable to equate the two conditions on presentation duration for triplets, rather than for individual letters. Thus a pause of 200 ms was added after every three letters in the grouped condition to create a temporal grouping effect. Following this, participants either saw a cue “?” indicating that they were to verbally recall the list exactly as they had just heard it; or a cue “-” indicating not to respond and to wait for 4 to 8 seconds for the next list (rest phase, Fig. 1A).

Subjects only had to recall the lists on half of the trials to allow the effects of encoding and retrieval to be modelled separately. Each participant was presented with 192 trials with equal numbers of all trial types pseudo-randomly ordered during each run. Forty eight trials were presented in each scanning run, and four runs were given to each participant, in addition to an initial practice

session outside the scanner. Participants were not informed that there were different types of trials. Participants' verbal responses were recorded during the session for behavioral data analysis. We used both adaptive online noise reduction based on direct comparison of signals from two optical microphones placed in the scanner, one facing towards and one away from the volunteers mouth, and post-scanning noise reduction [Cusack et al., 2005] before transcribing participants' responses.

Stimuli and Behavioral Measures

The letters were spoken by a male native speaker of British English and recorded at 44.1 kHz sampling rate with 16-bits per sample. Recordings were made in a sound-proofed booth with recording equipment located outside the room, then down-sampled to 22.05 kHz mono sound files for playback using headphones in the scanner. The sound files were all adjusted so that the letters appeared equally spaced in all presentation orders and such that the addition of a 200 ms pause provided a rhythmic cue in grouped lists.

For each trial, positional recall was measured as a modified Levenshtein edit distance [Levenshtein, 1966]. A common way to score written recall in ISR is to score items correct only if they are recalled in the correct serial position. However, according to this criterion, a list where the first item was omitted would be scored as completely incorrect. The Levenshtein edit distance calculates the smallest number of edit operations that are necessary to modify one string to obtain another string, where an edit operation is an insertion, deletion, or substitution of a single character, thus ensuring that participants are given maximum credit for those items recalled correctly in each sequence.

Data Acquisition

Participants were scanned at the Medical Research Council Cognition and Brain Sciences Unit (Cambridge, UK) on a 3T Siemens TIM Trio MRI scanner using a head coil. Functional images were collected using 32 slices covering the whole brain (slice thickness 3 mm, 25% slice gap, in-plane resolution 3×3 mm) with an EPI sequence (TR, 2 s; TE, 30 ms; flip angle, 78°). In addition, high-resolution MP-RAGE structural images were acquired at a resolution of $1 \times 1 \times 1$ mm (available at: <http://imaging.mrc-cbu.cam.ac.uk/imaging/ImagingSequences> for additional detailed information.)

Each participant performed four scanning runs, each of which lasted 692 seconds. Three hundred forty-six scans were acquired per run, including 16 dummy scans. Stimulus presentation was controlled by DMDX software version 3 [Forster and Forster, 2003]. Visual cues for sequence onset and recall were rear projected onto a translucent screen outside the bore of the magnet and viewed via a mirror system attached to the head coil. Auditory stimuli were delivered with magnet-safe noise-blocking headphones (NordicNeuroLab, noise attenuation of +30dB).

Data Preprocessing

All fMRI data were preprocessed using SPM5 software (Wellcome Trust Centre for Neuroimaging, London). Before analysis, all images were corrected for slice timing, with the middle slice in each scan used as a reference. Images were realigned with respect to the first image using trilinear interpolation, creating a mean realigned image. The mean realigned image was then coregistered with the structural image and the structural image was normalized to the MNI average brain using the combined segmentation/normalization procedure in SPM5. The normalization parameters were then applied to the functional images before spatial smoothing with a 10 mm full width, half maximum Gaussian kernel.

Analysis Design

Each trial was modeled with a single regressor for a retrieval phase and multiple regressors for the encoding phase, depending on the load and grouping conditions. The first three letters which appeared before the first pause were modeled as one regressor (E1). The second triplet was modeled as sub-span and the last as supra-span encoding regressor of either grouped or ungrouped list (subspan grouped (SubGr), subspan ungrouped (SubUn), supra-span grouped (SuprGr) and supra-span ungrouped (SuprUn); see Fig. 1B).

This model allowed us to accurately model grouping effects: all relevant activation differences between the hemodynamic response as a function of grouping had to occur after the presentation of the third item in the list. Similarly, we split the nine-item sequences into subspan and supra-span portions to separate neural activity due to letters four to six that are usually within the human STM span, and letters seven to nine that are typically supra-span. Although behavioral grouping results are only observed for nine-item sequences, as participants' performance is usually close to ceiling with six or fewer items [see Henson et al., 2000], we may nonetheless see neural correlates of grouping before this period.

The regressors were generated by convolving a box-car representation of the onset and duration of each trial phase with a single canonical response function. A sufficient degree of de-correlation between all of the regressors was insured by: (1) jittering the length of the rest phase (between 4 and 10 seconds), (2) varying the length of the recall period (5–7 seconds) and (3) omitting the retrieval phase for half of the trials. Correlation between the encoding regressors was sufficiently low (0.42 mean) for reliable coefficient estimation with the general linear model implementation in SPM5.

During the encoding phase of the trials, events followed each other rapidly and at the same intrastimulus intervals (ISI), which might cause saturation (or "under-additivity") of BOLD response [Friston et al., 1999; Henson, 2006]. However, the size of this non-linearity appears to be relatively small until the time between the onsets of events within a trial (stimulus onset asynchronies, or SOAs) is

less than a second: according to Friston et al. [1999] optimal SOA can be as low as 1s without such non-linearities having a major impact on the predicted design efficiency. The duration of the grouped and ungrouped letter triplets in the task design (1,700 ms) is comfortably above this lower-limit. Hence the task design should ensure that potential nonlinear trends in the BOLD response are of neural origin and not due to design efficiencies.

Additional regressors to account for large head movements were added to the first-level model, mostly for recall periods when participants were speaking during scanning. In addition to six motion parameters (corresponding to translations and rotations of the image due to movement in the scanner), additional movement parameters were modeled to account for extreme inter-scan movements which exceeded a translation threshold of 0.5 mm, rotation threshold of 1.33 degrees, and between-images difference threshold of 0.035, calculated by dividing the summed squared difference of consecutive images by the squared global mean. A separate movement spike regressor was added to the general linear model for each scan that exceeded these thresholds.

Activation Analysis

Single subject statistical contrasts were set up by using the general linear model to fit each voxel with a combination of functions derived by convolving the standard haemodynamic response with the time series of the events and removing low-frequency noise with a high-pass filter (128 seconds). This resulted in voxel-wise t -statistic images displaying differences in activation levels for the two presentation modes (grouped, ungrouped) across the three encoding phases (one for each letter triplet) and a single recall phase (see Fig. 1B).

Activation during the encoding phase was assessed by adding together all encoding regressors (E1 + SubGr + SubUn + SuprGr + SuprUn) and contrasting the regression coefficients (β) against the implicit resting baseline expressed by the residual errors (ϵ). Activity during overt recall was modelled with a single retrieval regressor. Differential activation during encoding was assessed by contrasting grouped encoding events with the ungrouped encoding events, and subspan and supra-span periods. This led to two contrasts that expressed the main effects of grouping ($[\text{SubGr} + \text{SuprGr}] > [\text{SubUn} + \text{SuprUn}]$) and load ($[\text{SuprUn} + \text{SuprGr}] > [\text{SubGr} + \text{SubUn}]$). Given previous behavioral data, we expected grouping effects to be more pronounced for supra-span lists and hence predicted a grouping by load interaction. This interaction was tested by contrasting load effects for grouped and ungrouped lists, i.e. $([\text{SuprGr} > \text{SubGr}] > [\text{SuprUn} > \text{SubUn}])$.

Each t -statistic image was projected onto the subject's cortical surface representation with the FreeSurfer *tk-surf* program. Group t -statistics with subject as a random effect could then be computed for each node of the average surface mesh. All reported peaks were from the group analysis,

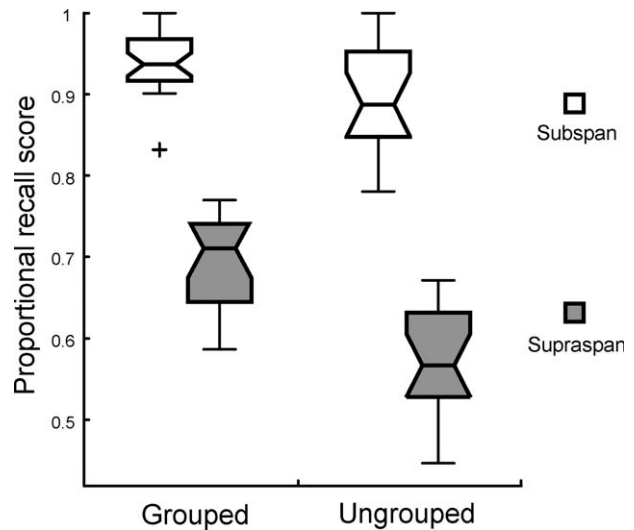


Figure 2.

Average proportional recall in the scanner measured by Levenshtein distance. For all box-plots, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually using a +.

had to pass a whole-brain false detection rate [FDR; Genovese et al., 2002] threshold of $P < 0.05$. The FDR approach controls for the expected proportion of false positives among supra-threshold voxels. An FDR threshold is determined from the observed P -value distribution, and hence is adaptive to the amount of signal within a given contrast.

Regression Analysis

To assess whether differential brain activation predicted recall performance, we calculated simple regressions (correlations) between activation in a number of regions of interest (ROIs) and recall scores. The ROIs were defined using the voxels that showed either a main effect of grouping, load, or a grouping \times load interaction in the group level analysis. The $[\text{beta}]$ values from the peak voxel (the voxel with the highest t -value within the activated cluster) were correlated with the recall score for individual subjects in the supra-span condition. We used this behavioral effect as the dependent measure since grouping effects in the subspan condition are hard to measure due to participants performing at near ceiling levels.

RESULTS

Behavioral Results

Positional recall performance in the scanner (see Fig. 2) showed the expected main effects of load ($F_{(1,79)} = 368, P < 0.001$) and grouping ($F_{(1,79)} = 31.04, P < 0.001$), and an interaction between load and grouping ($F_{(1,79)} = 19.05, P < 0.01$).

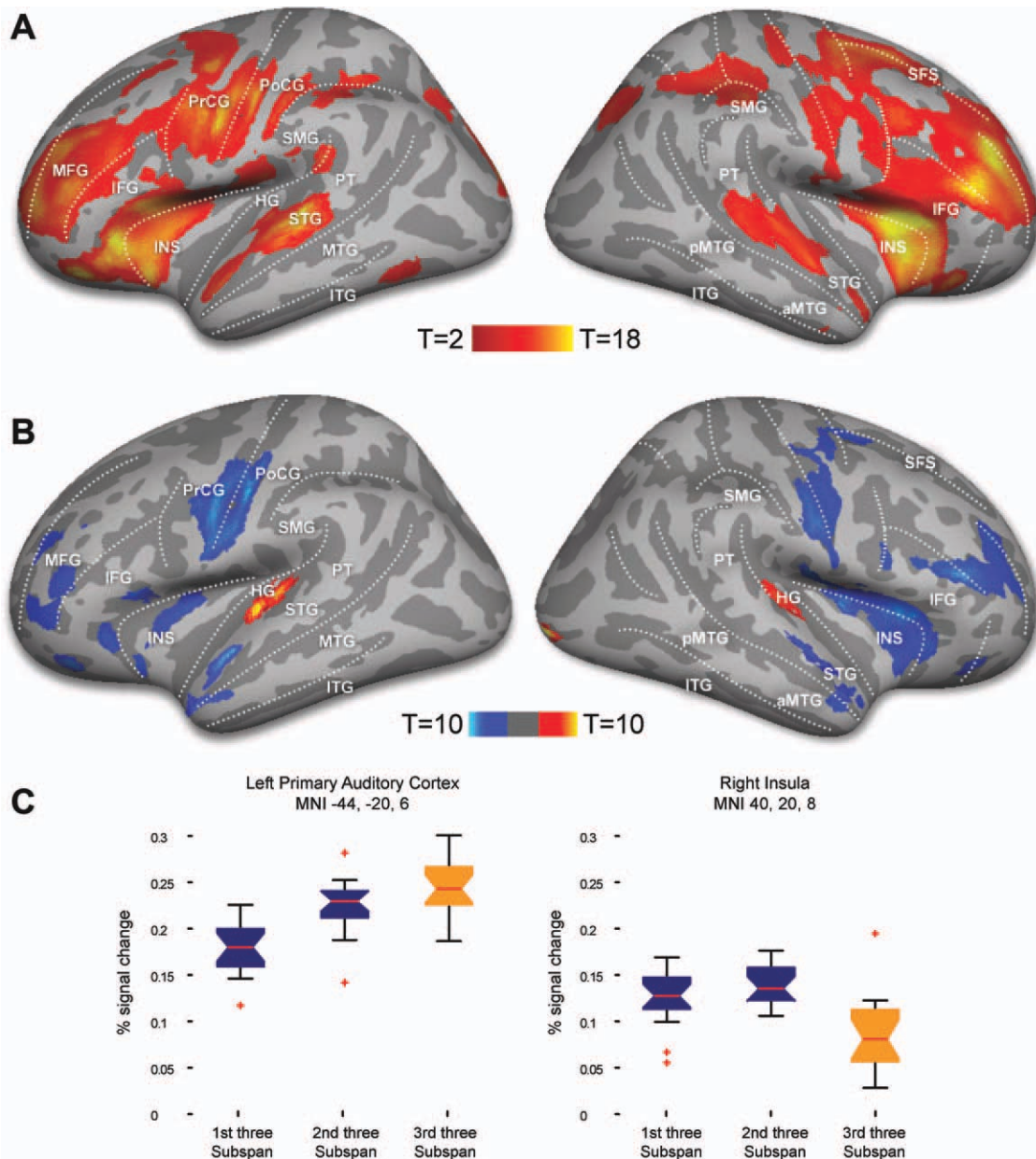


Figure 3.

(A) Increased activation during encoding compared to unmodelled baseline, $p < 0.05$ FDR corrected. (B) Increased activation during supraspan trials (red | yellow) as compared to decrease in activation during subspan trials (dark blue | light blue), $p < 0.05$ FDR corrected. (C) Activation profiles for load contrasts: peak voxel activation profiles in (1) left primary auditory cortex and (2) right insula during the encoding stages (% signal change).

As outlined on Figure 2, this interaction was expected since participants tend to not make errors when lists are subspan [Henson et al., 2000]. In addition, the majority of participants ($N = 13$), when asked afterwards, recognized that some lists were grouped during encoding, and found these trials easier to perform.

Imaging Results

Activation across trial phases

Figure 3A shows multisubject surface-based functional activation for the encoding phase. During the encoding phase substantial proportions of the frontal, parietal and

TABLE I. Peak voxel activations for contrasts

Contrast	Hemi	Location	x	y	z	T	Z-score
Encoding > unmodelled rest							
	L	Superior temporal gyrus	-62	-26	2	17.75	6.88
	L	Insula	-36	20	-2	14.58	6.44
	L	Inferior frontal gyrus	-32	36	-8	13.85	6.33
	L	Middle frontal gyrus	-46	38	28	12.91	6.16
	L	Precentral gyrus	-54	4	32	12.83	6.15
	L	Inferior parietal sulcus	-52	-38	36	11.92	5.98
	R	Superior temporal gyrus	52	14	-26	15.55	6.59
	R	Insula	34	20	10	12.07	5.75
	R	Cerebellum	32	-40	-38	11.92	5.98
	R	Inferior frontal gyrus	28	16	14	9.62	5.46
	R	Middle frontal gyrus	44	44	28	9.57	5.45
	R	Superior frontal gyrus	22	44	40	9.13	5.13
	R	Supramarginal gyrus	54	-20	38	8.01	4.82
Load: supraspan > subspan							
	L	Primary auditory cortex	-44	-20	6	10.49	5.45
	R	Primary auditory cortex	50	-12	8	9.89	5.53
	R	Primary visual cortex	18	-104	2	9.1	5.13
Load: subspan > supraspan							
	L	Superior temporal cortex	-56	-6	-4	9.3	5.38
	L	Anteriosuperior temporal cortex	-42	22	-30	8.31	4.92
	L	Precentral gyrus	-42	-8	38	7.95	4.82
	L	Inferior frontal gyrus	-34	38	-14	7.45	4.6
	L	Middle frontal gyrus	-34	44	20	6.73	4.43
	R	Anteriosuperior temporal cortex	58	12	-30	8.07	4.85
	R	Insula	40	20	8	8.21	4.89
	R	Precentral gyrus	54	10	28	7.25	4.6
	R	Inferior frontal gyrus	46	36	-6	7.12	4.56
Ungrouped > grouped							
	L	Superior temporal cortex	-54	2	-2	3.11	2.71
	R	Superior temporal cortex	62	-2	-6	2.94	2.59
	R	Superior temporal cortex	52	8	-24	2.51	2.27
Supraspan: grouped > ungrouped							
	L	Supramarginal gyrus	-56	-22	36	3.10	2.70
	L	Precentral gyrus	-48	2	26	2.89	2.51
Subspan: ungrouped > grouped							
	L	Superioposterior temporal gyrus	-56	-30	16	4.23	3.41
	L	Insula	-36	20	-2	3.89	3.22
	L	Premotor cortex	-28	-2	-58	3.69	3.10
	L	Middle frontal gyrus	-38	54	14	2.66	2.39
	R	Superioposterior temporal gyrus	60	-22	20	3.97	3.26

FDR corrected, $P < 0.05$.

temporal lobes showed robust and bilateral activity (Table I). During the retrieval phase, when items had to be recalled, considerable activation was observed in motor and premotor, frontal, and auditory areas. (Details of activations during retrieval can be obtained from the first author on request.)

Activation for load contrasts

The subspan versus supra-span contrasts revealed an extensive premotor-prefrontal network more active during subspan encoding (Fig. 3B). Only bilateral primary auditory areas were more active for supra-span stimuli. Also

the right primary visual cortex showed an increase in activation for the final, supra-span letter triplet, perhaps by virtue of these items predicting the onset of a subsequent visual cue for recall.

Activation for grouping contrasts

There were main effects of grouping (grouped vs. ungrouped) with more activation for ungrouped stimuli bilaterally in superior temporal areas anterior to Heschl's gyrus and in the anterior portion of the superior temporal sulcus (Fig. 4A).

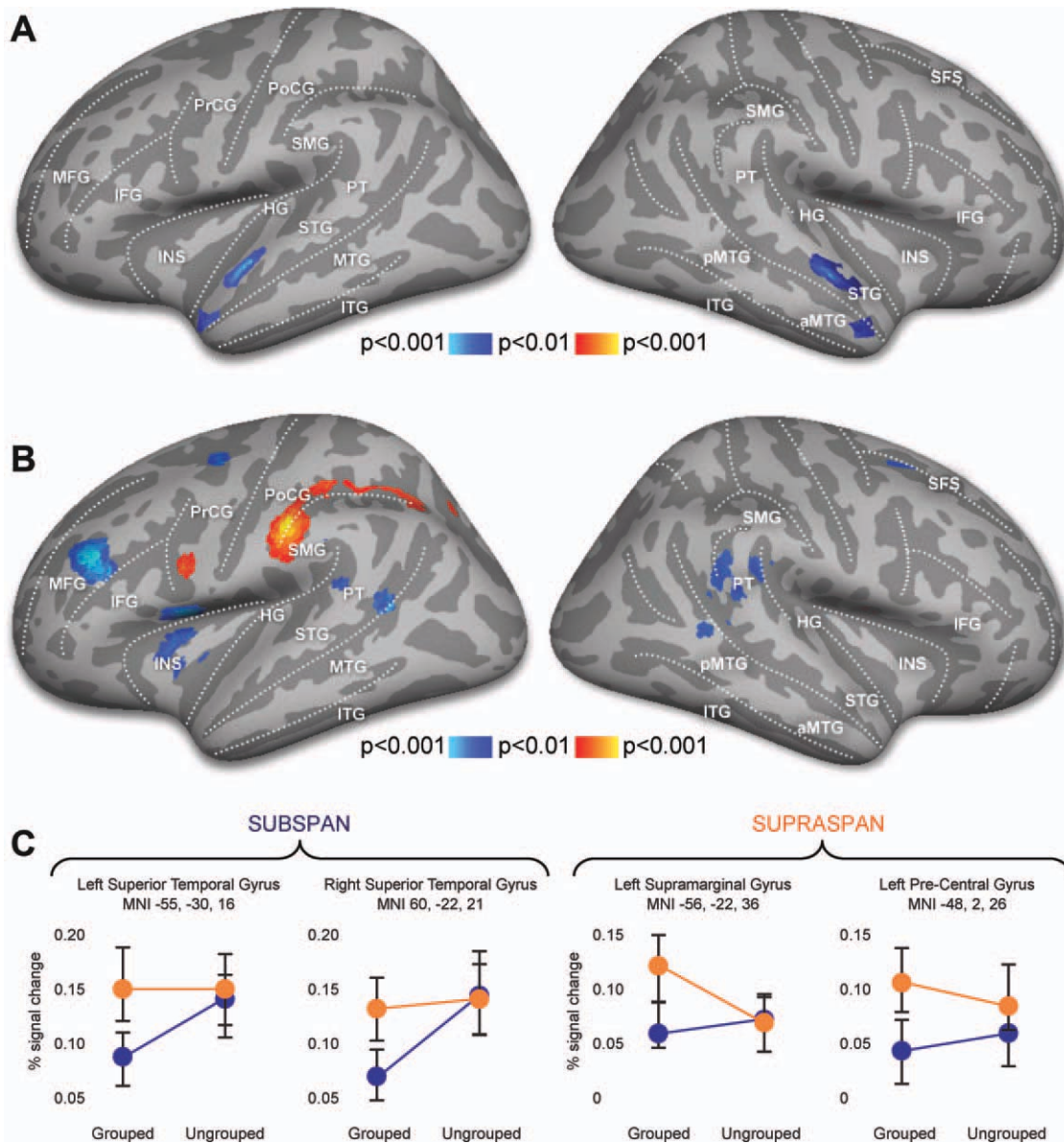


Figure 4.

Main effects of grouping: Decreased activation during grouped trials as compared to ungrouped trials during the encoding phase. (B) Grouping \times load interaction: Increased activation during grouping load effects compared to ungrouped load effects: supra-span phase (red | yellow) and subspan phase (dark blue | light blue). (C) Activation profiles for voxels showing grouping effects during subspan and supra-span phases of encoding (% signal change).

The load by grouping interaction revealed differential activity in a number of bilateral regions specific to sub- or supra-span portions of grouped and ungrouped lists (Fig. 4B). In particular, planum temporale (PT) showed bilaterally more activation for ungrouped stimuli during the sub-span phase, while supramarginal, precentral, and inferior parietal (IPL) areas were more active for grouped lists during the supra-span phase. These effects were more pro-

nounced in the left hemisphere, where middle frontal, premotor and postero-inferior temporal areas showed more activation.

A number of brain regions showed additional activation for ungrouped lists that was greater in the sub-span than supra-span phase: frontal areas including insula (INS), pars triangularis in inferior frontal gyrus (IFG), middle frontal gyrus (MFG), superior frontal sulcus (SFS), and

TABLE II. Results of simple regressions of behavioral effects against brain activity

Load									
Correlation between supraspan activation and supraspan recall performance									
MNI coordinates	-44 -20 6	50 -12 8	-42 -8 38	54 10 28	-42 12 30	58 12 30	-34 44 20	46 36 -6	
Region	L A1	R A1	L PrCG	R PrCG	L aSTG	R aSTG	L MFG	R MFG	
Pearson's <i>R</i>	0.46*	0.49*	0.37	0.39	0.20	0.10	0.25	0.19	
Correlation between supraspan > subspan activation and supraspan recall performance									
MNI coordinates	-44 -20 6	50 -12 8	-42 -8 38	54 10 28	-42 12 30	58 12 30	-34 44 20	46 36 -6	
Region	L A1	R A1	L PrCG	R PrCG	L aSTG	R aSTG	L MFG	R MFG	
Pearson's <i>R</i>	0.37	0.40	0.44*	0.49*	0.11	0.13	0.18	0.34	
Grouping									
Correlation between supraspan grouped activation and supraspan grouped recall performance									
MNI coordinates	-56 -22 36	-48 2 26	-36 20 -2	-56 -30 16	60 -22 20	-54 2 -2	52 8 -24	-38 54 14	
Region	L SMG	L PCG	L INS	L PT	R PT	L STG	R STG	R MFG	
Pearson's <i>R</i>	0.47*	0.27	0.28	-0.35	-0.31	-0.42*	-0.44*	-0.22	

* $P < 0.05$.

posterior temporal areas including PT and posterior superior temporal sulcus (pSTS).

During the supra-span phase grouped stimuli elicited more activation compared to the ungrouped lists in left premotor (inferior pre-central sulcus (PrCG)) and parietal (inferior postcentral sulcus and supramarginal gyrus (SMG)) areas. All the effects related to the supra-span phase were detected in the left hemisphere only.

Regression Analysis

We conducted cross-subject regression analysis of neural responses in peak voxels so as to relate neural effects of load and grouping to individual participant's behavioral performance.

Load

We extracted parameter estimates ($[beta]$ values) for supra-span and subspan encoding phases. For each subject, betas values from the peak voxels of the ROIs were correlated with the mean recall score of supra-span lists (collapsed over grouped and ungrouped conditions). Subspan recall scores were not included in the correlation as all the participants showed near-ceiling performance with sub-span lists (mean subspan recall score = 0.94). We interpreted this supra-span recall score as a participant's measure of STM span. Brain activity in four bilateral regions of interest (8 ROIs; A1, aSTG, PrCG, MFG; bilaterally), chosen according to clusters of activity in the load contrast (supra-span vs. subspan, "Results," Fig. 3B, Table I), were included in a simple regression with supra-span recall scores (see "Materials and methods," "Regression analysis").

We computed two simple regressions: first, we correlated $[beta]$ values from the supra-span encoding phase with the supra-span recall scores, to see how activity during the supra-span phase predicts subsequent performance. Second, we correlated the difference between supra-span and subspan betas with the supra-span recall score, to see whether

the change of activity from subspan to supra-span can predict recall performance. Activity during subspan periods supports (essentially) perfect recall, whereas presentation of further items (i.e. a supra-span lists) disrupts performance not only for the additional supra-span items, but also for items presented before span was exceeded. We therefore might expect that differences between responses during subspan and supra-span encoding to be more informative than absolute activation during either phase alone. The smaller the difference in $[beta]$ values, the less activity in this brain region is affected by exceeding the memory span. Thus, we predict a positive correlation between differential activation and recall scores; those participants for whom activity shows a large reduction when span is exceeded should perform more poorly on recall.

Supra-span activity in bilateral auditory cortex (A1) was reliably correlated with recall performance (Table II). Change of activity from subspan to supra-span phase was reliably correlated with subsequent recall performance in left and right pre-central gyri (L PrCG, R PrCG) (Table II).

Grouping

To assess whether brain activity was able to predict behavioral grouping effects, we correlated the supra-span recall scores with beta values of grouped and ungrouped trials from the supra-span encoding phase. In addition, we correlated the supra-span recall score with the difference between grouped and ungrouped betas, to see whether the difference in activity between grouped and ungrouped trials predicted recall performance.

Brain activity in eight ROIs (left SMG, left PCG, left INS, left PT, right PT, left STG, right STG, right MFG; see Results, Fig. 4, Table I) were included in simple regressions with recall scores. Only two correlations between grouped activation and recall performance were significant: left supramarginal gyrus (L SMA) showed reliable positive correlation, while anterosuperior temporal gyrus (STG) showed negative correlation bilaterally (Table II).

No significant correlations with supra-span recall performance were observed for ungrouped trials or for the difference between grouped and ungrouped trials.

DISCUSSION

We examined the neural correlates of grouping in immediate serial recall using sequences of sub- and supra-span letters. The overall activation profiles for encoding and retrieval are in agreement with recent neuroimaging data on STM for auditorily presented verbal stimuli [Buchsbaum et al., 2005; Strand et al., 2008] in that there was robust bilateral activity in substantial proportions of the frontal, parietal and temporal lobes (Fig. 3A, Table I: “Encoding > Unmodeled rest”). Our results also showed a considerable overlap in activation between encoding and recall in both superior temporal areas associated primarily with auditory speech perception, and in premotor areas, associated with production.

Grouping Effects

For both subspan and supra-span lists of letters, two loci within the superior temporal lobe showed greater activation for ungrouped than grouped lists during encoding: an area just anterior to Heschl’s gyrus, and the anterior tip of the temporal lobe (Fig. 4A, Table I). Given that this peri-auditory activation is in a very similar area to that observed for supra-span stimuli, this suggests that ungrouped stimuli require additional auditory encoding within superior temporal lobe. The idea that superior temporal areas outside of Heschl’s gyrus play a role in low-level perceptual memory is supported by fMRI data [Buchsbaum et al., 2005] and electrophysiological recordings [Ojemann et al., 2009]. Such activation might reflect rapidly decaying auditory sensory memory, sometimes called echoic memory. From this viewpoint our results suggest that, independent of the list length, grouped lists require less auditory sensory memory.

It is also possible that activation in these superior temporal areas might be related to subvocal rehearsal. Previous imaging studies of verbal STM have shown that parts of the temporal cortex, especially along the middle and anterior portions of supratemporal plane, deactivate during silent speech production [Houde et al., 2002; Zheng et al., 2010]. Similarly, studies with monkeys and humans have shown that auditory neurons are suppressed during subjects own vocalizations [Numminen et al., 1999]. The relative deactivation of temporal areas with grouped stimuli may therefore be a consequence of subjects engaging in covert rehearsal during the pauses between groups.

There were no brain areas that were more active during the encoding of grouped than ungrouped stimuli in both subspan and supra-span phases. It has been proposed [Henson et al., 1996] that grouping is driven by the need for more short-term storage, which is only evident when

participants have to recall more than about six items. If so, grouping effects would be expected to be minimal in the subspan condition (as demonstrated with previous behavioral data; Henson et al., 2000). This proposal might explain both why we found no areas that were more active for grouped stimuli during the subspan phase, and why a previous imaging study by Henson et al. [2000] that used six-item lists did not show a main effect of grouping.

Load Effects

As item load increased (first, second, and third set of three items in letter sequences) there was a monotonically increasing BOLD response in primary auditory areas, and a non-linear response in pre-frontal and temporal areas (Fig. 3C). A nonlinear response to STM load has been observed and replicated in a number of neuroimaging studies using parallel presentation of visual objects [e.g. Mitchell and Cusack, 2008; Moran et al., 2009; Vogel and Machizawa, 2004]. The common interpretation is that these measures are indicative of asymptotic neural response to items in STM, reflecting participants individual STM span [Luck and Vogel, 1997; Moran et al., 2009]. However, since our sequences were presented sequentially rather than in parallel we must be more cautious in concluding that observed BOLD nonlinearities have a neural origin.

In a sequential task, like ours, stimulus duration inevitably increases with item load and thus the possible effects of nonlinear BOLD saturation cannot be entirely eliminated [Mechelli et al., 2001; Talavage and Edmister, 2004]. However, the dissociation between the activation of auditory and frontal areas (Results, Table I) suggests that our load effect is unlikely to be due to hemodynamic nonlinearities alone, since different brain areas respond differently to STM load increase. Furthermore, as with studies of visual STM, the significant correlation between supra-span BOLD signal and recall performance (Results, Table II) provides additional evidence for a neural origin of these nonlinearities. In particular, we observed a significant correlation between recall performance and BOLD response in pre-frontal areas (see “Results,” “Correlation analysis,” and “Load”) even with supra-span lists, which would not be expected if the BOLD response was saturated. Specifically, there was a reliable positive correlation between recall and the difference in activation between subspan and supra-span encoding phases in the bilateral precentral gyri. For these ROIs, subspan parameter estimates were greater than supra-span ones. A positive correlation with recall scores means that the participants who showed the smallest reduction in supra-span phase brain activity (compared with the subspan phase) were better at subsequent recall. This suggests that participants’ ability to encode the final three letters of supra-span stimuli into STM depends on the amount of de-activation observed in the precentral gyri during the supra-span phase (see Fig. 5B, Table II).

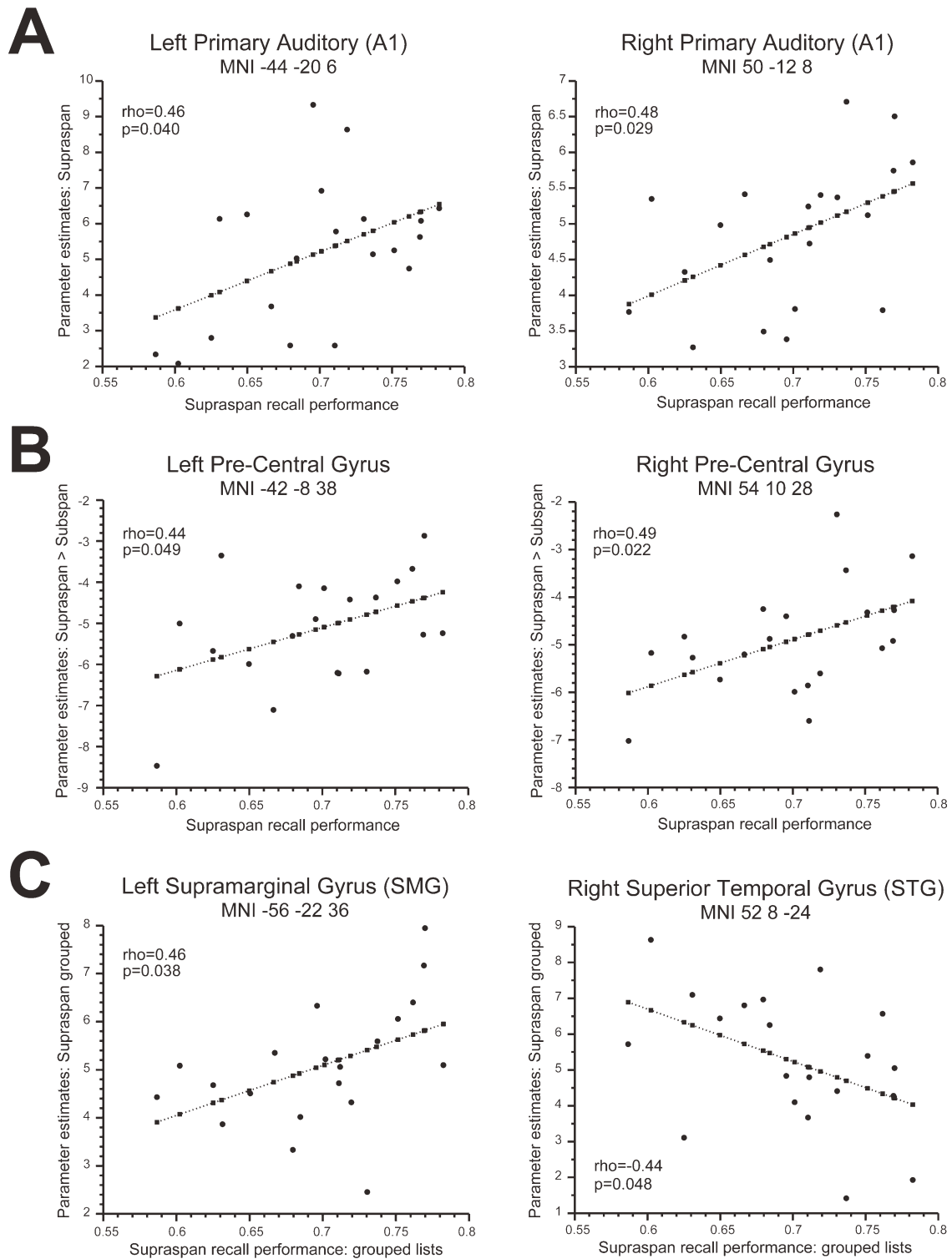


Figure 5.

Correlations between behaviour and brain activity: (A) Correlation between supraspan activation and supraspan recall performance (B) Correlation between 'supraspan>subspan' activation and supraspan recall performance (C) Correlation between activation during supraspan grouped trials and recall performance.

Our results are in agreement with other studies that have shown an increase in the BOLD response with item load before exceeding the memory span [Chein and Fiez, 2001; Henson and Fletcher, 2001; Rypma et al., 1999], and reports of an asymptotic neural response to item load [Mitchell and Cusack, 2008]. Thus our findings have the potential to differentiate load effects for subspan lists (for which activity in prefrontal and posterior regions is increased), and the effect of increased load that exceeds the STM span.

One of the fascinating aspects of our load results was the dissociation within the temporal lobe: bilateral primary auditory areas were more active for supra-span stimuli while more anterior temporal were less active (bilateral anterior temporal gyrus). Previous research has established that anterior regions within the superior temporal lobe seem to respond to abstract, linguistic information in speech. Obleser et al. [2006] showed that left anterior temporal cortex responded more to intelligible consonantal bursts compared with incomprehensible control sounds. Davis and Johnsrude [2003] have shown that anterior temporal areas show a response that is correlated with sentence intelligibility, but unlike periauditory regions, the anterior temporal lobe is largely insensitive to the nature of the stimulus degradation applied to speech. These findings suggest that the anterior temporal lobe participates in the larger ventral stream of speech processing [Hickok and Poeppel, 2007; Rauschecker and Scott, 2009]. Thus, the relative deactivation of anterotemporal regions for supra-span items suggests that exceeding the memory span affects primarily speech processing network but preserves auditory sensory memory.

The Interaction Between Grouping and Load

During the subspan phase, bilateral posterior superior temporal gyrus regions (including planum temporale) were less active for grouped than ungrouped lists, while during the supra-span phase supramarginal (SMG) and inferior parietal areas (IPL) were more active for grouped lists (Fig. 4B,C; Table I). No areas were more active for grouped than ungrouped lists during the subspan phase. So what is the most likely cause of this interaction between grouping and load?

In a number of functional imaging studies the IPL and the SMG have been linked to sequential speech and language processing [Buchsbaum, 2008; Marshuetz, et al., 2006; Marshuetz and Smith, 2006; Moser et al., 2009]. For example, in a task where participants were required to reverse the order of subspan auditory syllables and hummed notes, Gelfand and Bookheimer [2003] found a very similar pattern of neural activity to that which we observed for grouped supra-span stimuli. They found that while the posterior portion of Broca's area responded specifically to the sequence manipulation tasks regardless of the nature of the stimuli, activation of the left supramarginal gyrus (SMG) was more specific to sequencing phonemes. Conditions in which par-

ticipants were required to reverse the order of sequences produced signal increases in the left IPL and posterior superior parietal lobule (SPL), and in the right precentral gyrus. The common element in both grouping and sequence reversal might be the need for stimuli to be recoded or restructured, which may only be necessary in supra-span lists. In addition, many previous studies have also associated prefrontal regions with sequence and order processing. However, this frontal component appears to be largely modality and domain independent [Bor et al., 2003; Gelfand and Bookheimer, 2003; Henson and Fletcher, 2001].

A number of brain areas, including left middle frontal gyrus, left insula, and posterior superior temporal areas, including bilateral posterior PT, responded more for ungrouped stimuli during the subspan phase. This is consistent with the data from Henson et al. [2000], who observed more activation in the left dorsolateral premotor cortex (in BA 6) for visually-presented ungrouped subspan letter sequences. However, we found that this and other coactivated regions did not show an increased response to ungrouped stimuli for supra-span lists.

Activation in posterior temporal areas, specifically the PT, has been observed in previous studies of auditory maintenance and rehearsal. Buchsbaum et al. [2005] showed that for auditory-verbal stimuli the activity in PT was sustained throughout the delay period, during which participants covertly rehearsed the stimuli. Additionally, Hocking and Price [2008] observed that posterior superior temporal sulcus (pSTS) activation depended on the number of auditory and visual perceptual objects that needed to be simultaneously attended to during encoding and response. Our data show that, for verbal stimuli, posterior temporal activation depends not only on the absolute number of letters or syllables to be encoded, but also on whether or not these syllables can be grouped into chunks.

However, we found that that no brain areas responded more to grouped than ungrouped stimuli before span was exceeded, raising a question of which brain areas are responsible for coding grouping information in the subspan phase. In particular, why did we find no differential SMG-IPL activity for grouped and ungrouped lists during the subspan phase?

It is possible that the nature of the grouping process changes when span is exceeded [see MacGregor, 1987, and Yoshino, 1993, for a discussion of such possibilities) or that brain areas responsible for group encoding must work harder when span is exceeded, resulting in an increased BOLD response. Further studies are necessary to tackle this question as univariate analysis of fMRI data allows only limited inferences about the nature of neural representations within specific regions.

CONCLUSIONS

We have established with fMRI that grouping effects can be seen even during subspan encoding, showing that,

contrary to previous proposals [Henson et al., 1996] grouping effects are not simply a consequence of the demands of supra-span encoding. In addition, grouping does not lead to greater involvement of brain regions responsible for perceptual processing. If anything, the opposite is true: superior temporal brain regions are activated significantly more by ungrouped stimuli. The left inferior parietal lobe, including the supramarginal gyrus, areas previously associated with phonological sequencing and sublexical processing, are more active for grouped stimuli. This neural signature of grouping suggests that the benefit of presenting letters in triplets seems to reflect recoding the groups into chunks in parietal areas instead of reliance on perceptual-auditory processing in temporal regions.

REFERENCES

- Adams H (1915): A note on the effect of rhythm on memory. *Psychol Rev* 22:289–299.
- Baddeley A, Hitch G (1974): Working memory. In: Bower G., editor. *Recent Advances in Learning and Motivation*. New York: Academic Press. pp47–90.
- Bor D, Duncan J, Wiseman R, Owen A (2003): Encoding strategies dissociate prefrontal activity from working memory demand. *Neuron* 37:361–367.
- Botvinick M, Plaut D (2006): Short-term memory for serial order: A recurrent neural network model. *Psychol Rev* 113:201–233.
- Broadbent D, Broadbent M (1981): Articulatory suppression and the grouping of successive stimuli. *Psychol Res* 43:57–67.
- Brodie D, Murdock B (1977): Effect of presentation time on nominal and functional serial-position curves of free recall. *J Verbal Learn Verbal Behav* 16:185–200.
- Brown G, Preece T, Hulme C (2000): Oscillator-based memory for serial order. *Psychol Rev* 107:127–181.
- Buchsbaum B (2008): The search for the phonological store: From loop to convolution. *J Cogn Neurosci* 20:762–778.
- Buchsbaum B, Olsen R, Koch P, Berman K (2005a): Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron* 48:687–697.
- Buchsbaum B, Olsen R, Koch P, Kohn P, Kippenhan J, Berman K (2005b): Reading, hearing, and the planum temporale. *Neuroimage* 24:444–454.
- Burgess N, Hitch G (2006): A revised model of short-term memory and long-term learning of verbal sequences. *J Mem Lang* 55:627–652.
- Chein J, Fiez J (2001): Dissociation of verbal working memory system components using a delayed serial recall task. *Cereb Cortex* 11:1003–1014.
- Cusack R, Cumming N, Bor D, Norris D, Lyzenga J (2005): Automated post-hoc noise cancellation tool for audio recordings acquired in an MRI scanner. *Hum Brain Mapp* 24:299–304.
- Davis M, Johnsrude I (2003): Hierarchical processing in spoken language comprehension. *J Neurosci* 23:3423–3431.
- Forster K, Forster J (2003): DMDX: A Windows display program with millisecond accuracy. *Behav Res Methods* 35:116–124.
- Frankish C (1985): Modality specific grouping effects in STM. *J Mem Lang* 209:200–209.
- Frankish C (1989): Perceptual organization and precategorical acoustic storage. *J Exp Psychol* 15:469–479.
- Friston K, Zarahn E, Josephs O, Henson R, Dale A (1999): Stochastic designs in event-related fMRI. *Neuroimage* 10:607–619.
- Gelfand J, Bookheimer S (2003): Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38:831–842.
- Genovese C, Lazar N, Nichols T (2002): Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *Neuroimage* 15:870–878.
- Griffiths T, Warren J (2002): The planum temporale as a computational hub. *Trends Neurosci* 25:348–353.
- Henson R (1998): Short-term memory for serial order: The Start-End Model. *Cogn Psychol* 36:73–137.
- Henson R (2001): Short-term memory for serial order. *The Psychologist* 14:70–73.
- Henson R (2006): Efficient experimental design for fMRI. In: Friston K, Ashburner J, Kiebel S, Nichols T, Penny W, editors. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier: London. pp 193–210.
- Henson R, Burgess N, Frith C (2000): Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia* 38:426–440.
- Henson R, Fletcher P (2001): Frontal lobes and human memory. *Brain* 124:849–881.
- Henson R, Norris D, Page M, Baddeley A (1996): Unchained memory: Error patterns rule out chaining models of immediate serial recall. *Q J Exp Psychol A* 49:80–115.
- Hickok G, Poeppel D (2007): The cortical organization of speech processing. *Nature* 8:393–402.
- Hitch G, Burgess N, Towse J, Culpin V (1996): Temporal grouping effects in immediate recall: A working memory analysis. *Q J Exp Psychol* 49.4:116–139.
- Hocking J, Price C (2008): The role of the posterior superior temporal sulcus in audiovisual processing. *Cereb Cortex (New York, NY)* 18:2439–2449.
- Houde J, Nagarajan S, Sekihara K, Merzenich M (2002): Modulation of the auditory cortex during speech: An MEG study. *J Cogn Neurosci* 14:1125–1138.
- Jahnke J (1968): Presentation rate and the serial-position effect of immediate serial recall. *J Verbal Learn Verbal Behav* 7:608–612.
- Levenshtein V (1966): Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics-Doklady* 10:707–710.
- Luck S, Vogel E (1997): The capacity of visual working memory for features and conjunctions. *Nature* 390:279–281.
- MacGregor J (1987): Short-term memory capacity: Limitation of optimization? *Psychol Rev* 94:107–108.
- Marshuetz C, Reuterlorenz P, Smith E, Jonides J, Noll D (2006): Working memory for order and the parietal cortex: An event-related functional magnetic resonance imaging study. *Neuroscience* 139:311–316.
- Marshuetz C, Smith E (2006): Working memory for order information: Multiple cognitive and neural mechanisms. *Neuroscience* 139:195–200.
- Mechelli A, Price C, Friston K (2001): Nonlinear coupling between evoked rCBF and BOLD signals: A simulation study of hemodynamic responses. *Neuroimage* 14:862–872.
- Miller G (1956): The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol Rev* 63:81–97.
- Mitchell D, Cusack R (2008): Flexible, capacity-limited activity of posterior parietal cortex in perceptual as well as visual short-term memory tasks. *Cereb Cortex* 18:1788–1798.

- Moran R, Strange B, Campo P, Dolan R (2009): Theta bandwidth determines human visual working memory capacity. *Neuroimage* 47:S144–S144.
- Moser D, Baker J, Sanchez C, Rorden C, Fridriksson J (2009): Temporal order processing of syllables in the left parietal lobe. *J Neurosci* 29:12568–12573.
- Numminen J, Salmelin R, Hari R (1999): Subject's own speech reduces reactivity of the human auditory cortex. *Neurosci Lett* 265:119–122.
- Obleser J, Zimmermann J, Van Meter J, Rauschecker J (2006): Multiple stages of Aud Speech Perception fMRI. *Cereb Cortex* 17:2251–2257.
- Ojemann G, Schoenfield-McNeill J, Corina D (2009): The roles of human lateral temporal cortical neuronal activity in recent verbal memory encoding. *Cereb Cortex* 19:197–205.
- Page M, Norris D (1998): The primacy model: A new model of immediate serial recall. *Psychol Rev* 105:761–781.
- Rauschecker J, Scott S (2009): Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nat Neurosci* 12:718–724.
- Ryan J (1969): Grouping and short-term memory: Different means and patterns of grouping. *Q J Exp Psychol* 21:137–147.
- Rypma B, Prabhakaran V, Desmond J, Glover G, Gabrieli J (1999): Load-dependent roles of frontal brain regions in the maintenance of working memory. *Neuroimage* 9:216–226.
- Strand F, Forssberg H, Klingberg T, Norrelgen F (2008): Phonological working memory with auditory presentation of pseudo-words—An event related fMRI Study. *Brain Res* 1212:48–54.
- Talavage T, Edmister W (2004): Nonlinearity of fMRI responses in human auditory cortex. *Hum Brain Mapp* 22:216–228.
- Vogel E, Machizawa M (2004): Neural activity predicts individual differences in visual working memory capacity. *Nature* 428:748–751.
- Wickelgren W (1964): Size of rehearsal group and STM. *J Exp Psychol* 68:413–419.
- Yoshino R (1993): Magical systems of human STM—Efficient designs of biological memory systems? *Behaviormetrika* 20:171–186.
- Zheng Z, Munhall K, Johnsrude I (2010): Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *J Cogn Neurosci* 22:1770–1781.