

Neural network for encoding immediate memory in phonological processing

Xiaojian Li, Donald Wong,¹ Jack Gandour,^{2,CA} Mario Dzemidzic,³ Yunxia Tong,² Thomas Talavage⁴ and Mark Lowe³

South China Normal University, Guangzhou, PR China; ¹Department of Anatomy and Cell Biology, Indiana University School of Medicine; ²Department of Speech, Language and Hearing Sciences, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038; ³Cleveland Clinic Foundation, Cleveland, OH; ⁴School of Electrical and Computer Engineering, Purdue University, 1353 Heavilon Hall, 500 Oval Drive, West Lafayette, IN 47907-2038, USA

^{CA}Corresponding Author: gandour@purdue.edu

Received 10 August 2004; accepted 18 September 2004

The aim of this fMRI study was to identify neuroanatomical substrates of immediate memory underlying phonological processing. To distinguish encoding of immediate memory from rehearsal, participants were required to match tones from the first and last positions of a three-syllable list to their following probes in an immediate-recognition paradigm. The first position task included intervening distractors between the target and probe. Increased activations were found in the left inferior frontal gyrus,

right lateral cerebellum, and medial frontal gyrus for the target tone in first position. This network mediates articulatory encoding in immediate-response, and articulatory rehearsal in delayed-response paradigms. These findings support a working memory model in which rehearsal is optional, while encoding is an obligatory component of the phonological loop. *NeuroReport* 15:2459–2462
© 2004 Lippincott Williams & Wilkins.

Key words: Auditory; Encoding; fMRI; Immediate memory; Lexical tone; Mandarin Chinese; Phonetics; Phonology; Rehearsal; Verbal working memory

INTRODUCTION

A prevailing view of the neuroanatomical substrates of verbal working memory (VWM) and language processing links the left inferior parietal lobe (IPL) with phonological storage, Broca's area and right lateral cerebellum with rehearsal, and the prefrontal cortex with different executive processes [1,2]. The relationship between VWM and language processing, however, remains debatable [2], especially the role of parietal cortex in phonological storage and the role of Broca's area in rehearsal.

Using delayed-response paradigms to enhance rehearsal is a key design feature in functional neuroimaging studies of VWM. In contrast, behavioural VWM studies use immediate-response paradigms to detect short-term effects without rehearsal [3]. Immediate memory is the initial short-term trace from encoding. Recent electrophysiological studies have demonstrated that immediate memory does exist, and is encoded within a few hundred milliseconds after the presence of target stimulus [4,5]. Neuropsychological evidence shows that it is necessary to distinguish encoding and rehearsal in working memory systems [6]. Immediate memory performance has been shown to vary as a function of encoding during stimulus input [7]. Immediate memory has a very limited capacity of about three to four novel items [8], lasts only a few seconds [9,10], and must be re-activated periodically, presumably through subvocal rehearsal, for extended processing [10]. Since delayed recall cannot increase memory span [11], rehearsal during the delay interval can only re-activate immediate memory. Thus,

delayed recall in VWM studies may involve a potential confound between language and memory processing beyond immediate memory.

VWM has been implicated in phonological processing of lexical tones [12–14]. Tonal processing typically engages a left-sided, frontoparietal cortical network that includes the middle frontal gyrus (MFG), inferior frontal gyrus (IFG), supplementary motor area (SMA), and superior parietal lobule (SPL). However, these aforementioned experiments were not specifically designed to isolate immediate memory.

The aim of this fMRI study was to identify neuroanatomical substrates of immediate memory that subserve subsyllabic processing of Mandarin tones (e.g., *ma*¹ 'mother', *ma*² 'hemp', *ma*³ 'horse', *ma*⁴ 'scold'). To distinguish the encoding of immediate memory from rehearsal, we employ an immediate recognition paradigm in which a lexical tone of a syllable in a list is matched to the tone of a following probe. The short, filled interval between target and probe falls within the time limit for a span length. Subjects must therefore execute the tone matching task rapidly without rehearsal before the immediate memory fades out. This paradigm effectively minimizes articulatory rehearsal. The presence of intervening distractors maximizes articulatory encoding of immediate memory.

Specifically, we compare two experimental conditions distinguished by the presence or absence of distractors. In one condition, two distractor syllables intervene between the target syllable and its probe; in the other condition, there are no intervening distractors. The distractors act as a suffix

to the targets and eliminate their auditory sensory memory [10], as well as divide the listeners' attention. To reduce these interference effects, listeners are likely to enhance their encoding of auditory input by covert articulation in order to increase their memory performance [10]. In the no-distractor condition, auditory traces of the targets immediately preceding the probes remain intact, requiring less articulatory encoding. Such an immediate recognition paradigm should elicit differential strengths of memory encoding by comparing trials with and without distractors intervening between targets and their probes. We hypothesize that even in the absence of rehearsal, the same network that subserves encoding, retrieval, and storage in VWM should similarly be recruited for immediate memory processing.

MATERIALS AND METHODS

Subjects: Twelve adult native speakers of Mandarin Chinese (six male; six female) from mainland China, ranging in age from 23 to 32 years, participated in this study. All subjects were strongly right-handed ($\geq 78\%$) as measured by the Edinburgh Handedness Inventory [15] and exhibited normal hearing sensitivity at 0.5, 1, 2, and 4 kHz. All subjects gave informed consent in compliance with a protocol approved by the Institutional Review Board of Indiana University Purdue University Indianapolis and Clarian Health Partners.

Stimuli: A list of three Chinese monosyllables followed by a monosyllabic probe made up a sequence for each trial. All stimuli were produced by a native Chinese male speaker from Beijing. In matching sequences for tone in the first syllable and its probe (FP; F=first, P=probe), only the first syllable in the syllable list shared a tone in common with its probe (Table 1). In matching sequences for tone in the last syllable and its probe (LP; L=last, P=probe), only the last syllable in the syllable list shared a tone in common with its probe. No other phonetic units in any of the syllables matched those in the probe. In non-matching sequences for either LP or FP, no phonetic units (i.e., onset, rhyme, tone) matched those in the probe. Matching and non-matching sequences were equal in number. No syllable was used more than once for a task. The FP and LP tasks shared the same stimuli except for the target positions (Table 1). Thirty-two sequences of a syllable list and probe were constructed for FP and LP, respectively.

Tasks: The scanning run consisted of two tasks (FP, LP) presented in blocked format (36s) in an alternating boxcar design with 18s rest intervals separating the task blocks. The order of presentation of FP and LP blocks was counter-balanced across subjects. Half of the subjects started with FP, the other half with LP. Total duration of the run was 7.5 min. There were four blocks per task. Each block contained eight

4.5s trials, four matching and four non-matching, in random order. The timing architecture of a trial consisted of the syllable list and probe (2300ms) and the response interval (2200ms). Stimulus onset asynchrony within the syllable list was 500ms. A silent interval of 350ms was inserted between the syllable list and its probe. Each sequence of four syllables (list + probe) fell within the span for immediate memory and attention [8,16]. Instructions were delivered to subjects via headphones near the end of the rest intervals immediately preceding each task (in Chinese): "first tone" for FP and "last tone" for LP. For FP, subjects were asked to ignore the two intervening distractors in the syllable list and make a matching judgment on the tone in the first syllable in comparison to the probe. For LP, subjects were asked to ignore the first and second syllables in the syllable list and make a matching judgment on the tone in the last syllable in comparison to the probe. They responded by pressing a mouse button (left=yes, right=no). Prior to the actual scanning session, subjects were trained to a high level of accuracy ($\geq 90\%$) on both tasks using stimuli different from those used during image acquisition.

Image acquisition: Scanning was performed on a 1.5T Signa GE LX Horizon scanner (Waukesha, WI) equipped with birdcage transmit-receive radiofrequency head coils. One 200-volume echo-planar imaging (EPI) series was begun with a rest interval consisting of eight baseline volumes (18s), followed by 184 volumes during which the two comparison tasks (36s) alternated with 18s rest intervals, and ended with a rest interval of 8 baseline volumes (18s). Functional data were acquired using a gradient-echo EPI pulse sequence with the following parameters: repetition time (TR) 2.25s; echo time (TE) 50ms; matrix 64×64 ; flip angle (FA) 90° ; field of view (FOV) 24×24 cm. Sixteen 7.5mm contiguous axial slices were used to image the entire cerebrum and superior aspects of the cerebellum. Prior to functional imaging runs, high-resolution, anatomic images were acquired in 124 contiguous axial slices using a 3D Spoiled-Grass (3D SPGR) sequence (slice thickness 1.2–1.3mm; TR 35ms; TE 8ms; 1 excitation; FA 30° ; matrix 256×128 ; FOV 24×24 cm) for purposes of anatomical localization and co-registration to a standard stereotactic system [17]. Subjects were scanned with eyes closed and room lights dimmed. The effects of head motion were minimized by using a head-neck pad and dental bite bar.

Image analysis: Image analysis was conducted using the AFNI software package. For each subject, all image volumes in a functional imaging run were motion-corrected to the fourth acquired volume. The time series of each voxel was detrended using a second order polynomial fit to account for temporal variation in the signal baseline, and then normalized to its mean intensity. The hemodynamic response functions for FP and LP tasks were separately deconvolved from the baseline (rest) for each voxel in the imaging volume. Magnitude of the response for each condition, corresponding to the area under the hemodynamic response curve, was evaluated by a general linear model approach. This analysis yielded differential responses illustrated by Student's *t*-statistic maps (FP vs baseline, LP vs baseline). These activation datasets, and the anatomic datasets for the corresponding subjects, were then interpolated to isotropic 1mm^3 voxel volumes and transformed to a stereotaxic coordinate system [22]. Student's

Table 1. Sample auditory stimuli for tone matching.

Task	Target position	Syllable list	Probe	Match
FP	First	duo ⁴ rou ² bei ³	na ⁴	yes
		tu ¹ bo ² ka ³	meng ⁴	no
LP	Last	rou ² bei ³ duo ⁴	na ⁴	yes
		bo ² ka ³ tu ¹	meng ⁴	no

Note. Chinese syllables are written in *pinyin* transcription. Tones are indicated in superscript. Target tones are in bold.

t-statistic values in each 1 mm^3 voxel were spatially smoothed by a 5.2 mm full-width half-maximum (FWHM) Gaussian filter chosen to reflect functional image resolution and minimize effects of the intersubject anatomic variability. The *t*-values of each voxel on the resulting statistical maps were converted to *z*-scores via the corresponding *p* value, and averaged over all 12 subjects. These average voxel-by-voxel activation (FP vs LP, FP vs baseline, LP vs baseline) maps were generated solely to illustrate the results of ROI-based (region of interest) ANOVAs.

ROI analysis: ROIs were defined anatomically in each hemisphere of individual subjects for regions implicated in verbal working memory [1,2,18]. The borders of each anatomic ROI were defined by key sulcal landmarks used in parcellation schemes to delimit cortical gyri or cerebellar lobules [19–22]. Each ROI volume was obtained by manually tracing slice-by-slice either in the sagittal or axial (cerebellum) plane.

The IFG ROI captured the pars opercularis (BA 44) and a small part of the pars triangularis (BA 45) using as borders the inferior frontal sulcus (IFS), inferior precentral sulcus (PrCS), Sylvian fissure, and an anterior margin 20 mm from the PrCS/IFS junction. The MFG ROI captured mainly BA 9 using as borders the IFS, inferior PrCS, and an anterior margin 20 mm from the PrCS/IFS junction. The SPL ROI captured BA 5/7 using as borders the postcentral sulcus (PoCS), fundus of the PoCS, and a posterior margin 20 mm from the PoCS/IPS junction. The pre-SMA ROI captured part of the medial frontal gyrus (BA 6) using as borders the cingulate sulcus, and anterior and posterior margins at $y=0$ and 20 mm, respectively. The CBL ROI captured mainly the lateral cerebellum (simple lobule) using as borders the superior posterior sulcus and a medial margin at $x=20$ mm.

For each ROI, mean *z*-scores were calculated for each combination of task (FP, LP) and hemisphere (LH, RH) for every subject. These *z*-scores were analyzed using planned contrasts between the two tasks on each hemisphere and between the two hemispheres on each task.

RESULTS

Behavioral data: Mean (\pm s.d.) response accuracy was not significantly different between the FP ($98 \pm 4.1\%$) and LP ($98.3 \pm 2.4\%$) tasks ($t(22)=-0.37$, n.s.). Neither was mean response time FP (502 ± 163 ms); LP (535 ± 192 ms); $t(22)=-0.46$, n.s.).

ROI data: The FP task elicited greater activation than the LP in the left IFG ($F(1,10)=5.02$, $p<0.0489$), right lateral CBL ($F=5.58$, $p<0.0398$) and bilateral and medial pre-SMA (LH, $F=18.56$, $p<0.0015$; RH, $F=17.43$, $p<0.0019$; Fig. 1). Both tasks revealed a leftward asymmetry in the IFG (FP, $F=60.87$, $p<0.0001$; LP, $F=11.74$, $p<0.0065$), MFG (FP, $F=10.77$, $p<0.0083$; LP, $F=10.54$, $p<0.0088$), pre-SMA (FP, $F=6.63$, $p<0.0277$; LP, $F=5.96$, $p<0.0347$) and SPL (FP, $F=24.60$, $p<0.0006$; LP, $F=24.12$, $p<0.0006$). A rightward asymmetry was observed for both tasks in the CBL (FP, $F=13.06$, $p<0.0047$; LP, $F=5.84$, $p<0.0363$).

DISCUSSION

This study reveals a neural network that subserves articulatory encoding of auditory input for verbal immediate memory in

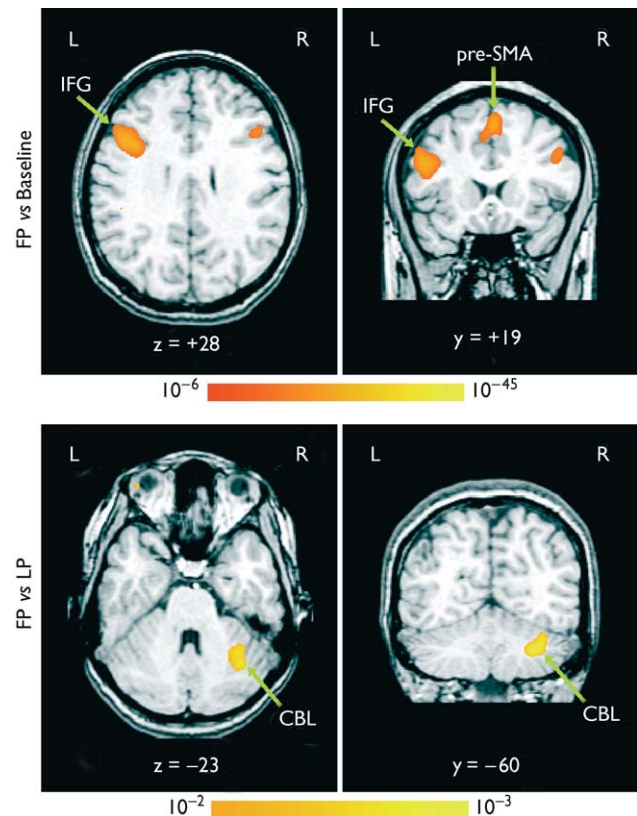


Fig. 1. Averaged fMRI activation maps obtained by comparing tone matching judgments in a distractor condition relative to a resting baseline (top panels) and relative to a no-distractor condition (bottom panels). Each row shows axial and coronal sections in stereotaxic space of activation maps superimposed onto a representative brain anatomy. The top panels (FP vs Baseline) show greater activation in the left IFG and the bilateral pre-SMA for the FP task. The bottom panels (FP vs LP) show greater activation in the right lateral CBL for tone matching judgments in a distractor relative to a no-distractor condition. IFG, inferior frontal gyrus; SMA, supplementary motor area; CBL, cerebellum. Stereotaxic coordinates (mm) are derived from the human brain atlas of Talairach and Tournoux (1988). Activation maps for FP vs Baseline and FP vs LP are thresholded at $p=10^{-6}$ and $p=10^{-2}$, respectively; *z*-values are represented by a color scheme where the significance levels increase from red to yellow.

the absence of rehearsal via an immediate-recognition paradigm. Increased activations were found in the left IFG, right lateral CBL, and medial pre-SMA for task FP relative to LP. The co-activation of the left IFG and right CBL has been implicated in rehearsal for VWM studies [1,2]. Their co-activation in the present study, however, reflects the articulatory encoding of auditory input induced by the major difference between tasks FP and LP. In task FP, distractors intervene between the target and probe. They act as a suffix to the target and eliminate the echoic memory of auditory stimuli [10]. To increase their memory performance, listeners likely encode the auditory input subvocally [10]. In task LP, on the other hand, the target is immediately followed by the probe. The auditory information is intact for encoding, and therefore requires less articulatory processing.

Pre-SMA has been associated with internal guidance [14], working memory delays [23], and voluntary selection of actions [24,25], as part of a frontoparietal network [25]. The FP task requires more procedural guidance than LP to handle the

intervening distractors. In the LP task, subjects listen to the last two syllables and make tone matching judgments. The LP task therefore requires less internal guidance. It appears that activation of the pre-SMA is due to task control demands instead of articulatory encoding itself. We conclude that the left IFG and right lateral CBL is a fronto-cerebellar network for articulatory encoding that is independent of rehearsal.

The identification of a fronto-cerebellar network for articulatory encoding that is isolated from rehearsal is of theoretical significance. In earlier models of working memory [10], rehearsal plays a prominent role in the phonological loop. Its role in VWM has been challenged recently to account for findings in behavioral and neuropsychological studies [3]. By the revised model, rehearsal becomes optional, whereas encoding from stimuli to memory is specified as a crucial component of the phonological loop [1]. Consistent with this revised model, our findings reveal neuroanatomical substrates for articulatory encoding of auditory input in the absence of rehearsal.

This fronto-cerebellar network underlies both encoding and rehearsal in VWM. In this study, encoding of verbal immediate memory in the FP task involves subvocal articulation. It is no coincidence that the same fronto-cerebellar network has been implicated in rehearsal for VWM delayed-recall tasks [1,2]. Rehearsal during a delayed recall task reactivates immediate memory. The net effect is that the cycles of rehearsal work in a way similar to periodic encoding mediated by articulation. In each cycle of rehearsal, stored phonological codes are necessarily retrieved by overt or covert articulation. This view is in agreement with neuropsychological and brain imaging data, which show that the left IFG is also crucially involved in delayed-response VWM tasks (see [2] for review).

We infer that MFG and SPL activity reflects attentional modulation of lexical tone processing. Because of the absence of a task effect, their increased activity relative to baseline is likely due to tone extraction and tone matching. The attentional demands associated with these processes are common to both the FP and LP tasks. This interpretation is compatible with previous findings which reveal increased activity in these same regions that was elicited by a task identical to LP [14].

CONCLUSIONS

This fMRI study demonstrates that tonal recognition in an immediate-response paradigm with intervening distractors recruits a fronto-cerebellar network for articulatory encoding. This network enhances articulatory encoding in immediate memory. This same network appears to mediate rehearsal in delayed-response paradigms. These findings converge to support a revised working memory model in which rehearsal is optional, while encoding is an obligatory component of the phonological loop.

REFERENCES

1. Baddeley A. Working memory: looking back and looking forward. *Nature Rev Neurosci* 2003; **4**:829–839.
2. Chein JM, Ravizza SM and Fiez JA. Using neuroimaging to evaluate models of working memory and their implications for language processing. *J Neurolinguistics* 2003; **16**:315–339.
3. Nairne JS. Remembering over the short-term: the case against the standard model. *Annu Rev Psychol* 2002; **53**:53–81.
4. Halgren E, Boujon C, Clarke J, Wang C and Chauvel P. Rapid distributed fronto-parieto-occipital processing stages during working memory in humans. *Cerebr Cortex* 2002; **12**:710–728.
5. Naatanen R. The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology* 2001; **38**:1–21.
6. Henson RN, Burgess N and Frith CD. Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia* 2000; **38**:426–440.
7. Cowan N, Sauls JS and Brown GDA. On the auditory modality superiority effect in serial recall: separating input and output factors. *J Exp Psychol Learn Mem Cogn* 2004; **30**:639–644.
8. Cowan N. The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behav Brain Sci* 2000; **24**: 87–185.
9. Schweickert R and Boruff B. Short-term memory capacity: magic number or magic spell? *J Exp Psychol Learn Mem Cogn* 1986; **12**:419–425.
10. Baddeley A. *Working Memory*. Oxford: Oxford University Press; 1986.
11. Turvey MT, Brick P and Osborn J. Proactive interference in short-term memory as a function of prior-item retention interval. *Q J Exp Psychol A* 1970; **22**:142–147.
12. Gandour J, Wong D, Lowe M, Dziedzic M, Saththamuwong N, Tong Y and Li X. A cross-linguistic fMRI study of spectral and temporal cues underlying phonological processing. *J Cogn Neurosci* 2002; **14**:1076–1087.
13. Gandour J, Xu Y, Wong D, Dziedzic M, Lowe M, Li X and Tong Y. Neural correlates of segmental and tonal information in speech perception. *Hum Brain Mapp* 2003; **20**:185–200.
14. Li X, Gandour J, Talavage T, Dziedzic M, Lowe M, Li X and Tong Y. Selective attention to lexical tones recruits left dorsal frontoparietal network. *Neuroreport* 2003; **14**:2263–2266.
15. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 1971; **9**:97–113.
16. Zhang G and Simon HA. STM capacity for Chinese words and idioms: chunking and acoustical loop hypotheses. *Mem Cogn* 1985; **13**:193–201.
17. Talairach J and Tournoux P. *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System: An Approach to Cerebral Imaging*. New York: Thieme Medical Publishers; 1988.
18. Newman SD, Just MA and Carpenter PA. The synchronization of the human cortical working memory network. *Neuroimage* 2002; **15**:810–822.
19. Caviness JV, Meyer J, Makris N and Kennedy D. MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *J Cogn Neurosci* 1996; **8**:566–587.
20. Makris N, Hodge SM, Haselgrove C, Kennedy DN, Dale A, Fischl B *et al*. Human cerebellum: surface-assisted cortical parcellation and volumetry with magnetic resonance imaging. *J Cogn Neurosci* 2003; **15**:584–599.
21. Rademacher J, Galaburda A, Kennedy D, Filipek P and Caviness JV. Human cerebral cortex: localization, parcellation, and morphometry with magnetic resonance imaging. *J Cogn Neurosci* 1992; **4**:352–374.
22. Schmahmann JD, Doyon J, McDonald D, Holmes C, Lavoie K, Hurwitz AS *et al*. Three-dimensional MRI atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage* 1999; **10**:233–260.
23. Petit L, Courtney SM, Ungerleider LG and Haxby JV. Sustained activity in the medial wall during working memory delays. *J Neurosci* 1998; **18**: 9429–9437.
24. Crosson B, Sadek JR, Maron L, Gokcay D, Mohr CM, Auerbach E *et al*. Relative shift in activity from medial to lateral frontal cortex during internally vs externally guided word generation. *J Cogn Neurosci* 2001; **13**:272–283.
25. Deiber MP, Ibanez V, Sadato N and Hallett M. Cerebral structures participating in motor preparation in humans: a positron emission tomography study. *J Neurophysiol* 1996; **75**:233–247.

Acknowledgements: NIH research grant R01 DC04584-04 (J.G.); NIH postdoctoral traineeship at Purdue University (X.L.). Portions of this research were presented at the 10th International Conference on Functional Mapping of the Human Brain, Budapest, June 2004, and the 28th International Congress of Psychology, Beijing, August 2004. We thank T. Osborn and J. Zimmerman for their technical assistance in the MRI laboratory.