Cortical organization of sensory corrections in visuomotor skill acquisition

S. Mitra [a],* A. Bhalerao [b] P. Summers [c] S. C. R. Williams [d]

[a] *Department of Psychology and Institute of Applied Cognitive Science, University of Warwick, Coventry, CV4 7AL, UK*

[b] *Department of Computer Science, University of Warwick, Coventry, CV4 7AL, UK*

[c] *Department of Neuroradiology, Medical Sciences, University of Oxford, Oxford OX3 9DH*

[d] *Department of Neurology, Institute of Psychiatry, King's College, London SE5 8AF, UK*

Abstract

During sensorimotor skill acquisition, early learning of the required neuromuscular pattern and sensorimotor mappings is followed by an intermediate stage of gradually increasing consistency and efficiency of execution, which gives way, with persistent practice, to the later stages of automatization. It has been suggested that the intermediate stage is distinguished by refinements in the background sensory corrections that support, stabilize and smoothen the fine motor adjustments required by the new coordination. While the later stages of motor refinement are thought to be sub-cortically organized, the neurophysiology of the proposed sensory learning component in the intermediate stage is not well understood. During explicit learning of a visually cued finger-tap sequence, the present research used fMRI to isolate those cortical activations that were significant in the immediate post-learning phase, but were not also observed during the corresponding pre-learning phase. Such exclusively post-learning activation occurred significantly more in visual and somatosensory association areas, than in primary somatosensory or primary and secondary motor areas. These results show that the intermediate stage of skill acquisition has a significant sensory learning component, and that the process has observable cortical correlates.

Key words: Motor Learning, skill acquisition, fMRI
PACS:

1 Introduction

During the acquisition of a sensorimotor skill, progress from the hesitant, inaccurate and inconsistent performance of the novice to the smooth, accurate and reliable execution of the expert is thought to occur in several identifiable stages [1–3]. In both Bernstein's [1] and Fitts' [2] classic theories, three stages are proposed. In the early phase, the learner gets acquainted with the essential nature of the required coordination (i.e., the correct sequence of what needs to be done). During this phase, there are explicit analyses of performance and consciously introduced changes in execution, and the task requires dedicated attention. Once the basic form of the task has been learned, learning goes through an intermediate phase of reducing performance variability, and increasing consistency and stability of coordination. The final stage of learning occurs over prolonged, persistent practice, whereby the attention and effort required at execution time is markedly diminished – a process referred to as automatization – and the neural support for the coordination is thought to become mostly sub-cortical in organization [4,5].

The present study was concerned with the suggestion that the intermediate stage of learning is distinguished by refinements in the background sensory corrections that are necessary to support, stabilize and smoothen the fine motor adjustments required by the new coordination. A key feature of skilled motor coordination is its resistance to perturbation. Execution-time perturbations can originate externally or internally. External sources include mechanical disturbances such as collisions or perceived imminent onset of such disturbances. Internal perturbations include effects of physiological noise or conduction delays on the unfolding sequence of motor neuron activations. The role of background sensory corrections is to detect perturbations at execution time and adaptively alter motor neuron activation patterns to keep the coordination stable and smooth [6,1,3]. Learning to identify coordination-relevant perturbations, and how these may affect performance, is an involved process – conduction time delays involved in perceptual information-gathering mean that corrections to motor neuron activations based upon such information must be anticipatorily applied. Yet, this process is fundamentally important in becoming skilled at a motor task. Before this intermediate-level learning takes place, modulations of motor neuron activations often require conscious intervention, and performance does not attain speed, smoothness or stability.

This view of intermediate learning has its origin in observed patterns of behavioural data [1–3], but its neural basis remains unspecified. Neurophysiolog-

———
* Department of Psychology and Institute of Applied Cognitive Science, University of Warwick, Coventry, CV4 7AL, UK
    Email address: Subhobrata.Mitra@warwick.ac.uk (S. Mitra).

ical studies of motor learning have tended to focus more broadly on either the differences in neural organization between the pre- and post-learning phases of performance [7–13], or the organizational changes that accompany the time course of learning [14,15]. The sensory corrections hypothesis predicts that, in the time period immediately following the early stage of learning (i.e., acquisition of the basic form of the required motor sequence), mappings between motor sequence generation and its various perceptual correlates would be acquired. As such, for a visually guided tapping task, for example, it should be possible to observe patterns of neural activity in visual and somatosensory association areas that were not also present in the early stage of learning. Several recent studies on (explicit) motor-sequence learning have reported practice-dependent activity in the parietal region [9,14,11,16,15], while others have reported greater parietal activity during novel motor-sequence learning compared to performance of well-learned sequences [10,12,13]. Even though this evidence does point to the existence of heightened visual and somatosensory association area activity during the time course of skill learning, it remains unclear whether this is specifically a characteristic of an identifiably intermediate stage that follows the early stage of learning.

The present study used explicit learning of a visually cued finger-tap sequence to address this question. A sequence of key presses across multiple fingers is a common everyday skill (e.g., typing), but the process of learning to rapidly, accurately, and smoothly generate such sequences presents many of the challenges involved in more complex coordinations. Performing this task involves accurate phasing of time-overlapped waves of motor neuron activity across the involved fingers. Since relative timing is crucial, perturbations (e.g., due to physiological noise) to motor neuron activity at one location may need to be counteracted by rephasing of activity at multiple locations and time points. In the visually cued version of the task used here, completion of each key press generates visual feedback. Somatosensory feedback is, however, considerably richer as it is modulated by nearly continuous changes in pressure applied by the fingers on the keys. Once the sequence of required key presses is memorized during early learning, further improvements in speed and stability of performance during the intermediate phase involve learning to accurately detect relative phase (and its deviations) and effect subtle changes in motor neuron activity as a function of these deviations.

Thus, during explicit learning of a visually cued finger-tap sequence, we used fMRI to isolate, for each individual participant, those cortical activations that were significant in the period immediately following the early learning phase, but were not also observed during the early phase itself. For each participant, the end of the early learning phase (and the start of the intermediate phase) was defined as the point beyond which the participant showed perfect recall of the required sequence of finger-taps. The specific hypothesis we tested was this: If there does exist an intermediate stage of learning that is distinguished from

3

the early phase by the refinement of background sensory corrections, then the intermediate-stage activations we isolated would be localized particularly in the occipital-parietal-temporal association areas. Thus, activations specific to the intermediate stage would not occur mostly in primary and secondary motor and somatosensory regions, as would be the case in early learning. The regions of interest included primary and secondary motor and somatosensory areas, where early learning activity would be expected, and the occipital-temporal and parietal-temporal association areas (Fig. 3), where, by hypothesis, activity specific to the intermediate stage of skill acquisition would be concentrated. Acquisition of explicit sequence recall, as required in the task used here, has been associated with increasing pre-frontal activity that is thought to reflect a working memory process [17]. Tertiary motor areas of the prefrontal cortex were, therefore, also included in the analysis.

The physiological changes expected in the transition to late-stage learning were not of particular interest in this study because the duration of single-session learning used here was too short for late-stage learning to occur. However, we still tested post-learning brain activations for habituation effects that would signal late learning processes.

2 Methods and Materials

The task resembled the well-known serial reaction time (SRT) task, and involved explicit learning of a single sequence of 10 taps across four digits of the dominant, right hand (e.g., 2-3-2-1-4-1-3-4-2-4). Participants were told that they would be repeatedly presented with the same sequence of 10 finger taps, and their task would be to learn to reproduce the sequence as accurately and quickly as possible. They were told that throughout the session, each presentation of the sequence would be followed by a test of recall.

The four digits of the participant's right hand rested on four buttons of a keypad. On the viewing screen, a motif of a right hand remained visible throughout the 19.2 minute scanning session. The session contained alternating REST, LEARN, and TEST blocks, each lasting 12 s, and cycling throughout in the order REST-LEARN-REST-TEST. During the 12 s REST blocks, only the hand motif appeared and the participant was instructed to simply look at it. In each LEARN block, the sequence of 10 finger-presses was visually cued by a sequence of green dots appearing above the fingers to be tapped. The interval between successive cues was 1.2 s. During the TEST blocks, a green dot was present above all the four digits of the hand motif. As the participant reproduced the learned sequence by tapping the keypad buttons, the green light corresponding to each tap blinked to confirm which button had been pressed. Participants were instructed to 'soldier' through the sequence

4

(or the portion of it they remembered) as many times as they could during each 12 s TEST block. A novel tapping sequence was generated randomly for each participant, and a total of 9 right-handed, male participants were tested (in accordance with the guidelines and permit of the Institute of Psychiatry's ethics committee for research).

Functional MR imaging was performed on a 1.5T GE Signa-LX system. The fMRI and 3D anatomical reference images were acquired in an axial orientation, prescribed from an initial sagittal localizer scan. Single-shot, gradient-echo, echo-planar MR images covering the whole head were acquired repeatedly over the 20 minute task session. Scanning parameters for the EPI data were: 16 slices, 7 mm thick, 0.7 mm slice separation, 24 cm FOV, 64 x 64 matrix, repetition time (TR) of 3 s with data acquisition centred at an echo time (TE) of 40 ms. For each participant, a total of 388 image volumes were acquired over 19.2 min. A high resolution, anatomical reference scan consisting of a 124 slice 3D IR-SPGR sequence at the same FOV as the fMRI scan was also acquired. The sequence parameters were: TR 15.9 ms, TE 5.2 ms, inversion time 300 ms, slice thickness 1.5 mm, and a 256 x 192 imaging matrix.

Data preprocessing and GLM-based analysis was performed using SPM99b [18]. The image data were first realigned to minimize head motion between scans. The high-resolution structural MR image (3D IR-SPGR) was then aligned to the mean of the T2* weighted functional images which were then spatially normalized to the MNI template [7]. The functional data were then smoothed using a Gaussian smoothing kernel of 5 mm at FWHM to improve the SNR and 'condition' the data to conform more closely to the Gaussian field model [18].

Experimental conditions were modelled using a box-car design (Fig. 2) and a linear least-squares regression was performed using a time-locked square-wave function convolved with an estimated haemodynamic response function in the standard way on the pre-processed data. Over the course of each participant's session, the REST-LEARN-REST-TEST cycle of the 10-tap sequence was presented 24 times. Each participant's LEARN (and TEST) sequences were then divided into a PRE-LEARN (PRE-TEST) and a POST-LEARN (POST-TEST) phase at the point where the participant began to reproduce the entire 10-tap sequence correctly during TEST (Fig. 2). Across the 7 participants who succeeded in learning their sequence, this occurred after 3 to 14 repetitions of the REST-LEARN-REST-TEST cycle (Fig. 1A). The POST-LEARN (and POST-TEST) phase was further subdivided into periods of 6 presentation cycles each (e.g., POST-LEARN1, POST-LEARN2; see Fig. 2) to allow testing for any habituation effects. Due to the different times at which participants learned the sequence, data for only the first 2 POST-LEARN (and POST-TEST) periods were available for all participants. The effects of interest were POST-LEARN1 and POST-LEARN2 each masked by PRE-LEARN, and POST-TEST1 and POST-TEST2 each masked by PRE-TEST.

In each case, an exclusive mask was applied using a Z threshold of $p < .25$ on the PRE-LEARN (or PRE-TEST), followed by an SPM{Z} threshold at $p < .0025$ (Fig. 4). This exclusive masking removed all voxels that also reached the default level of significance in the respective masking contrast (i.e., in PRE-LEARN or PRE-TEST).

In view of the different learning speeds across participants, SPM results of participants were not averaged. For each participant, we grouped the significant voxel activations (i.e., above the SPMZ score $p < .0025$) during PRE-LEARN, PRE-TEST, POST-LEARN1, POST-LEARN2, POST-TEST1 and POST-TEST2, in each hemisphere, into six functional regions of interest: primary motor (Brodmann Area 4), secondary motor (BA 6, 8, 44), tertiary motor (BA 9, 10, 11, 45, 46, 47), primary somatosensory (BA 1, 2, 3), parieto-temporal association (BA 7, 22, 39, 40), and occipito-temporal association (BA 18, 19, 20, 21, 37), using the Talairach coordinates of SPM activation [19]. The Talairach daemon provided Brodmann area labels with a spatial accuracy within a 5 mm voxel size. Each activation was binned to the nearest region if within a 5mm radius.

3 Results and Discussion

3.1 Summary of Results

Seven out of 9 participants learned to accurately produce the assigned sequence within the 24 presentation cycles. Across these participants, perfect recall of the sequence occurred after 3 to 14 presentation cycles (Fig. 3). Analysis of tapping latencies showed that, in the LEARN condition, tapping latencies reduced significantly only over the first quarter of the session, whereas, in the TEST condition, participants kept getting faster over the entire session (Fig. 1B). Analysis of cortical activations exclusive to the intermediate learning phase showed significantly more activation in the parieto-temporal and occipito-temporal association areas than in the primary motor and primary somatosensory areas. Activation in the pre-frontal tertiary motor areas was marginally greater than in the primary motor and somatosensory areas. When LEARN and TEST condition activation data were analysed separately, these effects were significant in the TEST but not in the LEARN condition.

3.2 Behavioural Data

As indicated in Figure 1A, response latencies in both LEARN and TEST conditions reduced over the course of the session. In order to quantitatively evaluate tapping latencies, the 24 LEARN and TEST blocks were grouped into 4 periods of 6 blocks each (see Fig. 1B), and a 2 (condition: LEARN, TEST) $\times$ 4 (period) within-subjects, repeated measures analysis of variance (ANOVA) was performed. The main effect of condition was significant ($F(1,6) = 15.99$, $p < .05$; TEST latencies were longer). The main effect of period was also significant ($F(3, 18) = 63.77$, $p < .05$; participants became faster with time), as was the condition $\times$ period interaction ($F(3, 18) = 3.52$, $p < .05$; response latency reduced more sharply and consistently in the TEST condition). Means comparisons showed that, in the LEARN condition, latency in the first period was significantly greater than in any of the next three ($p < .001$), but latencies in the later periods did not differ significantly from each other. In the TEST condition, latencies in each of the four periods differed from the others ($p < .01$). Note that the above grouping of LEARN and TEST cycles into 4 periods (Fig. 1B) was solely for the purpose of analyzing tapping latency data. These periods are unrelated to the division of each participant's session into one pre- and two post-learning periods (Fig. 2), as used below in analyses of cortical activation data.

3.3 Imaging Data

Pooling the LEARN and TEST data together, 2(learn/test) $\times$ 2(post-learn1/post-learn2) $\times$ 2 (left/right hem) $\times$ 6 (regions of interest) within-subjects, repeated-measures analysis of variance (ANOVA) was performed to analyze the distribution of activated voxels that survived exclusive masking across the regions of interest. Only the main effect of regions of interest was significant ($F(5, 30) = 3.652$, $p < .05$). Fisher's protected least significant difference (PLSD) post-hoc test showed that the number of activated voxels in the occipito-temporal and parieto-temporal association regions were both significantly greater ($p < .05$), and in the tertiary motor region marginally greater ($p < .06$), than in either the primary motor or primary somatosensory regions (Fig. 3). The number of activated voxels in the secondary motor region was significantly smaller than in the occipito-temporal association region, but no other pairwise differences in activation between the regions of interest were significant. When LEARN and TEST data were separately subjected to 2(post-learn1/post-learn2) $\times$ 2 (left/right hem) $\times$ 6 (regions of interest) within-subjects, repeated-measures ANOVAs, it was found that all the above effects were significant in TEST, but none was significant in LEARN.

These results show that, in the period that immediately followed the start of correct reproduction of the tapping sequence, there was significant activation in the visual and somatosensory association areas that did not also occur in the period before the required motor sequence had been learned. Our methodology allowed for individual differences in the time it took for correct reproduction of the sequence to begin. While the exclusive masking procedure left little primary and secondary motor, and primary somatosensory, activation unique to the post-acquisition period, there was significant such activation in pre-frontal tertiary-motor, and parieto-temporal and occipito-temporal association regions. Notably, this effect was statistically significant in the TEST conditions in which participants reproduced the learned sequence from memory (with visual feedback of results but not errors), but not in the LEARN conditions in which participants were being visually led through the sequence.

In the post-learning phase, the same sequence of motor activity was executed in both the LEARN and TEST conditions. However, in the LEARN condition, participants continued to passively follow the visual cue for each key tap, whereas, in the post-learning TEST condition, they were generating the sequence from memory. One explanation for significantly greater association region activation in TEST but not in LEARN could be that the association region activations that are exclusive to the post-learning phase are linked more strongly to the spontaneous generation of a newly learned skill than to the act of performing the movements of the skill per se. In the LEARN condition, the key coordination is between the visual onset of cues and their corresponding taps. Quick responding to visual cues, even for familiar sequences, is likely not nearly as dependent upon the relative phasing of motor neuron activations across the four fingers as is the spontaneous generation of the same sequence. As discussed earlier, accurate detection of relative phase and its deviations, and applying changes to motor neuron activations as a function of these deviations, is likely to be the main role of background sensory corrections in this task. If so, we would expect association region activations exclusive to the post-learning phase to be greater in the TEST than in the LEARN condition.

Even though the 10-tap sequence was not a particularly difficult task to learn, as shown in Figure 1, the skill acquisition process was far from complete once the sequence itself had been memorized – average tapping latency in the TEST condition continued to fall in nearly linear fashion well past the point of sequence-memorization. Thus, the post-learning phase studied here preceded the late stage of skill acquisition during which task performance becomes fluent and relatively effortless, and performance improvements, if at all detectable, slow down considerably. Here, it is worth noting that the absence of significant effects involving the two stages of post-learning practice suggests that habituation [17] did not play a role over the time-scale studied here. We conclude, therefore, that the post-learning association cortex activations isolated by our exclusive-masking procedure belong to an intermediate stage of

skill acquisition during which the background sensory corrections required to stabilize and automate the new coordination are acquired. Our results also suggest that the cortical network subserving online sensory corrections may develop mostly when practicing an already learned task, rather than during learning itself.
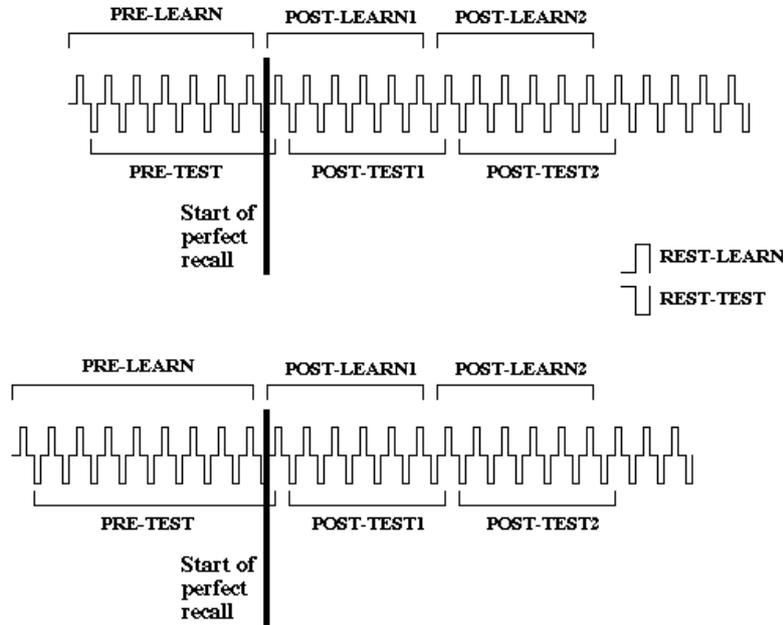

Fig. 1. Schematic of the tap-sequence learning task and determination of experimental conditions for two representative participants. Each participant received a total of 24 LEARN and TEST blocks in the order shown (REST-LEARN-REST-TEST). Each block lasted 12 s. For each participant, the starting point for perfect recall differed (e.g., after 7 presentation cycles for the top, and 9 cycles for the bottom participant). The blocks preceding this point were assigned to the PRE-LEARN (PRE-TEST) condition. For each participant, the following 12 block-cycles were broken into two POST-LEARN (POST-TEST) phases of 6 block cycles each.

Acknowledgements

References

[1] N. Bernstein. *Dexterity and its Development.* Erlbaum, Mahwah, NJ, 1996.

[2] P. M. Fitts. Perceptual-motor skill learning. In A. W. Melton, editor, *Categories of human learning*, pages 253–285. Academic Press, 1964.

[3] S. Mitra, P. Amazeen, and M.T. Turvey. Intermediate motor learning as decreasing active (dynamical) degrees of freedom. *Human Movement Science*, 17:17–65, 1998.

[4] V. B. Penhune and J. Doyon. Dynamic cortical and subcortical networks in learning and delayed recall of timed motor sequences. *Journal Neuroscience*, 22(4):1397–1406, 2002.

[5] I. Toni, N. Ramnani, O. Josephs, J. Ashburner, and R. E. Passingham. Learning arbitrary visuomotor associations: temporal dynamic of brain activity. *Neuroimage*, 14:1048–1057, 2001.

[6] N. Bernstein. *The coordination and regulation of movements*. Pergamon: Oxford, 1967.

[7] O. Hikosaka, K. Miyashita, S. Miyachi, K. Sakai, and X. Lu. Differential roles of the frontal cortex, basal ganglia, and cerebellum in visuomotor sequence learning. *Neurobilogy of Learning and Memory*, 70:137–149, 1998.

[8] O. Hikosaka, H. Nakahara, M. K. Rand, K. Sakai, X. Lu, K. Nakamura, S. Miyachi, and K. Doya. Parallel neural networks for learning sequential procedures. *Trends in Neuroscience*, 22:464–471, 1999.

[9] S. T. Grafton, E. Hazeltine, and R. B. Ivry. Abstract and effector-specific representations of motor sequences identified with PET. *Journal of Neuroscience*, 18:9420–9428, 1998.

[10] I. H. Jenkins, D. J. Brooks, P. D. Nixon, R. S. Frackowiak, and R. E. Passingham. Motor sequence learning: a study with positron emission tomography. *Journal of Neuroscience*, 14:3775–3790, 1994.

[11] K. Sakai, O. Hikosaka, S. Miyauchi, R. Takino, Y. Sasaki, and B. Putz. Transition of brain activation from frontal to prarietal areas in visuomotor sequence learning. *Journal Neuroscience*, 18:1827–1840, 1998.

[12] R. J. Seitz, P.E. Roland, C. Bohm, T. Greitz, and S. Stone-Elander. Motor learning in man: a positron emission tomography study. *Neuroreport*, 1:17–20, 1990.

[13] H. van Mier, L. W. Tempel, J.S. Perlmutter, M. E. Raichle, and S. E. Petersen. Changes in brain activity during motor learning measured with PET: Effects of hand of performance and practice. *Journal of Neurophysiology*, 80:2177–2199, 1998.

[14] R.-A. Müller and N. Kleinhans and K. Pierce and N. Kemmotsu and E. Courchesne. Functional MRI of motor sequence acquisition: effects of learning stage and performance. *Cognitive Brain Research*, 14:277–293, 2002.

[15] I. Toni, M. Krams, R. Turner, and R. E. Passingham. The time course of changes during motor sequence learning: a whole-brain fMRI study. *Neuroimage*, 8:50–61, 1998.

[16] W. R. Staines, M. Padilla, and R. T. Knight. Frontal-parietal event-related potential changes associated with practising a novel visuomotor task. *Cognitive Brain Research*, 13:195–202, 2002.

[17] J. C. Eliassen, T. Souza, and J. N. Sanes. Human brain activation accompanying explicitly directed movement sequence learning. *Experimental Brain Research*, 141:269–280, 2001.

[18] K.J. Friston, A.P. Holmes, K.J. Worsley, J.P. Poline, C.D. Frith, and R.S.J. Frackowiak. Statistical parametric maps in functional imaging: A general linear approach. *Human Brain Mapping*, 2:189–210, 1995.

[19] J. L. Lancaster, J. L. Summerln, L. Rainey, C. S. Freitas, and P. T. Fox. The Talairach Daemon, a database server for Talairach atlas labels. *Neuroimage*, 5(4):S633, 1997.

[20] H. A. Drury D. C. van Essen. Structural and functional analyses of human cerebral cortex using a surface-based atlas. *Journal of Neuroscience*, 17(18):7079–7102, 1997.
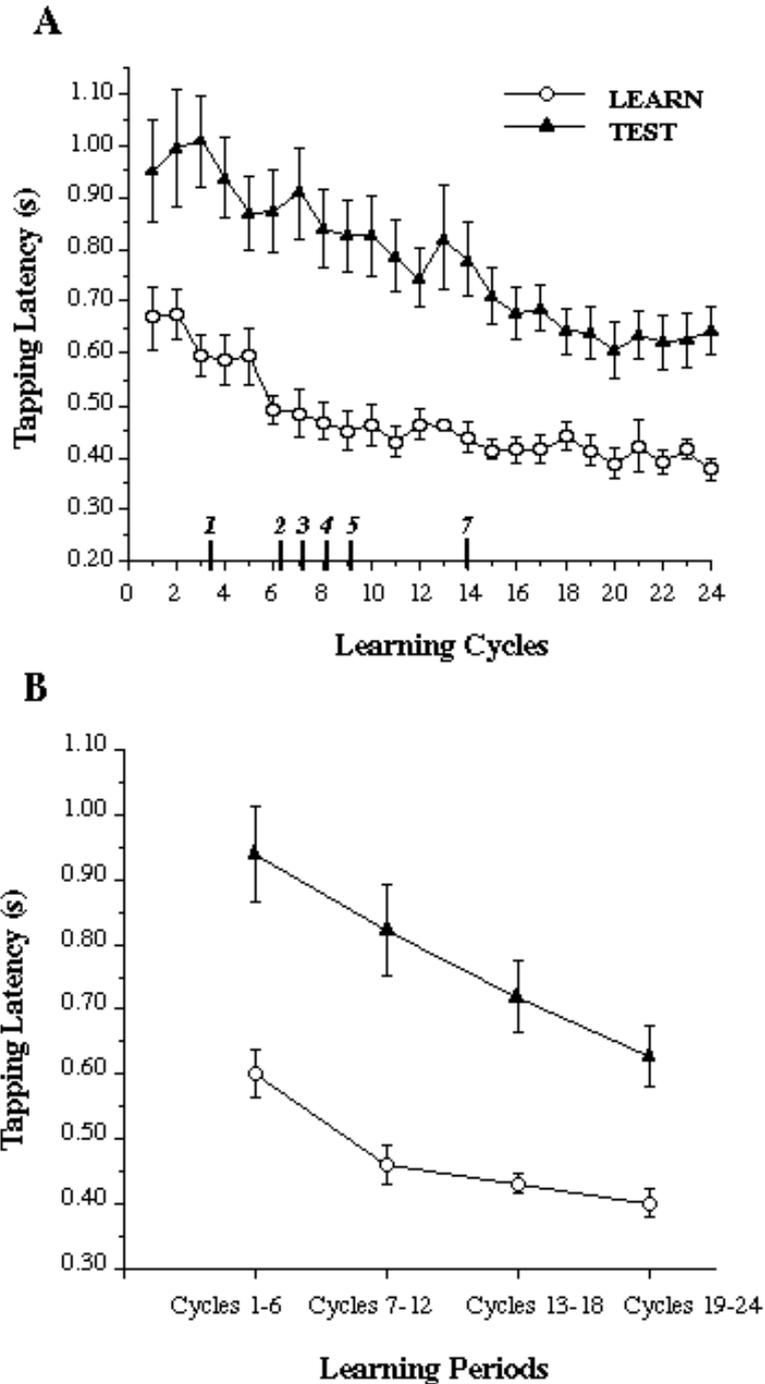
Fig. 2. **A**: Mean behavioural data for LEARN and TEST blocks over all 24 REST-LEARN-REST-TEST cycles. In LEARN, the latencies shown are between visual onset of the finger-tap cues and the participant's corresponding keypad-presses. In TEST, the shown latencies are inter-tap intervals. Numbers shown just above the horizontal axis are the number of participants who had learned the tap-sequence by that learning cycle. **B**: Mean response latencies grouped into periods of 6 cycles each. Error bars indicate standard error.
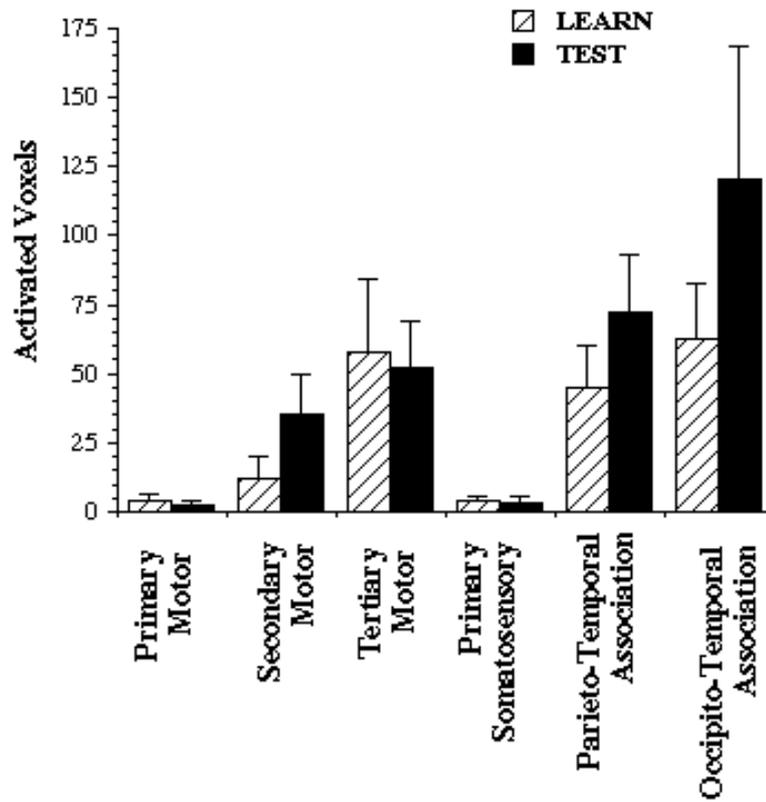
12

Fig. 3. Means of post-learn and post-test activations exclusive-masked by pre-learn and pre-test activations, respectively. Error bars indicate standard error.
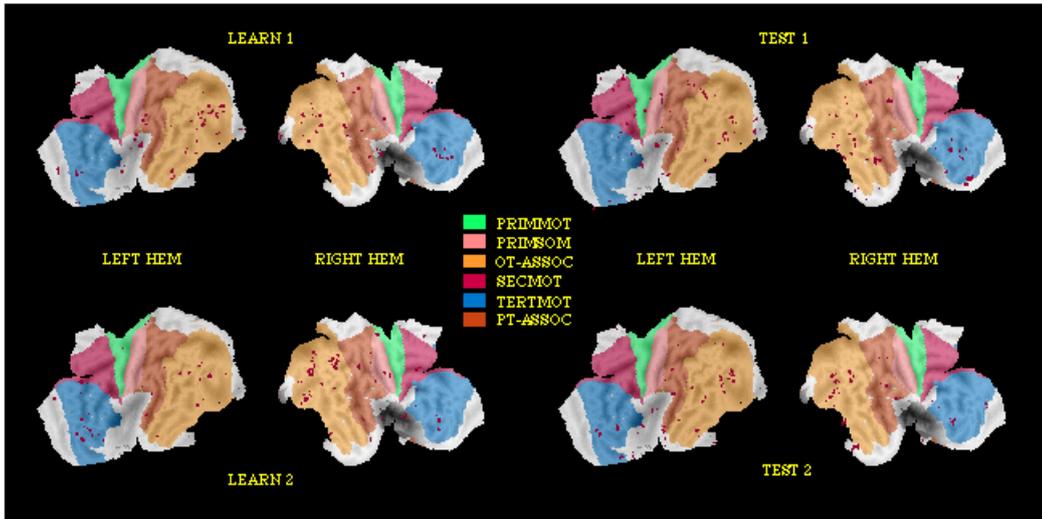
Fig. 4. Surface reconstructions [20] of the cortex showing activated voxels (across all 7 participants) by regions of interest. PRIMMOT: primary motor (BA 4); PRIM-SOM: primary somatosensory (BA 1, 2, 3); OT-ASSOC: occipito-temporal association (BA 18, 19, 20, 21, 37); SECMOT: secondary motor (BA 6, 8 44); tertiary motor (BA9, 10 11, 45, 46, 47); PT-ASSOC: parieto-temporal association (BA 7, 22, 39, 40).