RESEARCH ARTICLE

Katsuyuki Sakai · Katsuya Kitaguchi ·
Okihide Hikosaka

Chunking during human visuomotor sequence learning

Abstract Motor sequence learning is a process whereby a series of elementary movements is re-coded into an efficient representation for the entire sequence. Here we show that human subjects learn a visuomotor sequence by spontaneously chunking the elementary movements, while each chunk acts as a single memory unit. The subjects learned to press a sequence of 10 sets of two buttons through trial and error. By examining the temporal patterns with which subjects performed a visuomotor sequence, we found that the subjects performed the 10 sets as several clusters of sets, which were separated by long time gaps. While the overall performance time decreased by repeating the same sequence, the clusters became clearer and more consistent. The cluster pattern was uncorrelated with the distance of hand movements and was different across subjects who learned the same sequence. We then split a learned sequence into three segments, while preserving or destroying the clusters in the learned sequence, and shuffled the segments. The performance on the shuffled sequence was more accurate and quicker when the clusters in the original sequence were preserved than when they were destroyed. The results suggest that each cluster is processed as a single memory unit, a chunk, and is necessary for efficient sequence processing. A learned visuomotor sequence is hierarchically represented as chunks that contain several elementary movements. We also found that the temporal patterns of sequence performance transferred from the nondominant to dominant hand, but not vice versa. This may suggest a role of the dominant hemisphere in storage of learned chunks. Together with our previous unit-recording and imaging studies that used the same learning paradigm, we predict specific roles of the dominant parietal area, basal ganglia, and presupplementary motor area in the chunking.

Keywords Motor learning · Sequence · Chunk

K. Sakai (✉)
Wellcome Department of Cognitive Neurology,
Institute of Neurology, 12 Queen Square, London,
WC1N 3BG, UK
e-mail: ksakai@fil.ion.ucl.ac.uk
Tel.: +44-20-78337484
Fax: +44-20-78131420

K. Kitaguchi
Department of Education, School of Letters,
Mukogawa Women's University,
6–8 Ikebirakicho, 663–8558 Nishinomihya, Hyogo, Japan

O. Hikosaka
Laboratory of Sensorimotor Research, National Eye Institute,
National Institute of Health, Rm 2A50, Bldg 49,
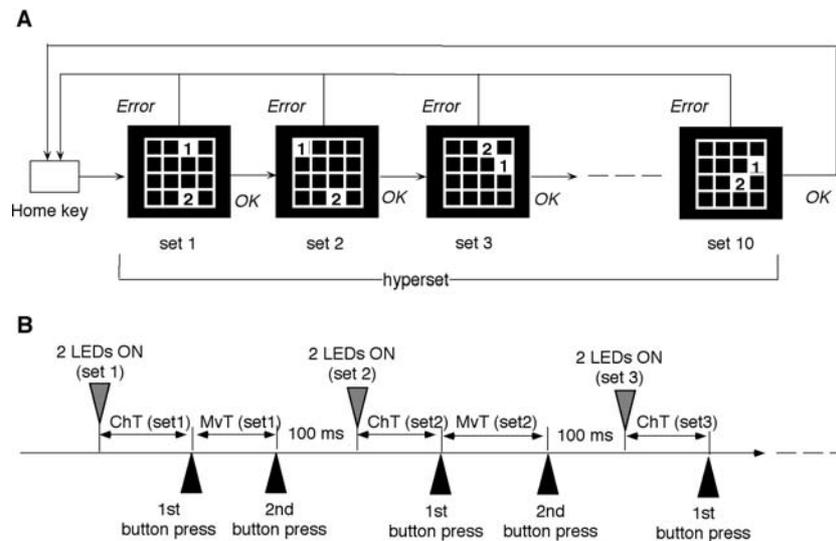49 Convent Drive, Bethesda, MD, 20892–4435, USA

Introduction

A motor sequence is performed with specific patterns of timing. These temporal patterns are thought to reflect the organization in which the motor sequence is represented. For example, Rosenbaum et al. (1983) found that a sequence of finger movements is performed as a group of subsequences that are separated by long time gaps and increased number of errors. A similar cluster pattern of motor sequence performance has been consistently observed in explicit learning of visuomotor sequence (Restle and Burnside 1972; Povel and Collard 1982), implicit learning of visuomotor sequence (serial reaction time task; Nissen and Bullemer 1987; Stadler 1989, 1993; Cohen et al. 1990; Keele and Jennings 1992; Curran and Keele 1993; Koch and Hoffmann 2000), and speech production (Gordon and Meyer 1984). The findings are taken to suggest a chunk representation for a motor sequence: a motor sequence is hierarchically organized with chunks of subsequences.

However, in all of those studies, the observed chunk patterns correspond to the specific structural patterns of the sequence: chunks have been identified according to the positions at which there is a change in the pattern of movements (repetition, inversion, and transposition; Koch and Hoffmann 2000), a unique pattern of transitions between movements (Cohen et al. 1990), or a temporal delay in the response–stimulus interval (Stadler 1993).

Fig. 1. A The 2×10 task. Sixteen LED buttons were arranged in 4×4 matrices. Subjects had to sequentially press the two LED buttons that were illuminated simultaneously (*set*). The correct order of the button presses was predetermined by the computer, and the subjects had to find it by trial and error. Ten consecutive sets (*hyperset*) were presented in a fixed order. If the subjects pressed a button in the wrong order at any set, or completed the hyperset, they started the same hyperset from the first set (*set 1*). **B** Behavioral measures: choice times (*ChT*) and movement times (*MvT*). ChT was expressed by the time between the illumination of the two LED buttons and the pressing of the first button. MvT was expressed by the time between the first and second button-pressing in a set

Thus, the chunk patterns are externally specified by physical parameters that determine the sequence structure. The same chunk patterns are observed across subjects who perform the same motor sequence. This raises the question as to whether the chunk patterns merely reflect the change in the movement parameters such as the distance of finger movements or switch between the hands. What remains unsolved is whether chunk patterns reflect sequence organization at a representational level, independent of physical parameters. Strong evidence for it would be that different subjects form different chunk patterns even if they perform the same sequence. Also, it remains open whether chunks have operational significance, that is, whether chunks are necessary for efficient performance of a visuomotor sequence. Finally it remains open whether chunk patterns are effector-specific or not.

To answer these questions, we have used a behavioral paradigm called a "2×10 task," an explicit visuomotor sequence learning through trial and error (Fig. 1A; Hikosaka et al. 1995). We investigated the temporal patterns in the performance of 2×10 visuomotor sequences in three experiments. In experiment 1, subjects performed a hyperset for 1 or 4 sessions. We show an emergence of chunk patterns that are independent of physical parameters. We also show that different subjects chunk the same hyperset differently. In experiment 2, subjects learned a hyperset and then performed a shuffled hyperset, that is, a new hyperset that was created by shuffling the 10 sets in the learned hyperset. We show that the shuffled hyperset is performed more efficiently when the chunks in the original hyperset are preserved than when they are distorted. A similar shuffling procedure has been used by Koch and Hoffmann (2000). In experiment 3, subjects learned a hyperset with one hand and then performed the hyperset with the other hand. We show that the chunk patterns in the learned hyperset can be transferred only from the nondominant to the dominant hand.

We have extensively used this 2×10 task in behavioral (Hikosaka et al. 1995, 2002b; Rand et al. 1998, 2000), single-unit recording (Nakamura et al. 1998), local inactivation (Miyachi et al. 1997; Lu et al. 1998; Nakamura et al. 1999), and imaging studies (Hikosaka et al. 1996; Sakai et al. 1998, 1999). We have also proposed a neural network model based on this learning paradigm (Nakahara et al. 2001). Together with the anatomical evidence obtained in these studies of our own, we make specific predictions on the neural mechanisms of chunking of a visuomotor sequence.

Materials and methods

Subjects

We conducted experiments on 26 (experiment 1), 6 (experiment 2), and 8 (experiment 3) normal human subjects. All gave informed consents prior to the study.

Task paradigm

We used a sequential button-press task, the 2×10 task (called "2×5 task" for monkeys; Hikosaka et al. 1995). The subjects were seated in a chair and faced with a button-pressing device. The device was a black box on which were placed 16 red light-emitting diode (LED)

buttons arranged in a 4×4 matrix (Fig. 1A). Below the 16 buttons was another button called a "home key." The task started upon pressing a home key for 1 s, after which a pair of buttons randomly chosen from the 16 buttons was illuminated simultaneously (a "set"). The subjects were asked to press the illuminated buttons one by one using the index finger on the instructed side of the hand. The correct order of the button presses was predetermined and the subjects had to find it through trial and error. When the order of pressing a pair of buttons was correct, the LED was turned off upon each button pressing, and another pair of buttons (next set) was illuminated. Again, the subjects had to press the two buttons in the correct order. By so doing, the subjects had to complete a sequence of 10 sets with different combinations of two buttons (a "hyperset"). When the subjects pressed the buttons in the wrong order in any set, the whole 16 buttons were illuminated with a beep and the subjects had to start a new trial from the first set of the same hyperset. After completion of a hyperset, the subjects started another trial of the same hyperset (the same 10 sets presented in the same order) until they completed the hyperset successfully for 20 consecutive trials (a "session"). Here a "trial" indicates the performance of a hyperset until subjects made errors at some point of the hyperset or completed the hyperset. Instruction was given to the subjects to perform the task as quickly as and as accurately as possible.

As described, the task is an explicit learning of a visuomotor sequence through trial-and-error processes. Importantly, the pressing of the second button in a set was followed by the illumination of LEDs for the next set, with only a small gap of 100 ms (Fig. 1B), and subjects were allowed to make anticipatory movements before the next pair of LEDs was turned on. Also of note is that we can create an almost infinite number of different hypersets by changing the combinations of buttons. Therefore, we can test the same subject many times using different hypersets.

Before starting the experimental sessions, we explained the procedure of the 2×10 task to the subjects and asked them to go through a 5-min practice session. The hyperset used in the practice session was different from the ones used in the experimental sessions. We also made sure that the hypersets used in the practice and experiment did not share the same set. For all the hypersets used, a particular set did not appear more than once in a hyperset regardless of the order of pressing the two buttons, that is there is no repetition, inversion and transposition in a hyperset. Also there was no temporal delay in the response–stimulus interval at any point in the hyperset.

Behavioral measures

We analyzed the temporal patterns of performance of a hyperset using two behavioral measures, choice time (ChT) and movement time (MvT; Fig. 1B). ChT was expressed by the time between the illumination of the two LED buttons and the pressing of the first button. Because the subjects were allowed to start moving their fingers before illumination of the LEDs, ChTs could be shorter than ordinary choice reaction times, that is, less than 100 ms. MvT was expressed by the time between the first and the second button-pressing in a set. As shown in Fig. 1B, we introduced a 100-ms gap between the pressing of the second button in a set and presentation of the next set. This gap did not hamper the smooth performance of button-pressing, because subjects were allowed to make anticipatory movements and 100 ms was short enough for the next pair of buttons to illuminate before the subjects pressed those buttons. Since ChT reflects the time taken to select the correct button based on memory, it reflects cognitive components. By contrast, MvT reflects purely motor components, because when the first button is correctly pressed the next button is automatically determined, and there is no need of selection or memory retrieval.

For statistical analysis, ChTs and MvTs were logarithmically transformed in order to achieve normal distributions of the data. The skewness of the raw data of the ChTs and MvTs for the 20 subjects in experiment 1 were 1.196 and 0.361, respectively, indicating that the tail of the distribution extended toward the high reaction time values. After the transformation, the skewness of the log(ChT) and log(MvT) was close to zero (−0.144 and −0.121, respectively), confirming the proximity to the normal distribution of the data.

Experiment 1: Emergence of chunks

We examined whether a hyperset was processed as several clusters of sets separated by long processing times. Where there were significantly larger processing times between set-clusters, we called these set-clusters "chunks." We used the term "chunk patterns" to describe the patterns in which the 10 sets were organized into chunks, e.g., set (1-2-3)-(4-5)-(6-7-8-9-10). We also examined whether the chunk patterns were unique to each subject. We then examined whether the observed chunk patterns were independent of the distance of finger movements.

Twenty subjects (2 left-handed; aged 20–26 years; 16 men, 4 women) learned a hyperset for 1 session (20 successful trials) using the dominant hand. Among them, 8 subjects performed a same hyperset in order to assess the intersubject difference in the performance. The other 12 subjects performed a hyperset different from each other. In addition, 6 subjects (aged 21–34 years, 5 men, 1 women, all right-handed) learned one hyperset for 4 sessions, taking a 3-min rest between sessions.

First, across-set difference in the number of errors, ChTs, and MvTs was analyzed. We conducted analysis of variance (ANOVA) with factor of Set on the number of errors, log(ChT), and log(MvT) for each set. The specific test was whether these behavioral measures differed significantly between sets. As a post-hoc test, we used Newman-Keuls test to account for the multiple comparisons across the 10 sets.

Second, we analyzed the intersubject difference in the across-set patterns of ChTs and MvTs for the 8 subjects who performed the same hyperset. We performed a two-factor ANOVA with factors of Set and Subject on log(ChT) and log(MvT). The key point of the analysis was whether there were significant Set × Subject interactions, which would indicate a significant across-subject difference in the ChT and MvT patterns. We also tested whether the log(ChT) patterns were significantly correlated across the 8 subjects.

Third, we analyzed the correlation of ChTs and MvTs with the distance of finger movements by calculating the Pearson's product-moment correlation coefficient (r). For ChTs, we measured the distance between the second button in the previous set and the first button in the present set, called "ChT-distance." For MvTs, we measured the distance between the first and the second buttons in a given set, called "MvT-distance." We also analyzed the correlation of ChTs and MvTs with the number of errors committed at the corresponding set.

Experiment 2: Operational significance of chunks

We examined the performance of a shuffled hyperset to test whether the chunks were essential for efficient sequence performance. Six subjects (all right-handed, aged 20–34 years, four men, two women) participated in the experiment. First, they learned a hyperset for 3 sessions (60 trials), with a 3-min rest in between. Then, the sequence of 10 sets in the learned hyperset was split into three segments, which were then shuffled to create a new hyperset. On session 4, the subjects performed the shuffled hyperset. Each subject was tested on two conditions, "between" and "within," using different learned hypersets. The between condition preserved the set clusters in the original hyperset. The learned hyperset was split before the sets with the longest and the second longest ChTs among the 9 sets (from set 2 to 10). In other words, the hyperset was split between the set-clusters separated by long ChTs. In contrast, the within condition destroyed the set clusters. The learned hyperset was split before the sets with the shortest and the second shortest ChTs. In other words, the hyperset was split within the set-clusters. Subjects had been informed that the hyperset on session 4

was the shuffled one, but did not know whether it was between or within. Three subjects were tested on "between" first and then on "within", while the other three subjects were tested on "within" first and then on "between"; thus the order of conditions was counterbalanced across subjects.

We first examined the difference in the effects of the two shuffling procedures (between and within) on the number of errors and the performance time. We performed two-factor ANOVA with factors of Condition (between and within) and Session (pre- and post-shuffle sessions; that is, sessions 3 and 4) on the number of errors and the mean performance time for a trial of a hyperset. The key point of the analysis was whether there were significant Condition × Session interactions, which would indicate a difference in the effect of the shuffling "between" and "within" conditions. If the set-clusters in the learned hyperset had operational significance, a shuffled sequence would be performed more accurately and more quickly when the original set-clusters were preserved in a shuffled hyperset (between condition) than when they were destroyed (within condition).

We then examined the difference in the effects of the two shuffling procedures (between and within) on the temporal patterns to perform a hyperset. We compared the degree of changes in the across-set patterns of ChTs and MvTs between the two conditions. To do this, the 10 sets for the shuffled hyperset were rearranged in the same order as in the original hyperset to enable the comparison of ChTs and MvTs corresponding to the identical sets between the pre- and post-shuffle sessions. To quantify the changes in the across-set ChT patterns, we regarded the across-set patterns of log(ChT) from set 1 to 10 as a 10-dimensional vector and calculated an angle formed by the two vectors for pre and post. The ChT angle thus calculated reflects the difference in the across-set ChT patterns between pre- (session 3) and post-shuffle sessions (session 4). A large ChT angle would indicate a large change in the across-set ChT patterns. Similarly, we calculated MvT angles from log(MvT) for sessions 3 and 4. The ChT angles and MvT angles for the six subjects were compared "between condition" and "within condition" using a paired *t*-test.

Experiment 3: Intermanual transfer

We examined whether chunks were independent of hands to perform the hyperset. Eight subjects participated in the experiment (aged 21–34 years; five men, three women). Six of them were strongly right handed (mean lateralization quotient of 89; Oldfield 1971), whereas the other two were left handed (lateralization quotient of −64 and −56, respectively). Each subject was tested on four conditions: dominant hand to dominant hand, D→D; nondominant hand to nondominant hand, ND→ND; dominant hand to nondominant hand, D→ND; and nondominant hand to dominant hand, ND→D. Different hypersets were used for the four conditions. In D→D condition, the subjects performed a hyperset for 4 sessions using the dominant hand. In ND→ND condition, they performed a hyperset for 4 sessions using the nondominant hand. In D→ND condition, the subjects used the dominant hand for the first 3 sessions and used the nondominant hand on session 4. In ND→D condition, they used the nondominant hand for the first 3 sessions, and used the dominant hand on session 4. For each subject the hand transfer conditions were tested after the non-hand transfer conditions, that is D→ND was tested immediately after D→D, and ND→D was tested immediately after ND→ND. The order of testing D→ND and ND→D conditions was counterbalanced across subjects, that is, four of the subjects were tested on D→ND first and then on ND→D, while the other four were tested on the reverse order.

First, we analyzed the effect of hand-switch on the temporal patterns of the performance at a subject level. We analyzed the effect of hand-switch on the temporal patterns of the performance at a subject level. We performed three-factor ANOVA with factors of Set (set 1–10), Session (pre and post transfer sessions, that is, session 3 and 4), and Condition (D→D, ND→ND, D→ND, and ND→D) on the log(ChT) for each subject. The analysis was focused on whether the size of interactions between Set and Session differed significantly across the Condition, that is, whether there were significant Set × Session × Condition interactions. We also performed, for each subject and for each hand-transfer condition, two-factor ANOVA with factors of Set and Session on the log(ChT). The key point of the analysis is under which condition there were significant Set by Session interactions, which would indicate significant differences in the across-set ChTs patterns between the pre- and post-hand transfer sessions.

Second, we analyzed the effect of hand-switch at a group level. As in experiment 2, we calculated the ChT angles and MvT angles between pre- and post-transfer sessions for each of the eight subjects. Large angles indicate large changes in the across-set temporal patterns due to hand-switch. To examine the difference in the ChT angle and MvT angles across conditions, we performed ANOVA on the dataset of eight subjects with a factor of Condition (D→D, ND→ND, D→ND, and ND→D).

Results

Experiment 1: Emergence of chunks

We examined the temporal patterns with which subjects performed a visuomotor sequence to see whether and how the subjects developed chunk patterns. Figure 2B, C, and D show the learning performance of one subject on a new hyperset shown in Fig. 2A. The subject learned to perform the hyperset for four consecutive sessions. In Fig. 2B are plotted the numbers of completed sets (abscissa) against trial numbers (ordinate, from top to bottom). At the beginning of trials, the subject made errors at the earlier sets in a hyperset. As the trials progressed, the number of completed sets in a trial gradually increased and, at trial 16, the subject successfully completed a hyperset (10 sets). The subject eventually became able to perform the hyperset with few errors. As seen from the figure, the number of completed sets often terminated at set 5, indicating that he made errors frequently at set 6.

Figure 2C and D show the temporal patterns of the performance of the hyperset. Broken lines in Fig. 2C show the cumulative sums of the ChTs up to 1–10 sets, which decreased gradually as trials progressed. Uneven spacing between the broken lines indicates across-set difference in ChTs. As trials progressed, the broken lines became clustered into three groups: set 1–3, set 4-5, and set 6–10. This suggests that the subject performed the hyperset in three clusters. ANOVA has shown significant main effect of Set on ChTs, suggesting across-set difference in ChTs on session 4 ($F_{9, 190}=19.4$, $P<0.01$). Post-hoc testing has shown that the ChTs for set 1, set 4, and set 6 were significantly larger than the ChTs for the other sets ($P<0.01$). In contrast, the cumulative sums of the MvTs (Fig. 2D) remained relatively unchanged and were less variable across the 10 sets.

Such clustering of sequence performance was consistently observed in the 6 subjects who performed a hyperset for 4 sessions. For all the subjects, ANOVA has shown significant across-set difference in ChTs on session 4 ($P<0.01$), and post-hoc testing has shown that several sets had significantly larger ChTs than the other sets ($P<0.01$). The size of the set-cluster varied from 1 to

Fig. 2A-D Emergence of chunks for one subject (*experiment 1*). **A** The hyperset used. **B-D** Learning-related changes (*from top to bottom*) in the number of completed sets (**B**) and in the cumulative sums of ChTs (**C**) and MvTs (**D**). The subject performed the hyperset (**A**) for four sessions, their borders indicated by the *thick horizontal lines* on each panel. **C**, **D** *Broken lines* indicate the cumulative sums of ChTs and MvTs, up to set 1, set 2, and set 10, respectively, *from the left to right*; the spacing between two adjacent broken lines corresponds to ChTs (**C**) and MvTs (**D**) for each set. In **C** and **D** are shown only the trials in which subjects successfully completed a hyperset, whereas in **B** are shown all trials. In **C**, three chunks separated by long ChTs (set 1–3, set 4–5, and set 6–10) became evident as sessions progressed (*from top to bottom*)
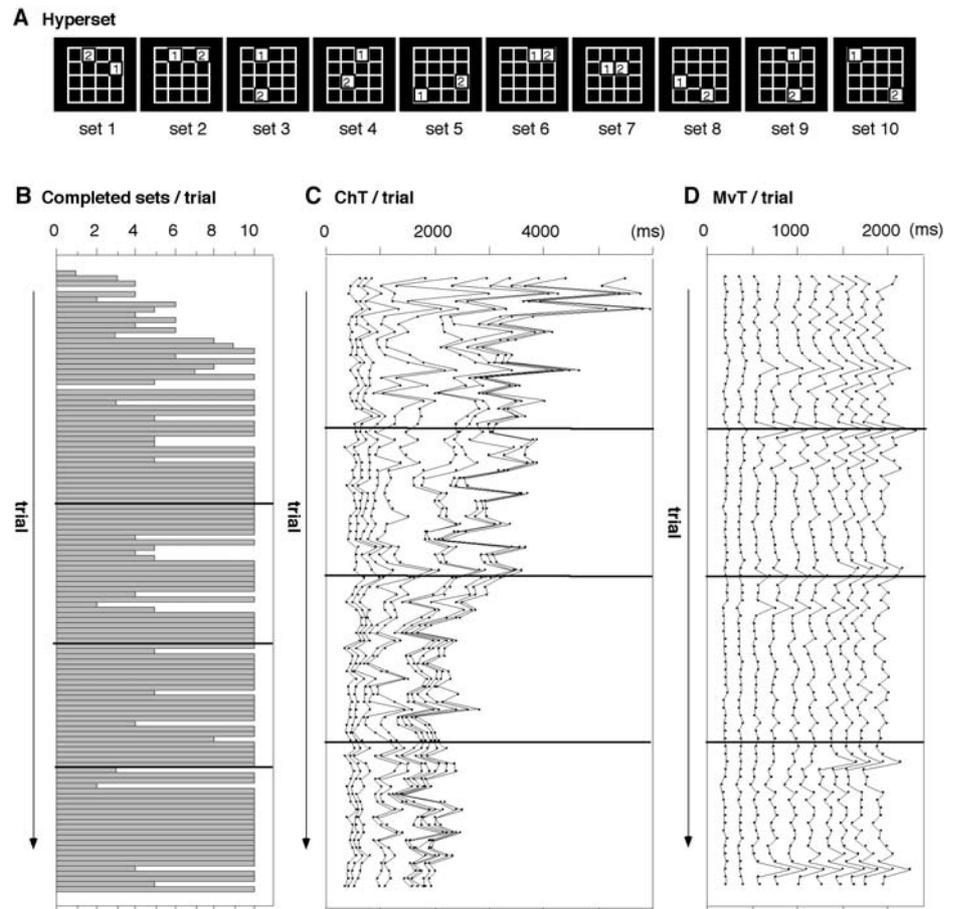
5 sets (mean 2.5 sets). For 20 subjects who performed a hyperset for 1 session, ANOVA showed significant across-set difference in ChTs even on session 1 ($P<0.01$). For 8 of them, post-hoc testing showed that several sets had significantly larger ChTs than the other sets ($P<0.01$). Thus, a hyperset was performed as several set clusters separated by long ChTs and this tendency became clearer as subjects repeated the hyperset.

We then examined the intersubject difference in the temporal patterns to perform a hyperset. Figure 3 shows the performance of 3 subjects who learned a new hyperset for one session. Although they performed the same hyperset (shown in Fig. 3A), the subjects made errors at different sets (Fig. 3B). The across-set pattern of ChTs was also different among the subjects (Fig. 3C), suggesting formation of cluster patterns that were different among the subjects. Subject Y.Y. chunked the hyperset as 1-(2,3)-(4,5)-6-7-8-(9,10), subject H.I. chunked the same hyperset as (1,2)-3-(4,5)-(6,7)-8-(9,10), and subject TS chunked the hyperset as (1,2,3)-4-(5,6,7)-8-(9,10). In contrast, the across-set MvT patterns were similar across the subjects (Fig. 3D). A two-factor ANOVA with factors of Set and Subject has shown that, for 8 subjects who performed the same hyperset, the Set × Subject interaction was significant for log(ChT; $F_{63, 1520}=4.65$, $P<0.01$), indicating across-subject difference in the ChT patterns. Thus subjects chunked the same sequence in different

manners. By contrast, the interaction was not significant for log(MvT; $F_{63, 1520}=1.15$, $P>0.1$), indicating a similar MvT patterns across subjects.

We further performed correlational analysis on the across-set patterns of log(ChT) for the 8 subjects who performed the same hyperset. Of the 28 possible pairs of the 8 subjects, in only one pair of subjects was there a significant correlation in the pattern of log(ChT; $P<0.05$ corrected). No other pair showed significant correlation even with less stringent threshold of $P<0.05$ uncorrected. Thus most of the subjects performed the hyperset with temporal patterns unique to their own even though they performed the same hyperset. By contrast, for all pairs of the 8 subjects, there was a significant correlation in the pattern of log(MvT; $P<0.05$ corrected). Thus the MvT patterns were determined by the hyperset itself.

Consistent with the idea, we found that the across-set distributions of log(MvT; left three graphs in Fig. 3D) were similar to the distribution of the distances between the first and second buttons (MvT-distance: right-most graph in Fig. 3D; also see Fig. 3A). For the 20 subjects who performed a hyperset for 1 session, MvTs were significantly correlated with the MvT-distance ($F_{1, 198}=57.7$, $P<0.01$, $r=0.48$).This suggests that the MvT was determined largely by the distance of finger movement. In contrast, the distributions of log(ChT; left three graphs in Fig. 3C) were dissimilar to the distribution of the
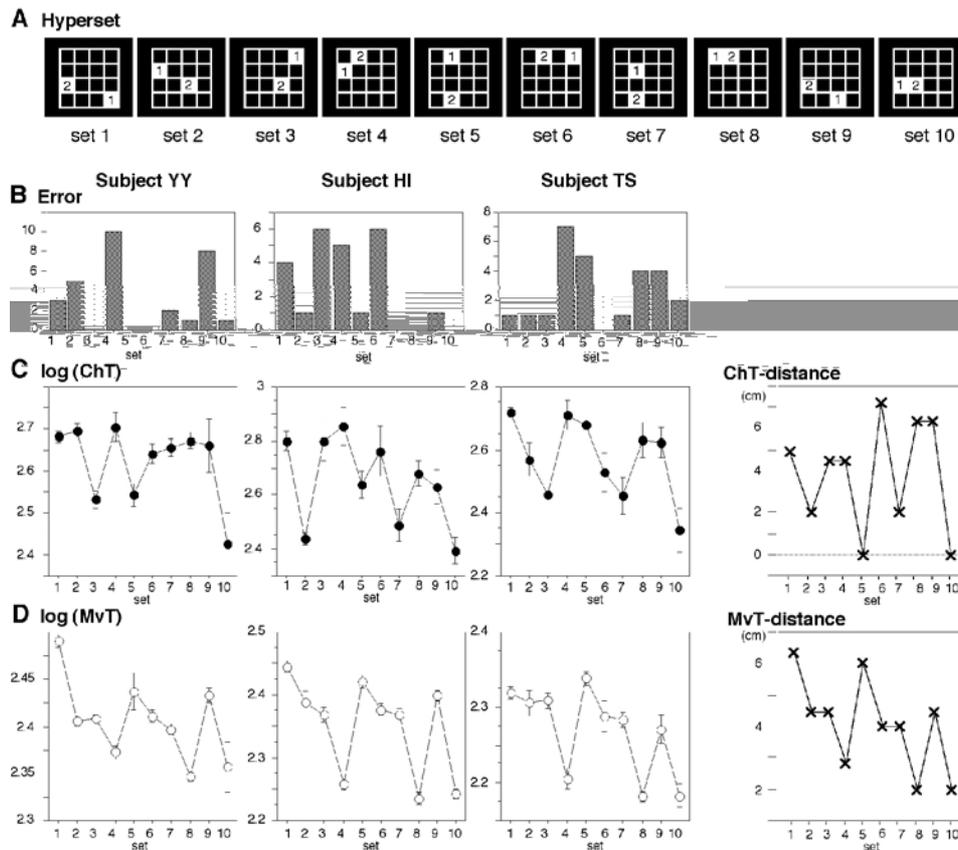
Fig. 3A-D Different chunk patterns for three subjects who learned the same hyperset for one session (*experiment 1*). **A** The hyperset used. **B** Across-set distribution of number of errors in a session. **C, D** Means and standard errors of log(ChTs) (**C**) and log(MvTs) (**D**). The data for **C** and **D** are based only on the trials in which subjects successfully completed a hyperset. On the *right-most* panels of **C** and **D** are shown the distances between the first button in the set (*n*) and the second button in the set (*n*−1;), and distances between the

first button and the second button in the set (*n*; *MvT-distance*). The ChT-distance for set 1 was measured from the home key to the first button in set 1 (see Fig. 1). Note that, in set 5, the ChT-distance was 0 cm, because the second button in set 4 and the first button in set 5 were identical (see **A**). Similarly for set 10. The ChT patterns differed across subjects and also differed from the ChT-distance pattern, whereas the MvT patterns were similar across subjects and were also similar to MvT-distance pattern

movement distances (ChT-distance: right-most graph in Fig. 3C). For the dataset of 20 subjects, ChT was not significantly correlated with the ChT-distance ($F_{1, 146}$= 1.05, $P>0.1$). Instead, log(ChT) was correlated with the number of errors committed at the corresponding sets ($F_{1, 198}$=15.8, $P<0.01$, r=0.47). Indeed, Fig. 3B, C shows that longer ChTs were observed for those sets at which subjects committed large number of errors even though only the trials in which subjects made correct button press were analyzed. By contrast, log(MvT) was not correlated with the number of errors ($F_{1, 198}$=1.26, $P>0.1$).

We also performed the same correlation analysis for the 6 subjects who performed a hyperset for 4 sessions. As in the learning for one session, in the advanced stage of learning (session 4) there was no significant correlation between the ChTs and ChT-distance ($F_{1, 46}$=0.005, $P>0.1$), suggesting that the chunk patterns were independent of physical constraints even after extensive learning. The MvTs in session 4, on the other hand, were correlated with the MvT-distance ($F_{1, 58}$=23.9, $P<0.01$, r=0.54), as in the performance in session 1. Interestingly, the ChT

patterns in session 4 were significantly correlated with the number of errors in session 1 ($F_{1, 58}$=19.8, $P<0.01$, r=0.42), though there were no or few errors in session 4.

Although the hyperset used in the practice session was completely different from the ones used in the experimental session, there might be the possibility that the structure of the hyperset used in the practice session affected the performance in the experimental session. We have excluded this possibility by performing correlational analysis between the performance in practice and performance in experiment. For all the subjects, there was no significant correlation in the across-set patterns of log(ChT) between the practice session and experimental session ($P>0.1$). Thus the ChT patterns or chunk patterns were specific to the subject and also specific to the hyperset. In addition, 5 of the 8 subjects were trained with the same hyperset in the practice and were tested with the same hyperset in the experiment, but the chunk patterns in the experimental session differed significantly across these subjects.

Fig. 4A, B Results of a split-and-shuffle experiment for one subject (*experiment 2*). The subject learned a hyperset for 3 consecutive sessions, and then performed a shuffled hyperset that was created by splitting the original hyperset. This was done for two different hypersets (**A**, **B**). The first hyperset was split before the two sets with long ChTs (between condition; **A**); the second hyperset was split before the two sets with short ChTs (within condition; **B**). The same format as in Fig. 2, except that the border between sessions is indicated only before the shuffling (**thick horizontal lines** in **A, B**). For both hypersets, there were three clusters of sets (*chunks*) on session 3 (set 1–2, set 3–7, and set 8–10 for **A**; and set 1–2, set 3–4, and set 5–10 for **B**). In session 4 of the between condition (**A**), the 10 sets were split between the clusters, and the first and last clusters (sets 1–2 and sets 8–10) were swapped. Thus the set-clusters in session 3 were preserved in session 4. In the within condition (**B**), the 10 sets were split within the second and third clusters, and the first and last segments (sets 1–3 and sets 6–10) were swapped. Thus the set-clusters in session 3 were destroyed in session 4
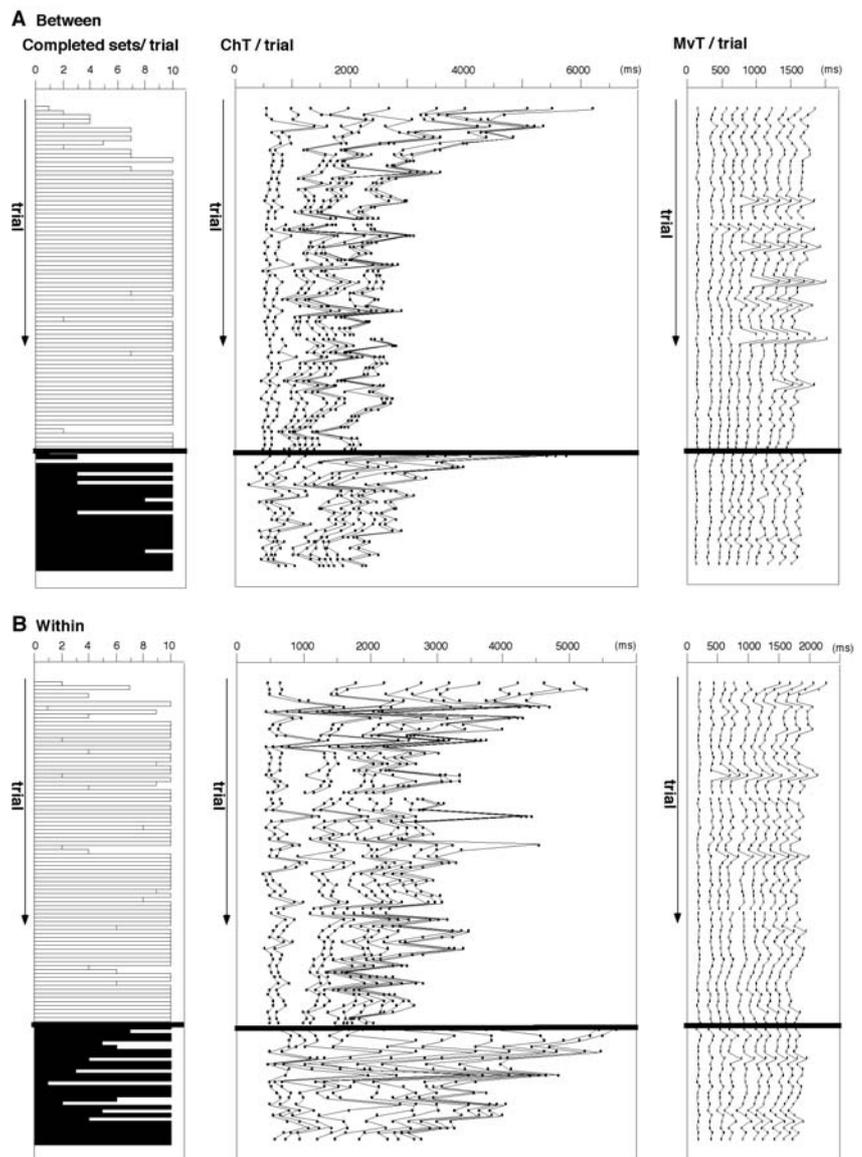
In sum, a hyperset was performed as several clusters of sets as learning progressed. The clusters were separated from each other by long time gaps (ChTs), which were associated with large number of errors at the beginning of learning. The across-set ChT patterns, an index of chunk patterns, differed from the patterns of the distance of finger movements. The patterns also differed across subjects even when they performed the same hyperset. Thus the chunking occurred independently of sequence structure, suggesting sequence organization at a representational level

Experiment 2: Operational significance of chunks

In experiment 1, we have shown that a learned hyperset was performed as several set-clusters, chunks, which were separated by long ChTs. However, the meaning of the chunks, i.e., their operational significance, was still unclear. To answer this question, we examined the performance on a shuffled hyperset. The learned hypersets were shuffled while the clusters were preserved (between condition) or were destroyed (within condition). The subject learned a hyperset for three sessions and then performed a shuffled hyperset on session 4. Each subject was tested twice, in the between condition (top) and in the within condition (bottom); an example is shown in Fig. 4. In the between condition, the hyperset was cut at the two longest ChTs, that is, to cut between the three set-clusters, set (1, 2), set (3, 4, 5, 6, 7), and set (8, 9, 10; Fig. 4). The first and last clusters were then swapped to create a new hyperset. In the within condition, the hyperset was cut at the two shortest ChTs, that is, to cut within the set-clusters. Thus created three segments, set (1, 2, 3), set (4, 5), and set (6, 7, 8, 9, 10), invalidated the set-clusters on session 3, set (1, 2), set (3, 4), and set (5, 6, 7, 8, 9, 10;
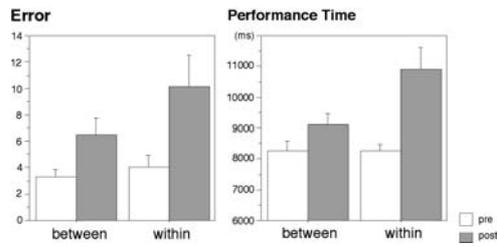
Fig. 5 Effect of shuffling on the number of errors (*left*) and mean performance time for a trial of a hyperset (*right*; experiment 2). Mean and standard error for the 6 subjects on pre- (*open bars*) and post- (*filled bars*) shuffling sessions are shown for between and within conditions

Fig. 4). The first and last segments were then swapped to create a new hyperset. In this way, the positions of the split were determined based on the performance of each of the 6 subjects tested.

The shuffling profoundly affected the subjects' performance. The number of errors and ChTs increased immediately after shuffling (thick horizontal lines in the graph), while MvTs showed little changes. The number of errors and ChTs then began to decrease as the subject repeated the shuffled hyperset, but the rate of decrease was smaller in the within condition than in the between condition. The results for all six subjects were consistent with this observation. Figure 5 shows the mean number of errors and performance time for the 6 subjects before (pre) and after (post) the shuffling. A two-factor ANOVA has shown a significant Condition (between and within) × Session (pre and post) interaction on the number of errors ($F_{1, 5}$=8.62, $P<0.05$) and performance time ($F_{1, 5}$=12.0, $P<0.05$). The increase in the number of errors and performance time after shuffling was significantly larger in within than in between (errors, 154% vs 55% increase; performance time, 32% vs 10% increase for within and between, respectively; $P<0.05$).

The shuffling also affected the temporal patterns to perform a hyperset, as shown in Fig. 6. For the data shown in Fig. 4, the across-set patterns of log(ChT) and log(MvT) were compared between pre- (session 3, thin line) and post-shuffling (session 4, thick line). Note that the ChTs and MvTs for the shuffled hypersets (thick line) have been rearranged in the order of the original hypersets to enable the comparison of ChTs and MvTs for the identical sets. The log(ChTs) were similar between the pre- and post-shuffling sessions in the between condition (Fig. 6A), but differed considerably in the within condition (Fig. 6B). As for log(MvTs), the difference was small between the two sessions for both within and between conditions.

We then quantified the difference in the across-set log(ChT) and log(MvT) patterns by calculating the ChT angles and MvT angles (Fig. 6C). The 10 sets for the shuffled hyperset were rearranged in the same order as in the original hyperset to enable the comparison of ChTs and MvTs corresponding to the identical sets between the pre- and post-shuffle sessions. We regarded the across-set
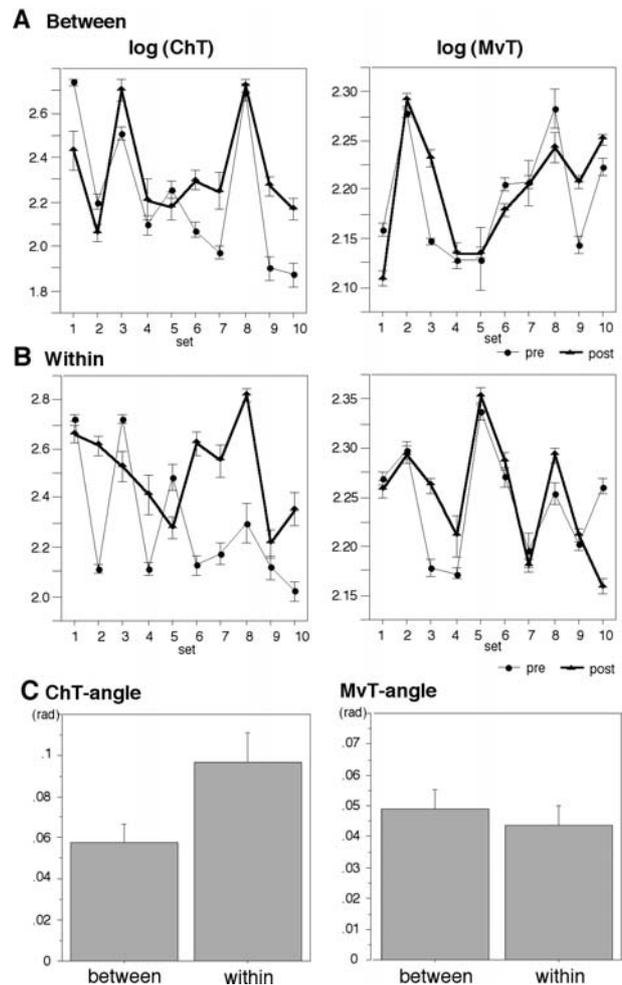

Fig. 6A–C Effects of the shuffling on the temporal pattern of performance in between condition (**A**) and within condition (**B**; experiment 2). Shown are the across-set patterns of log(ChTs; left) and log(MvTs; *right*) for one subject in pre- (*filled circle with thin lines*) and post- (*filled triangle with thick lines*) shuffling sessions (session 3 and 4). Note that the post-shuffling data have been rearranged according to the original order of the 10 sets. **C** The differences in ChT-angles (*left*) and MvT-angles (*right*) between the pre- and post-shuffling sessions. These angles reflect the difference in the across-set patterns of ChTs and MvTs between the pre- and post-shuffling sessions. The mean and standard error across the 6 subjects are shown

patterns of log(ChT) from set 1 to 10 as a 10-dimensional vector and calculated the angle formed by the two vectors for pre and post. A large ChT angle would indicate a large change in the across-set ChT patterns. Across the 6 subjects, the mean ChT angle was significantly larger in the within condition than in the between condition ($F_{1,5}$= 19.8, $P<0.05$), whereas the mean MvT angles were not significantly different between the two conditions ($F_{1,5}$= 0.67, $P>0.1$). Thus the changes in the temporal patterns were greater when chunks were destroyed than when chunks were preserved

In sum, a shuffled hyperset was performed more accurately and quickly when chunks in the original

hyperset were preserved than when they were destroyed. The temporal patterns to perform the sequence changed less when chunks were preserved. This suggests that each chunk was necessary for accurate and quick performance of a sequence and that each chunk was processed independently of its ordinal position within a sequence. A chunk works as a single memory unit.

Experiment 3: Inter-manual transfer

We examined whether the across-set ChT pattern, an index of chunk pattern, was specific to the hand used to learn the sequence. The subjects learned a hyperset for 3 sessions using one hand, and then performed the same hyperset using the same hand or the other hand. Each subject was tested on 4 conditions: dominant hand throughout the 4 sessions (D→D), nondominant through-out the 4 sessions (ND→ND), dominant then nondominant (D→ND), nondominant then dominant (ND→D). Figure 7A shows the data for a single subject. As shown, only in D→ND condition was there a large difference in the across-set log(ChT) patterns for the pre and post transfer sessions (session 3 and 4). The patterns were similar in D→D, ND→ND, and ND→D conditions. In contrast, the across-set log(MvT) patterns were consider-ably different in ND→D and D→ND conditions. The patterns were similar in D→D and ND→ND conditions.

This observation was confirmed statistically. On that subject shown in Fig. 7A, we performed three-way ANOVA with factors of Set (set 1–10), Session (pre and post transfer sessions), and Condition (D→D, ND→ND, D→ND, and ND→D) on the log(ChT)s. We found significant Set × Session × Condition interactions ($F_{27,\ 513}$=20.6, $P<0.01$). The size of Set × Session interactions was significantly larger in the D→ND condition than in other three conditions ($P<0.01$). To examine the nature of interactions in detail, we further performed a two-factor ANOVA with factors of Set and Session separately for each condition. The Set × Session interaction was significant only for D→ND condition, that is, when the hand was switched from the dominant to the nondominant hand ($F_{9,\ 171}$=5.02, $P<0.01$). In this D→ND condition, the sets having long ChTs changed by switching the hand; on the pre-transfer session (thin line), the ChTs for set 1, set 5, and set 8 were longer than those for the other sets (i.e., chunks: set 1–4, set 5–7, set 8–10; $P<0.05$); whereas, on the post-transfer session (thick line), the ChT for set 1, set 4, and set 6 were longer than those for the other sets (i.e., chunks: set 1–3, set 4–5, set 6–10; $P<0.05$). In contrast, the Set × Session interaction was not significant for D→D, ND→ND, and ND→D conditions, that is, when the same hand was used or when the hand was switched from the nondominant to the dominant hand ($F_{9,\ 171}$=0.81, 0.72, and 0.92, respectively; $P>0.1$ for all conditions). The sets having long ChTs did not change in these conditions.

The same finding was consistently observed for all the eight subjects tested. The right-handed subjects (n=6)
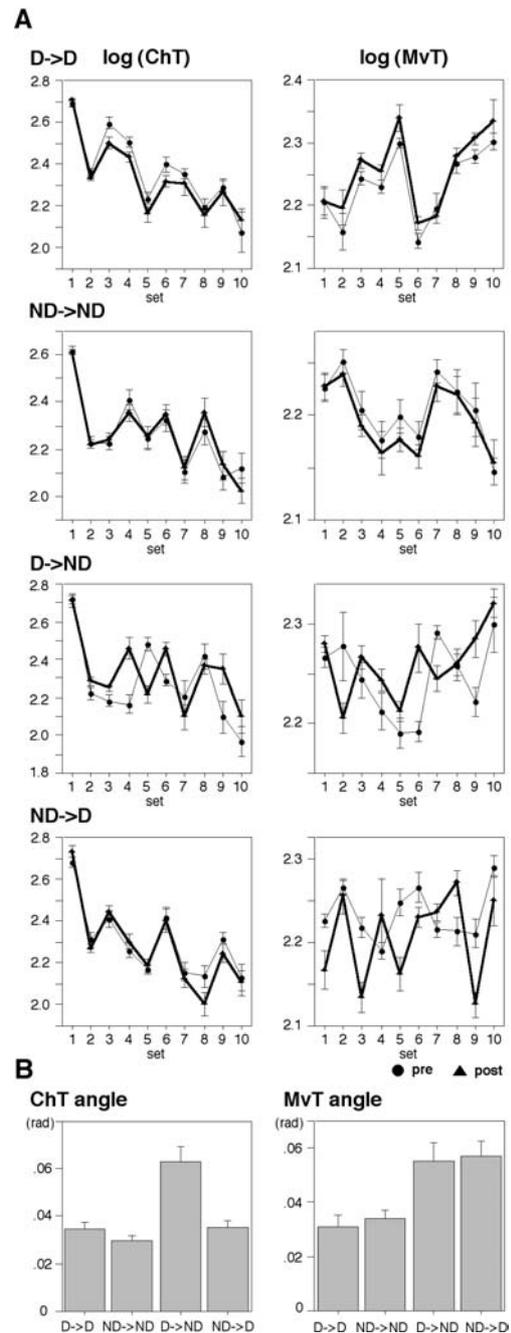

Fig. 7A, B Differential effects of intermanual transfer in one subject (experiment 3). For four conditions (D→D, ND→ND, D→ND, ND→D), the across-set patterns of log(ChT; *left*) and log(MvT; *right*) are shown for the pre- (*filled circle and thin lines*) and post- (*filled triangle and thick lines*) transfer sessions (session 3 and 4). In the D→D condition, the subject used the dominant hand throughout the 4 sessions; in the ND→ND condition, he used the nondominant hand throughout the 4 sessions. In the D→ND condition, the subject switched the hand from the dominant to the nondominant side at session 4; in the ND→D condition, he switched the hand from the nondominant to the dominant side. At the *bottom* are shown the ChT-angles (*left*) and MvT-angles (*right*) between the pre- and post-transfer sessions

showed significant changes in the ChT patterns when the hand was switched from the right to the left hand, whereas the left-handed subjects ($n=2$) showed significant change when the hand was switched from the left to the right hand. As for the log(MvT) patterns, the Set × Session interaction was significant for ND→D and D→ND conditions ($P<0.01$). It was not significant for D→D and ND→ND conditions ($P>0.1$).

The finding was also supported by the "angle" analysis based on the data for all eight subjects (Fig. 7B). Both the ChT angles and MvT angles, indices of the changes in the across-set patterns of ChTs and MvTs, were significantly different across the four conditions ($F_{3, 21}=18.8$, $P<0.05$). Post-hoc analysis showed that the mean ChT angle for D→ND condition was significantly larger than any other conditions ($P<0.05$), and that the angles for the other 3 conditions were not significantly different from each other ($P>0.1$; Fig. 7B, left). In contrast, the mean MvT angle was significantly larger in D->ND and ND->D conditions than D→D and ND→ND conditions ($P<0.05$; Fig. 7B right).

In sum, across-set ChT patterns (temporal patterns that reflect chunks) can only be transferred from the nondominant to the dominant hand, but not from the dominant to the nondominant hand. In contrast, MvT patterns (temporal patterns that reflect movement distances) changed significantly in both directions of hand transfer.

Discussion

The idea of chunk was originally proposed by Miller (1956), who defined chunk as a memory symbol with which several memory items can be treated as a single processing unit. Here we have shown that a visuomotor sequence was learned as chunks of several elementary movements. In previous studies on chunking of motor sequences (Restle and Burnside 1972; Povel and Collard 1982; Nissen and Bullemer 1987; Stadler 1989, 1993; Cohen et al. 1990; Keele and Jennings 1992; Curran and Keele 1993; Koch and Hoffmann 2000), the chunking pattern was externally imposed by generating a sequence with repetition, inversion, and transposition of elements or insertion of temporal delay between elements. We made no attempt to impose such a structure, and yet a chunking pattern emerged. And the chunking pattern was different between different subjects even when they performed the same sequence. This arbitrary and spontaneous nature seems to represent a key feature of motor habit: Motor habit develops with no obvious reason, varies among individual subjects, and often looks bizarre. More importantly, we showed that chunking makes processing of a visuomotor sequence more efficient.

Learning a visuomotor sequence

Our subjects learned the correct order of pressing 2×10 buttons that were indicated by the illumination of LEDs.

It is possible to learn the sequence by using nonmotor strategies. For example, subjects could verbally code the correct order of buttons. However we think this unlikely for the following reasons. First, all the subjects reported that they did not use the verbal coding strategy. Secondly, it is quite disadvantageous to put a verbal label on each of the 16 buttons and remember the sequence of 20 button presses based on the verbal codes. It is unlikely that, in performing a hyperset, decoding of the verbal labels to the correct movements could have occurred so quickly as shown in our data, sometimes with less than 100 ms of response time. Thirdly, our previous imaging studies using the same behavioral paradigm did not show activity in the language-related areas such as Broca's area or Wernicke's area (Sakai et al. 1998, 1999). It is also unlikely that the subjects learned the hyperset solely based on the visuospatial patterns of illuminated buttons. We have evidence that the learning also occurs in motor domain. In the present study, we found worsening of performance after changing the hand to perform the same hyperset even though the visuospatial patterns of the stimuli remained the same (experiment 3). Rand et al. (1998, 2000) have used the same behavioral paradigm on monkeys, and have shown that changing the hands affected the speed of performance for both early and late stages of learning. Furthermore, Bapi et al. (2000) have conducted a transfer experiment using a similar paradigm on human subjects and found that the response time was significantly shorter when the finger movement was the same compared with when the spatial sequence was the same. These results suggest that a hyperset was learned as an effector-specific motor sequence. In previous papers, we argued that a hyperset is learned based on spatial coordinates and motor coordinates and that the two mechanisms operate in parallel with significant interactions in between (Hikosaka et al. 1999a, 2002a; Nakahara et al. 2001).

Another possibility is that a hyperset is learned solely as associations between stimuli (a pair of buttons) and the correct response for each set. However, our data indicated that learning occurred for the sequence across sets. We found that shuffling of sets significantly affected the performance of a hyperset (experiment 2), indicating that a hyperset was learned as a sequence, not merely as a collection of stimulus-response associations. The same finding has been obtained in studies using monkeys (Rand et al. 1998, 2000). Thus it is the visuomotor sequence learning that has been examined in the present study.

Another important point is whether the hyperset was learned implicitly or explicitly. There can be 2^{10} ways of pressing buttons in a hyperset (2 ways of pressing the two buttons in a set × 10 sets) and yet our subjects learned a hyperset within 20 trials. At recognition test the day after the learning, the subjects were fully aware that they performed the hyperset (Hikosaka et al. 2002b), suggesting that a hyperset is learned explicitly. However, we think that implicit learning also occurs in parallel. When subjects were tested on recognition after 16 months of the learning, they were not aware of the hyperset at all, and

yet there was a significant effect of previous learning as indicated by the shortening of performance time (Hikosaka et al. 2002b). The present study is not aimed at segregating the explicit and implicit components of learning. It remains open whether the chunking of a visuomotor sequence reflects either of the two components.

Chunk representation for a motor sequence

We have shown that a motor sequence was performed as several clusters of sets, which were separated from each other by long time gaps. Moreover, we have shown that such set clusters are formed independently of the distance of finger movement (experiment 1). We call these set-clusters chunks. The chunk patterns differed across subjects even when they performed the same hyperset, supporting the view that the chunks are independent of sequence structures. Of note is that the subject-specific chunk patterns were evident in the patterns of ChTs, the time taken to select the first button in a set. The patterns of MvTs, that is the time taken to press the second button in a set, were determined by the distance of finger movements, and all the subjects showed the same pattern of MvTs as long as they performed the same hyperset. The chunking occurred between sets, not within a set. Since the chunking occurs as a result of delays in the selection/memory retrieval process (ChTs), the chunk patterns identified in the present study may reflect sequence organization in memory representation. In other words, the chunk patterns reflect how the hyperset is coded. As discussed, the coding can be neither verbal nor spatial. The hyperset is coded as an integrated product of spatial and motor sequences.

The fact that the long processing gap between chunks corresponds to the set where there was an increased number of errors might suggest that the chunk pattern is merely a by-product of making errors. However, this is not the case. The chunk patterns were observed even when there was no error in the performance of a hyperset. Much stronger evidence is obtained in experiment 2, showing the functional significance of chunks. The results in experiment 2 suggest that chunks work as functional units that altogether comprise a higher-order sequence and that the long ChTs do not merely reflect the delayed processing due to errors but processing gaps between functional, memory units. Our subjects were able to perform a shuffled sequence more accurately and more quickly when chunks in a learned sequence were preserved than when they were destroyed. The result suggests that a chunk is a necessary component for efficient processing of a sequence. Important here is that the chunk patterns were preserved even after their order was shuffled. Suppose a subject learned a sequence in three chunks, (1, 2, 3)-(4, 5)-(6, 7, 8, 9, 10), and then is asked to perform a shuffled sequence, 6-7-8-9-10-4-5-1-2-3, the subject performed it with the three chunks, (6, 7, 8, 9, 10)-(4, 5)-(1, 2, 3), but not with any of the other possible chunk patterns. The results suggest that each chunk is treated as a single memory unit, and a long sequence can be learned by linking chunks. As long as the individual chunks are preserved, the performance of a shuffled sequence requires only the rearrangement of the chunks.

A chunk representation can overcome the limitation of working memory capacity by forming a hierarchical structure for many memory items. Ericsson et al. (1980) reported a subject who became able to remember sequences of more than 100 digits by using chunks and hierarchically representing the sequences. The usage of chunks in serial behaviors has also been observed in rats (Fountain 1990; Dallal and Meck 1990; Macuda and Roberts 1995) and pigeons (Terrace 1987, 1991; Terrace and Chen 1991a, 1991b), and thus is considered to be the fundamental strategy to acquire long, complex sequences of actions. Our results thus suggest that the chunk formation is an efficient or crucial step toward development of a motor skill.

Spontaneous nature in the emergence of chunks

Importantly, the chunk pattern in the present study emerged spontaneously. There was no attempt to externally reinforce chunks, unlike the paradigm used by Koch and Hoffmann (2000) and Cohen et al. (1990), which introduced a change in the pattern of movements, a unique pattern of transitions between movements; the one used by Stadler (1993), which introduced a temporal delay between chunks; and the one used by Terrace (1991), which used different colors and patterns of stimuli for different chunks. The internally generated nature of chunks was evident in the fact that different subjects generated different chunks for the same sequence. One could then ask which subject showed the most optimal pattern of chunks. Learning-related changes in the chunk pattern may give some clues to this question. For example, the subject shown in Fig. 2 initially showed four chunks, (1, 2, 3), (4, 5), (6, 7), and (8, 9, 10) on session 2, but the last two chunks appeared to be fused to form a large third chunk, (6, 7, 8, 9, 10), on session 4. Thus, as learning progressed, subjects represented a sequence with fewer numbers of larger chunks. However, we have no indication that the chunk patterns of different subjects converged to the same pattern. It is possible that that the optimum is determined for each subject based on the history of performance unique to his/her own. Our results suggest that errors at the beginning of learning might be an important factor. Of note is that there still are significant chunk patterns in the error-less performance late in learning. Interruption of sequence performance caused by errors may have induced a breakpoint in sequence representation, which persisted even after extensive learning.

Such spontaneous nature in the emergence of chunk is similar to the formation of habit. Habit is formed through repetition of the same behavior without any intention.

Habit is unique to the individual. Once it is formed, habit remains unchanged for a long time. In our experiments, chunk became evident after extensive repetition of the same sequence. Our subjects had no obvious intention to form chunks. The chunk pattern was unique to each subject. The acquired chunk pattern remained unchanged even when a sequence was shuffled. It remains open whether a chunk pattern, like habit, is maintained as a long-term memory.

Asymmetric transfer of chunks

We also found that the temporal patterns with which subjects performed a motor sequence transferred from the nondominant to the dominant hand, but not from the dominant to the nondominant hand (experiment 3). Importantly, it is the patterns of response times that showed differential effects depending on the direction of hand transfer. The asymmetry cannot be accounted for by the difference in the skillfulness of the hands, because the significant change in the temporal patterns was observed only for the ChTs, not for the MvTs. The patterns of MvTs changed significantly after both ways of hand transfer. Since the MvT reflects purely the motor components of sequence performance, these changes may be due to the changes in muscular activity patterns in following the same spatial path with different hands/arms. By contrast, the ChTs are independent of movement distance and reflect selection/retrieval of correct sequence. Since the temporal patterns of ChTs reflect the patterns of chunk organization, the changes in the ChT patterns may suggest unidirectional transfer of chunks. The result indicates the changes in the patterns of ChTs rather than the changes in ChTs, thus it can not be accounted for by the difference in the skillfulness between the dominant and nondominant hands. In addition, since the MvT-angles, which reflect motor components, were similarly affected after either way of hand transfer, the effect on ChT-angles cannot be accounted for by the difference in the speed of performance. The result may further suggest an asymmetric role of cerebral hemispheres in the storage and/or expression of learned chunks. According to this hypothesis, when a sequence is learned by the dominant hand (e.g., right), the chunk pattern is represented in the dominant hemisphere (e.g., left), and the nondominant hand cannot gain access to the acquired chunk representation. In contrast, when a sequence is learned by the nondominant hand (e.g., left), there are at least two possibilities: (1) the chunk pattern is stored in the nondominant hemisphere (e.g., right), but even the dominant hemisphere can gain access to the chunk representation, or (2) the chunk pattern is stored in both hemispheres and therefore both hands can gain access to the chunk representation.

Alternatively, the asymmetric transfer could be explained by the difference in the grade of learning between the dominant and nondominant hands. At the time of hand transfer (after session 3), the dominant hand, compared with the nondominant hand, may have reached a more advanced stage of learning. As suggested by Rand et al. (1998, 2000), the initial learning may take place in both hemispheres, whereas, later in learning, only one hemisphere contralateral to the performing hand can control the sequence.

Neural substrates for chunking

We have used the same behavioral paradigm for monkey and human subjects, and have identified multiple brain areas that are involved in the learning. We demonstrated dissociation in the neural mechanisms between acquisition of a new motor sequence and execution of a learned motor sequence (Hikosaka et al. 1999a): the prefrontal cortex (Sakai et al. 1998), medial premotor cortex (Nakamura et al. 1998, 1999; Hikosaka et al. 1996; Sakai et al. 1998, 1999), and caudate nucleus (Miyachi et al. 1997) are involved in the acquisition of new sequences, whereas the parietal areas (Sakai et al. 1998), posterior putamen (Miyachi et al. 1997), and cerebellar dentate nucleus (Lu et al. 1998) are involved in storage of learned sequences. In the present study, we have shown that the learning of a motor sequence accompanies chunking of the sequence. Thus the areas active early and late in learning may, respectively, be involved in formation and storage of chunks. Several areas are of particular interest.

Our intermanual transfer experiment suggests a role of the dominant hemisphere in the storage of chunks. In the previous study, we have shown a learning-related increase of activity in the parietal areas on both sides (Sakai et al. 1998). The dominant side of the parietal areas may store the chunks acquired through learning. Lesions in the dominant parietal areas sometimes cause deficits in temporal structuring of a motor sequence, called apraxia (Harrington and Haaland 1991a, 1992). The deficits may be the result of disruptions of chunk patterns in learned daily behaviors. Rushworth et al. (1998) have proposed a role of the left hemisphere, including the parietal areas in selection of learned actions. In this sense, the dominant parietal areas may play roles in selecting appropriate chunks to perform a long motor sequence. Activation in the parietal areas has been shown to increase concomitant with an increase in sequence complexity (Sadato et al. 1996) and sequence length (Catalan et al. 1998). Since it is quite likely that an increase in sequence complexity and length is associated with an increase in the number of chunks to represent the sequence, the parietal activation may reflect the number of chunks to organize the sequence.

Another important structure may be the basal ganglia. Using the same behavioral paradigm on monkeys, Miyachi et al. (1997) have shown that the inactivation of the posterior putamen significantly affects the performance of well-learned sequences. Impairments in motor sequence performance have been demonstrated in parkinsonian patients (Harrington and Haaland 1991b). The deficits may result from failure to express chunked movements.

As suggested by Graybiel (1998), the basal ganglia may play a role in selection of the chunked representations that are stored in the cerebral cortex, possibly in the dominant parietal areas. It is also significant that, as discussed, a chunk has many properties similar to habit, and the basal ganglia play critical roles in formation of habit (Knowlton et al. 1996; Jog et al. 1999).

These arguments are, however, merely correlational. Direct evidence for neural chunking is specific patterns of neural activities that correspond to chunk patterns. We predict that such activities would be observed in the anterior part of the medial premotor area, called the presupplementary motor area (pre-SMA). Nakamura et al. (1998), using the same behavioral paradigm except that a sequence was comprised of 5 sets, have shown that neurons in the pre-SMA in monkeys are active at every set during the early stage of learning but later became active only at the first set of the learned sequence. In the early stage of learning, a chunk is thought to be of minimal size, 1 set, and five chunks comprise a sequence, whereas when the sequence is over-learned the 5 sets form a chunk and a sequence is represented as a single chunk. The learning-related changes in the pre-SMA activity may thus reflect the evolution of larger chunks to organize a sequence. A similar learning-related decrement of pre-SMA activation has also been observed in human subjects (Hikosaka et al. 1996; Sakai et al. 1998, 1999). Nakamura et al. (1998) have also found that when some of the sets in the learned sequence were replaced by new sets, the neurons in the pre-SMA showed increased discharge on these modified sets, as well as the unchanged sets. This manipulation has the same impact as the destruction of chunks in the present study: subjects had to relearn the entire sequence as a new one. Many pre-SMA neurons are active only at the first trial of a learned sequence (Nakamura et al. 1998). Such "first trial activity" was also observed by Shima et al. (1996). It was concluded that the pre-SMA may play roles in updating and retrieving a new motor sequence. In the case of a long motor sequence (10 sets, 20 moves) as in the present study, the entire sequence is comprised with several chunks, and it would be necessary to update and load the next chunks when going through a long sequence. A specific experiment to test this hypothesis is a test of phasic activation in the pre-SMA between chunks, but not within chunks. Recently, Kennerley et al. (2002) have trained human subjects with a 12-move motor sequence and demonstrated chunk patterns unique to each subject. They found that transcranial magnetic stimulation over the pre-SMA affected the performance of a learned motor sequence only when it was given between chunks, supporting a role of the pre-SMA in updating chunks.

To date, a large number of studies have investigated the neural representation of a learned motor sequence (for review, see Hikosaka et al. 1999b, 2002a). However, we know very little about the form in which a learned sequence is represented. The present study indicates that chunking is the critical step to represent a long motor sequence. Specific tests on the neural substrates of chunking will further clarify the sequence organization at a representational level.

References

Bapi RS, Doya K, Harner AM (2000) Evidence for effector independent and dependent representations and their differential time course of acquisition during motor sequence learning. Exp Brain Res 132:149–162

Catalan MJ, Honda M, Weeks RA, Cohen LG, Hallett M (1998) The functional neuroanatomy of simple and complex sequential finger movements: a PET study. Brain 121:253–264

Cohen A, Ivry RI, Keele SW (1990) Attention and structure in sequence learning. J Exp Psychol Learn Mem Cognit 16:17–30

Curran T, Keele SW (1993) Attentional and nonattentional forms of sequence learning. J Exp Psychol Learn Mem Cognit 19:189–202

Dallal NL, Meck WH (1990) Hierarchical structures: chunking by food type facilitates spatial memory. J Exp Psychol Anim Behav Process 16:69–84

Ericsson KA, Chase WG, Faloon S (1980) Acquisition of a memory skill. Science 208:1181–1182

Fountain SB (1990) Rule abstraction, item memory, and chunking in rat serial-pattern tracking. J Exp Psychol Anim Behav Process 16:96–105

Gordon PC, Meyer DE (1984) Perceptual-motor processing of phonetic features in speech. J Exp Psychol Hum Percept Perform 10:153–178

Graybiel AM (1998) The basal ganglia and chunking of action repertoires. Neurobiol Learn Mem 70:119–136

Harrington DL, Haaland KY (1991a) Hemispheric specialisation for motor sequencing: abnormalities in levels of programming. Neuropsychologia 29:147–163

Harrington DL, Haaland KY (1991b) Sequencing in Parkinson's disease. Abnormalities in programming and controlling movement. Brain 114:99–115

Harrington DL, Haaland KY (1992) Motor sequencing with left hemisphere damage: are some cognitive deficits specific to limb apraxia? Brain 115:857–874

Hikosaka O, Rand MK, Miyachi S, Miyashita K (1995) Learning of sequential movements in the monkey: process of learning and retention of memory. J Neurophysiol 74:1652–1661

Hikosaka O, Sakai K, Miyauchi S, Takino R, Sasaki Y, Putz B (1996) Activation of human presupplementary motor area in learning of sequential procedures: a functional MRI study. J Neurophysiol 76:617–621

Hikosaka O, Nakahara H, Rand MK, Sakai K, Lu X, Nakamura K, Miyachi S, Doya K (1999a) Parallel neural networks for learning sequential procedures. Trends Neurosci 22:464–471

Hikosaka O, Sakai K, Nakahara H, Lu X, Miyachi S, Nakamura K, Rand MK (1999b) Neural mechanisms for learning of sequential procedures. In: Gazzaniga MS (ed) The new cognitive neurosciences. MIT Press, Cambridge, MA, pp 553–572

Hikosaka O, Nakamura K, Sakai K, Nakahara H (2002a) Central mechanisms of motor skill learning. Curr Opin Neurobiol 12:217–222

Hikosaka O, Rand MK, Nakamura K, Miyachi S, Kitaguchi K, Sakai K, Lu X, Shimo Y (2002b) Long-term retention of motor skill in macaque monkeys and humans. Exp Brain Res 147:494–504

Jog MS, Kubota Y, Connolly CI, Hillegaart V, Graybiel AM (1999) Building neural representations of habits. Science 286:1745–1749

Keele SW, Jennings PJ (1992) Attention in the representation of sequence: experiment and theory. Hum Mov Sci 11:125–138

Kennerley SW, Sakai K, Rushworth MFS (2002) The pre-SMA and sequential organization of actions: a TMS study. Abstr Soc Neurosci 163.8

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399–1402

Koch I, Hoffmann J (2000) Patterns, chunks, and hierarchies in serial reaction-time tasks. Psychol Res 63:22–35

Lu X, Hikosaka O, Miyachi S (1998) Role of monkey cerebellar nuclei in skill for sequential movement. J Neurophysiol 79:2245–2254

Macuda T, Roberts WA (1995) Further evidence for hierarchical chunking in rat spatial memory. J Exp Psychol Anim Behav Process 21:20–32

Miller GA (1956) The magic number seven, plus or minus two: Some limits on our capacity for processing information. Psychol Rev 63:81–97

Miyachi S, Hikosaka O, Miyashita K, Karadi Z, Rand MK (1997) Differential roles of monkey striatum in learning of sequential hand movement. Exp Brain Res 115:1-5

Nakahara H, Doya K, Hikosaka O (2001) Parallel cortico-basal ganglia mechanisms for acquisition and execution of visuomotor sequences: a computational approach. J Cognit Neurosci 13:626–647

Nakamura K, Sakai K, Hikosaka O (1998) Neuronal activity in medial frontal cortex during learning of sequential procedures. J Neurophysiol 80:2671–2687

Nakamura K, Sakai K, Hikosaka O (1999) Effects of local inactivation of monkey medial frontal cortex in learning of sequential procedures. J Neurophysiol 82:1063–1068

Nissen MJ, Bullemer P (1987) Attentional requirements of learning: evidence from performance measures. Cognit Psychol 19:1-32

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113

Povel DJ, Collard R (1982) Structural factors in patterned finger tapping. Acta Psychologia 52:107–123

Rand MK, Hikosaka O, Miyachi S, Lu X, Miyashita K (1998) Characteristics of a long-term procedural skill in the monkey. Exp Brain Res 118:293–297

Rand MK, Hikosaka O, Miyachi S, Lu X, Nakamura K, Kitaguchi K, Shimo Y (2000) Characteristics of sequential movements during early learning period in monkeys. Exp Brain Res, 131:293–304

Restle F, Burnside BL (1972) Tracking of serial patterns. J Exp Psychol 95:299–307

Rushworth MFS, Nixon PD, Wade DT, Renowden S, Passingham RE (1998) The left hemisphere and the selection of learned actions. Neuropsychologia 36:11–24

Rosenbaum DA, Kenny SB, Derr MA (1983) Hierarchical control of rapid movement sequences. J Exp Psychol Hum Percept Perform 9:86–102

Sadato N, Campbell G, Ibanez V, Deiber M, Hallett M (1996) Complexity affects regional cerebral blood flow change during sequential finger movements. J Neurosci 16:2691–2700

Sakai K, Hikosaka O, Miyauchi S, Takino R, Sasaki Y, Putz B (1998) Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. J Neurosci 18:1827–1840

Sakai K, Hikosaka O, Miyauchi S, Sasaki Y, Fujimaki N, Putz B (1999) Presupplementary motor area activation during sequence learning reflects visuomotor association. J Neurosci 19: RC1–5

Shima K, Mushiake H, Saito N, Tanji J (1996) Role for cells in the presupplementary motor area in updating motor plans. Proc Natl Acad Sci USA 93:8694–8698

Stadler MA (1989) On learning complex procedural knowledge. J Exp Psychol Learn Mem Cognit 15:1061–1069

Stadler MA (1993) Implicit serial learning: questions inspired by Hebb (1961) Mem Cognit 21:819–827

Terrace HS (1987) Chunking by a pigeon in a serial learning task. Nature 325:149–151

Terrace HS (1991) Chunking during serial learning by a pigeon. I. Basic evidence. J Exp Psychol Anim Behav Process 17:81–93

Terrace HS, Chen S (1991a) Chunking during serial learning by a pigeon. II. Integrity of a chunk on a new list. J Exp Psychol Anim Behav Process 17:94–106

Terrace HS, Chen S (1991b) Chunking during serial learning by a pigeon. III. What are the necessary conditions for establishing a chunk? J Exp Psychol Anim Behav Process 17:107–118