A Temporal Signal Reveals Chunk Structure in the Writing of Word Phrases

Peter C-H. Cheng (p.c.h.cheng@sussex.ac.uk)
Hector Rojas-Anaya (h.rojas@sussex.ac.uk)
Representation and Cognition Research Group
Department of Informatics, University of Sussex
Brighton, BN1 9QH, UK

Abstract

Evidence is provided that the patterns of pauses that occur during the process of writing simple word phrases constitute a substantial temporal signal that reflects the structure of the chunks in working memory. This temporal signal is apparent in un-aggregated data for individual participants in single trials. Five and six word phrases, with word lengths of one to six letters were used. Pause durations were significantly longer between chunks (words) than for elements within a chunk (letters). Longer chunks, words with more letters, required significantly longer to process before they are written, and there is a trend for shorter pause durations for later chunks in the phrases. This provides a further demonstration that writing may be an effective approach to probe the structure of chunks in memory.

Keywords: Protocol analysis, writing, methodology, chunks, temporal signal, working memory, word phrases.

Introduction

The concept of chunks of information is fundamental to the understanding of the processing of information by the cognitive architecture. The notion is regularly invoked in theories of perception, memory, thinking and motor behaviour (e.g., Cowan, 2001; Miyake & Shah, 1999; Gobet & Simon, 1998; Chi, Glaser, & Farr, 1988; Rosenbaum, Hindorff & Munro, 1987). The range of tasks which can be coherently explained in terms of chunks is impressive (e.g., Chase & Simon, 1973; Goodnow & Levine, 1973; Reitman, 1976; Egan & Schwartz, 1979; Vincente, 1988, Cheng, McFadzean & Copeland, 2001). Despite the theoretical importance of chunks to Cognitive Science, standard methods to analyze chunks in behavioural output have limitations, because they are laborious to use (e.g., Ericsson & Simon, 1993) or because they rely upon inferences about chunk structures based on global task level performance measures, such reaction times or error rates.

This paper describes an experiment that is part of a program of research that is attempting to develop a method to identify the structure of chunks in working memory using the processes of writing and drawing – graphical protocol analysis. In previous work it has been shown that there is a distinct temporal signal that reveals whether written elements are constituents within a chunk or at the boundary between successive chunks (Cheng, McFadzean & Copeland, 2001; Cheng & Rojas-Anaya, 2005). Graphical protocol analysis appears to have several advantages over previous methods. It exploits modern graphics tablet technology, which is not only economical and relatively simple to use, but provides high sampling rates and spatial resolution that yield accurate and precise raw data in electronic form. Much of the initial extraction, analysis and coding of such digitally recorded writing and drawing actions can be done automatically with computer tools (although current tools are research prototypes). More importantly, however, graphical protocol analysis advocates the analysis of the production of continuous sequences of relatively long stimuli comprising multiple chunks. This permits relatively naturalistic drawing and writing tasks to be used even in an experimental context. Further, the density of data per trial can also be higher as each stimulus may contain multiple hierarchically structured chunks. In the case of the present experiment, phrases of five or six words, of up to six letters each, were used.

Our previous experiments using graphical protocol analysis revealed that there is a clear and substantial temporal signal that can be used to distinguish intra- and inter-chunk elements. This signal was found in the drawing of simple geometrical figures (Cheng, et al., 2001) and the writing of sequences of numbers (Cheng & Rojas-Anaya, 2005). The chunk structure of the stimuli was predefined and learned by the participants; induced in memory. The pause between the end of drawing/writing of one element and the beginning of the production of the next – the time the pen is lifted from the graphics tablet – is substantially shorter for elements within a chunk than for an element that begins a new chunk. For the geometric drawings the within and between chunk pause durations were approximately 410 and 620 ms, respectively, and for the written number sequences the pauses were approximately 280 and 440 ms, respectively. (The differences in durations between the tasks can in part be attributed to the greater distance the pen is moved between elements in the drawing task).

Graphical protocol analysis differs in a number of ways from previous work that uses handwriting to study chunk-based phenomena. Other approaches have typically used the response latency paradigm (e.g., Lochy, Pillon, Zesiger, & Seron, 2002; Hulstijn & van Galen, 1983), which focuses on a single chunk per trial with the manipulation of the size or complexity of the chunk between trials. Further, such studies have been particularly concerned with the nature of prepared motor program chunks, with reported latencies typically under 200 ms. In contrast our studies have dealt with the production of a succession of chunks from working memory. The pauses between chunks in the graphical protocol studies are at least double the latencies in the reaction times studies, Familiar Phrases:

F1. # ALL FOR ONE ONE FOR ALL

F2. # TO BE OR NOT TO BE

F3. # ONE SMALL STEP FOR A MAN

F4. # I SPY WITH MY LITTLE EYE

F5. # ALL YOU NEED IS LOVE

F6. # YOU ARE WHAT YOU EAT

Jumbled Phrases:

J1. # FOR NEED FOR SMALL SPY MAN

J2. # MY FOR WITH YOU YOU TO

J3. # EAT LITTLE NOT OR LOVE ONE

J4. # A BE TO WHAT STEP YOU

J5. # BE ALL ONE ALL EYE

J6. # ALL IS ONE I ARE

Fig. 1. Word phrases used in the experiment

which argues that different levels of phenomena are being address by the two approaches.

This paper continues the investigation of the scope and reliability of graphical protocol analysis by testing whether the temporal signal of chunks is apparent with simple word phrases. If the signal is another manifestation of the processing of chunks during writing, then it should be comparable in nature and strength to that found with written number sequences. (1) The durations of intra- and inter-chunk pauses should be comparable across the domains. (2) The durations of intra- and inter-chunk pauses should not only be significantly different but also substantially different. (3) In the number sequence experiment the pause duration before the beginning of a new chunk appeared to increase with the size of the chunk (Cheng & Rojas-Anaya, 2005). A similar pattern should be found with the word phrases.

As in the previous studies, the approach was first to induce a given chunk structure into working memory using stimuli with a predefined hierarchical structure and then to examine whether the patterns of pauses during production corresponded to that a priori structure. Fig. 1 shows the stimuli used in the experiment. It is assumed that: (a) a word will correspond to a chunk in memory (level 2 - L2); (b) its constituent letters will be its sub-chucks (L1); (c) the graphic elements, distinct written strokes, will be sub-sub-chunks of the word (L0). The first two assumptions are plausible because the phrases contain words that are not compounds (cf., 'evermore') and successive words cannot be put together as a single meaningful word (cf. 'man age').

Two sets of phrases, Fig. 1, were used as stimuli: six familiar and six jumbled (unfamiliar) phrases. The familiar phrases were obtained by pooling phrases from websites of common English sayings and having 10 native English speakers rate their familiarity. The six most familiar phrases that also met the above requirements were picked. The jumble phrases were created by randomly selecting words from the familiar phrases, without replacement, to make six phrases consisting of equivalent numbers of words. The jumbled phrases were included to test for a possible effect of familiarity of the word phrases.

Method

The methodology used was similar to that of Cheng & Rojas-Anaya (2005), so a brief account of the critical details is given here. The ten participants were postgraduate students and research staff at the University of Sussex. They wrote on a graphics tablet, Wacom Intuos$^{2®}$. Following familiarization with writing on the tablet and training on a set of dummy phrases, the participants wrote each of the twelve phrases alternating between familiar and jumbled phrases but otherwise in a random order. Each phrase was presented on a card and the experimenter checked the recall accuracy by asking each participant to recite the phrase until they were confident that they could write it in a continuous unhesitating manner. The phrase was then written in a horizontal row of squares, one letter per square, whilst simultaneous reciting the phrase again without seeing the card. A hash (#) was written at the beginning of each phrase to ensure that the writing process was well underway before the first letter was generated.

Specially designed software for drawing/writing analysis, TRACE, was used to record the writing actions, to extract the pen positions and times, and to analyze the duration of pauses between drawn elements (Cheng & Rojas-Anaya, 2004).

Results

Results will first be presented at the level of individual trials and then proceed towards more global analysis that covers data pooled across participant and phrases. Therefore, particular terms are introduced to help refer to the different levels of aggregation of the data. Data for an *individual-phrase* covers a single trial, consisting of one written phrase by one participant. Each participant produced twelve such individual-phrases (6 familiar and 6 jumbled), giving a total of 120 in the experiment. A *participant* set of data covers all of the 12 phrases written by one participant. There were ten such sets in the experiment. A *phrase* set of data covers all 10 participants writing the same phrase.

TRACE calculates the pauses between all of the elements. Each of the marks made were coded as being an element within a letter (L0), as a letter within a word (L1), or as the first letter of a word or chunk (L2). Median and non-parametric statistics were used for the analyses as data for chunking production behaviours is often skewed. L0 pauses were relatively rare and so are not considered here.

Patterns in individual-phrase data (single trials)

A screen snap shot from the TRACE graphical recording and analysis program, for the writing of the familiar phrase F2 by one participant, is shown in Fig. 2.

The small circles superimposed on each written letter indicate the beginning and end of the production of elements in those letters as the pen touches or leaves the paper. The

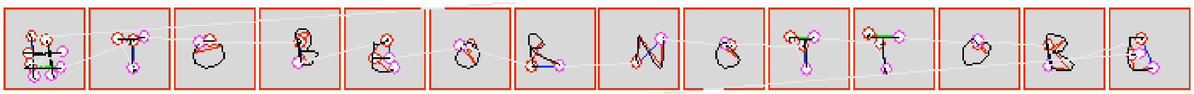

Fig. 2. A written individual-phrase (familiar F2, by DR), showing successive pauses.

lines between elements indicate transitions where the pen is off the paper. Note the two pairs of dots on the letters ‘T’ and ‘E’ as they were written in two parts. Hence, there will be a within letter (L0) pause associated with each of those letters. The distance between letters is comparable for letters with and between words.

Fig. 3 shows two graphs of the sequence of pause durations for the same participant writing phrases F2 and J3. The solid line gives the pause durations. The dashed line (arbitrary units) indicates the expected chunk level, whether it is an inter-chunk L2 data point (100 units), an intra-chunk L1 point (50 units), or a within letter L0 data point (zero units). The letters forming the sequence are shown along the dashed line (with each letter aligned to its respective data point). A sense of the closeness of the match between the expected chunk structure and the durations of the pauses can be judged by visually comparing the shape of the solid and dashed lines (not their absolute magnitudes). For both familiar and unfamiliar phrases, the pauses with the greatest magnitudes typically occur at the beginning of a chunk (new word), which suggests that the pattern of pauses is due to the chunk structure given by the words in the phrase.

The graphs shown in Fig. 3 have not been specially selected but are representative of all the individual-phrase graphs. It is noteworthy that the temporal signal reflecting the individual word structure of the phrases is apparent at this single trial level before any aggregation of the data.

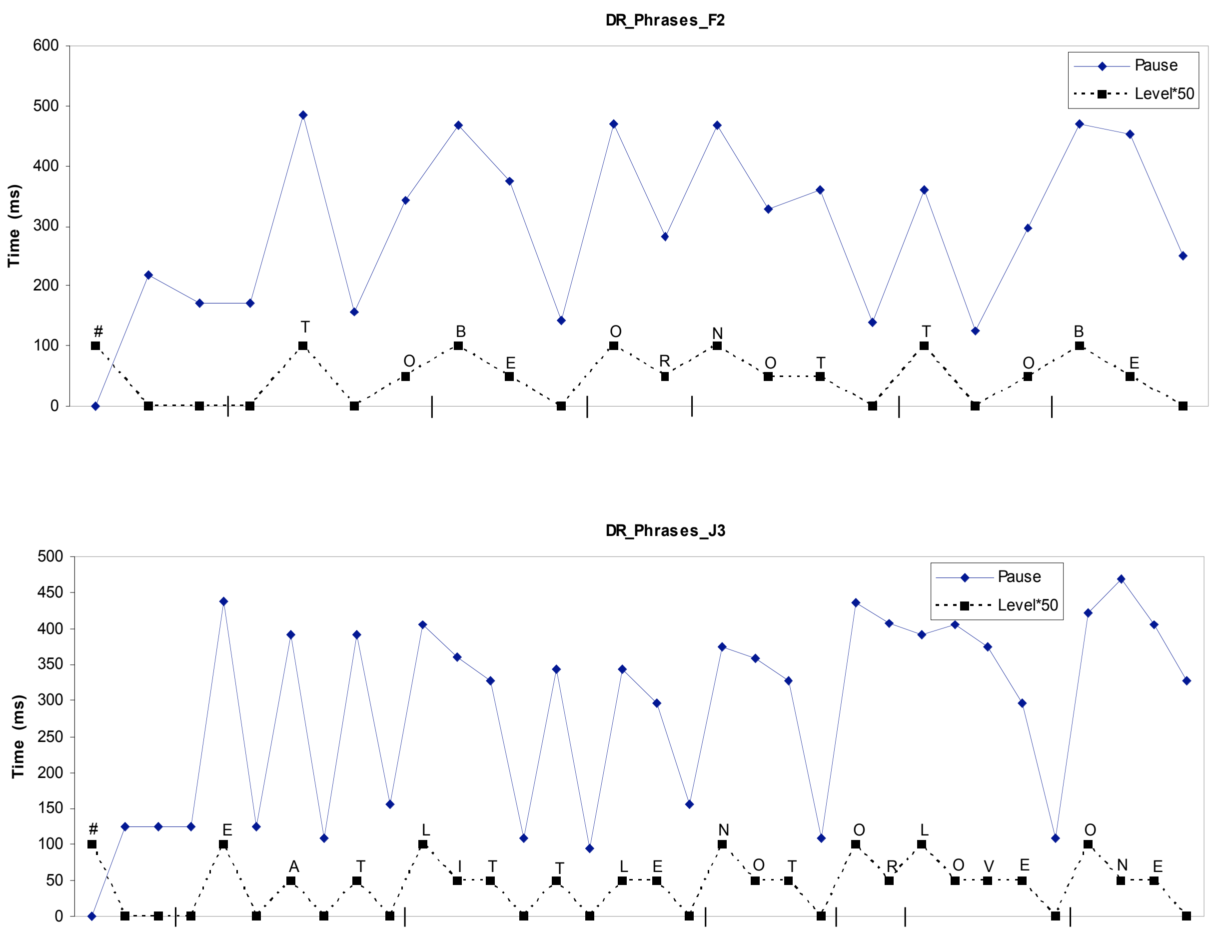

Figure 3: Successive pauses for one participant (DR), writing familiar phrase F2 (top) and jumbled phrase J3 (bottom).

Table 1. Pause durations for within (LI) and between chunk (L2) levels: medians for each participant

Familiar Phrases

Participant	DR	RG	NP	DL	II	MT	HR	RB	MJT	RK	Mean
Median of L1 medians	391	265	223	281	226	188	282	235	266	352	271
Median of L2 medians	445	516	282	387	383	359	383	469	324	398	394
Median of L2-L1 medians	62	266	59	129	156	165	78	258	59	31	126
L1—L2 significant difference											
Number of phrases with $p<.05$	3	6	3	5	3	6	4	5	3	1	3.9

Jumbled Phrases

Participant	DR	RG	NP	DL	II	MT	HR	RB	MJT	RK	Mean
Median of L1 medians	375	235	239	246	211	215	297	211	265	360	265
Median of L2 medians	411	515	328	438	304	325	407	485	332	387	393
Median of L2-L1 medians	31	289	90	180	90	109	102	270	74	20	125
L1—L2 significant difference											
Number of phrases with $p<.05$	1	6	2	5	4	3	5	6	4	1	3.7

Intra- and inter-chunk pause durations

For each individual-phrase, the medians of the L1 (within word/chunk) and L2 (between word/chunk) pauses were computed. Then participant medians were computed (i.e., the median of all the individual-phrase medians for each participant). These are shown in Table 1, along with the difference between the two levels (L2-L1). For every phrase across all the participants the median of the intra-chunk pause duration was less than the inter-chunk pause duration, for both familiarity conditions. To test whether the differences between L1 and L2 pauses were significant, Mann-Whitney U was for computed for each participant median. The number of phrases for each participant where the difference was significant at $p<.05$ is also shown Table 1. Of the 60 familiar individual-phrases, this difference was significant for 39 of the cases (65%). Of the 60 jumbled individual-phrases, this difference was significant for 37 of the cases (62%).

Overall, the magnitude of L2 pauses for the familiar phrases was 45% greater than the L1 pauses, and similarly 48% greater for jumbled phrases. The mean difference between the levels was 126 ms or 125 ms for the two conditions, respectively. Comparisons of the L1 and L2 participant medians (rows in Table 1) using one-tail t tests, shows that, for both the familiar and unfamiliar phrases, the difference was significant at the $p<.001$ level.

It is worth noting that although participant DR has a low L2-L1 (see Table 1), inspection of Fig. 3 shows that the pattern of pauses and the imposed chunk structure are comparable. The temporal signature of the chunks is more noticeable in the graphs of the other participants who have a larger difference between the L1 and L2 levels.

Contrary to expectation there is no difference between familiar and jumbled phrases in terms of their absolute pause values or the difference between L1 and L2. In the subsequent analyses the data for all the phrases are pooled across the two familiarity conditions.

Effect of chunk length and position in phrase

The words of all phrases were classified according to chunk length – 1 to 6 letters long; and according to their position in a phrase – the 1st to the 6th word in a phrase. The median of L2 pauses for words with the same chunk length and position in phrases are shown in Table 2a. The values are medians of all L2 values pooled across all phrases and all participants.

Table 2: Medians of L2 pause (ms) for words of particular lengths and position in phrases (top); number of words contributing to those values (bottom).

(a) L2 Median		Position in phrase					
		1	2	3	4	5	6
Chunk length	1	461	-	-	360	266	-
	2	422	360	344	422	337	368
	3	446	406	391	360	399	391
	4	-	438	406	422	414	-
	5	-	375	-	438	-	-
	6	-	462	-	-	437	-

(b) No. words		Position in phrase					
		1	2	3	4	5	6
Chunk length	1	2	-	-	1	1	-
	2	3	3	2	3	1	2
	3	7	6	5	6	6	6
	4	-	1	5	1	3	-
	5	-	1	-	1	-	-
	6	-	1	-	-	1	-

Table 3: Change of L2 medians (ms) with chunk length

		Positions in phrase					
		1	2	3	4	5	6
Chunk length transition	1-2	-39	-	-	+62	+71	-
	2-3	+24	+46	+47	-62	+62	+23
	3-4	-	+32	+15	+62	+15	-
	4-5	-	-63	-	+16	-	-
	5-6	-	+87	-	-	-	-

Table 4: Change of L2 medians (ms) with position in phrase

		Successive position in phrase				
		1-2	2-3	3-4	4-5	5-6
Chunk length	1	-	-	-	-94	-
	2	-62	-16	+78	-85	+31
	3	-40	-15	-31	+39	-8
	4	-	-32	+16	-8	-

For example, the cell at the coordinates (2, 3) in the Table has a median of 344 ms for both of the two-letter words that were in the 3rd position within all phrases, which includes 'OR' in F2, and 'TO' in J4. Empty cells in Table 2a and 2b means that no words of the given length and position occur among the stimuli.

On initial inspection, Table 2 shows a general trend of L2 pauses increasing with chunk length and decreasing with position in phrase. Table 3 quantifies the increase with chunk length by showing change in L2 values for words differing in length by one letter, for fixed positions in phrases. In Table 3 '1-2' denotes the transition from 1 to 2 letters, '2-3' from 2 to 3 letters, and so forth. Positive values indicate that, for a specific position in phrase, the L2 medians have increased, whereas negative values indicate a decrement in L2 medians. Table 3 shows a preponderance of increases in L2 values, indicated by the low proportion of negative changes; 3 out of 16. If it is assumed that it is equally likely that the values may increase or decrease (p=0.5), then by the binomial theorem the probability that at least 3 out of 16 values are decreasing is p=.01. Hence, an increase in the length of a word will result in an increase in the pause duration at the beginning of the word. The mean increase in Table 3 is 25 ms; in other words, adding an extra letter to a chunk will require an additional 25 ms to process the chunk before the word can begin to be written.

In a similar way, Table 4 examines the change in L2 medians of pause durations for successive positions of words in a phrase, with the chunk length kept constant. There are 14 cases where there are words of the same length in consecutive positions. Ten of those cases have decreasing pause durations: the remainder increasing. The average decrease is 16 ms per position. Assuming again that the chance of increasing or decreasing values are equal, then the binomial theorem probability of having up to 4 positive changes out of 14 is p=.09. Hence, although there appears to be a trend for a decreasing pause duration the later a word occurs in a phrase, the reliability of this finding must be treated with caution.

Discussion

In the context of simple word phrases, this experiment provides further evidence for the existence of a strong and robust temporal signal that reveals the structure of chunks in memory as pattern in the durations of pauses between written elements. In the majority of trials, with individuals writing a single phrase, the duration of inter-chunk (L2) pauses was significantly longer than intra-chunk (L1) pauses. The duration of L2 pauses was about 50% greater than the L1 pauses. Overall, this is consistent with our previous studies, in which the same temporal signal in writing and drawing has been shown to reveal chunk structure in memory (Cheng, et al., 2001, Cheng & Rojas-Anaya, 2005).

All together, this demonstrates that graphical protocol analysis has some potential as means to study chunk-based cognitive phenomena. The approach allows relatively long stimuli to be used, the phrases had five or six words and up 21 letters, which gives a good density of data per trial. The initial processing of data and the extraction of pause durations was done largely automatically by the TRACE tool. Although real words and familiar phrases were used, the judicious choice of words and phrases appears to have been successful in producing a consistent induced structure of chunks in the memory of the participants. A basis has now been established for exploring the use of the method, in reverse, to infer the structure of chunks from the patterns of pauses for words and phrases that are more complex.

The mean intra-chunk pauses were about 270 ms and the mean inter-chunk pauses were about 400 ms, for both the familiar and unfamiliar word phrases. These magnitudes are comparable to the within and between chunk pauses found in the previous experiment using number sequences (Cheng & Rojas-Anaya, 2005). Those L1 and L2 pauses were 280 and 440 ms, respectively. Further, the suggested increase in inter-chunk pause duration with larger chunks in the number sequences was also found with the word phrases, but this time the increase with each additional letter was significant. Overall, this argues that the writing of number sequences and word phrases is likely to share similar chunk processing mechanisms.

The magnitudes of the inter-chunk pause durations (400+ ms), contrasts with studies of writing that focus on programmed motor behaviour (e.g., Lochy, Pillon, Zesiger, & Seron, 2002; Hulstijn & van Galen, 1983). In those studies, the reaction time for the production of a chunk is no greater than 200 ms, which is also less than the present L1 intra-chunk pause. This supports the view that the current approach is concerned with different, or additional, process compared to those earlier studies. A major difference in the current approach is the successive production of multiple chunks, rather than a one-off reaction time. Hence, it is hypothesized that the longer durations may also incorporate processes that prepare chunks in working memory, prior to the programming of motor actions.

The lack of an effect of the jumbled versus familiar phrases is initially surprising, but in hindsight also theoretically explicable. Although familiarity often has an affect in many tasks, its absence here may be due to the specific experimental procedure that required participants to verbally rehearse a jumbled phrase. This may have made the words in the jumbled phrase as active in memory as the words in a familiar phrase, which sees to have required less repetition. The extra immediate processing of the jumbled phrase could compensate for its lack of presence in long-term memory. The lack of an effect is interesting, as it suggests, tentatively, that the temporal signal in this experiment largely reflects processing associated with working memory. This is an issue for further investigation.

Graphical protocol analysis relies fundamentally on the idea that the pause durations between written elements meaningfully reflect the structure of chunks in memory and is somehow due to the internal processing of the chunks. An alternative explanation is that the longer pauses are caused by the time needed for participants to physically move their hand, which may happen at the boundary between chunks rather than between elements within a chunk. This is an issue we are currently investigating using video recordings of the participants as they wrote the phrases. Preliminary analysis indicates that such movements are not strongly associated with transitions between words. Hence, the occurrence of physical movements is likely to be detrimental to the temporal chunk signals as a source of noise, rather than as a challenge to the underlying basis of the approach.

Acknowledgment

Our thanks go to the members of the Representation and Cognition Research Group for their assistance with this study.

References

Card, S. K, Moran, T. P., & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cowan, N. (2001). The Magical Number 4 in Short-term Memory: A Reconsideration of Mental Storage Capacity. *Behavioral and Brain Sciences* **24**(1): in press

Chase, W., & Simon, H.(1973). Perception in chess. *Cognitive Psychology* **4**,55-81.

Cheng, P. C.-H., McFadzean, J., & Copeland, L. (2001). Drawing out the temporal structure of induced perceptual chunks. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the Twenty Third Annual Conference of the Cognitive Science Society* (pp. 200-205). Mahwah, New Jersey: Lawrence Erbaum.

Cheng, P.C-H., & Rojas-Anaya, H. (2004). TRACE user guide (Unpublished Representational Systems Laboratory report).

Cheng, P. C. H., & Rojas-Anaya, H. (2003). Writing out a temporal signal of chunks: patterns of pauses reflect the induced structure of written number sequences. In B. G. Bara, L. Barsalou & M. Bucciarelli (Eds.), *Proceedings of the Twenty Seventh Annual Conference of the Cognitive Science Society* (pp. 424-429). Mahwah, NJ: Lawrence Erlbaum.

Chi, M. T. H., Glaser, R., & Farr, M. J. (Eds.). (1988). *The Nature of Expertise*. Hillsdale, N.J.: Lawrence Erlbaum.

Egan, D. E., and B. J. Schwartz (1979). Chunking in the recall of symbolic drawings. *Memory and Cognition*, **7**(2), 149-158.

Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Revised ed.). Cambridge, Mass: MIT Press.

Gobet, F. (1998). Expert Chess Memory: Revisiting the Chunking Hypothesis. *Memory*, **6**(3), 225-255.

Goodnow, J., & Levine, R. (1973). The Grammar in Action: Sequence and syntax in Children's Copying. *Cognitive Psychology*, **4**, 82-98.

Hulstijn, W., & van Galen, G. P. (1983). Programming in handwrting: reaction time and movement time as a function of sequence length. *Acta Psychologica*, **54**, 23-49.

Lochy, A., Pillon, A., Zesiger, P., & Seron, X. (2002). Verbal structure of numerals and digits handwriting: New evidence from kinematics. *Quarterly Journal of Experimental Psychology*, **55a**(1), 263–288.

Miyake, A., & Shah, P. (Eds.). (1999). *Models of working memory: Mechanisms of active maintenance and executive control*. New York: Cambridge University Press.

Reitman, J. (1976). Skilled perception in Go: Deducing Memory Structures from Inter-Response Times. *Cognitive Psychology*, **8**, 357-381.

Vicente, K. (1988). Adapting the memory recall paradigm to evaluate interfaces. *Acta Psychologica*, **69**, 249-278.