A Graphical Chunk Production Model:
Evaluation Using Graphical Protocol Analysis With Artificial Sentences

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

A model of the processes involved in the graphical production of chunks form memory is proposed. The primary processing constraints of the model are (a) serial processing (b) over a hierarchical structure of chunks (c) in a depth first manner. The model makes predictions about the relative durations of each individual pause that occurs between successive writing or drawing actions based on the particular chunk structure of each individual stimulus. An experiment is also described that tests the model using artificial sentences. The overall results of the experiment were consistent with main predictions of the model.

Keywords: chunks, model of chunk production, graphical protocol analysis, drawing, writing.

Introduction

A brief inspection of any text on Cognitive Science reveals that the perception, storage, retrieval and transformation of information by the cognitive architecture have been extensively studied but the transmission of information has been rather neglected. Much less is known about how chunks of information in the mind are processed for external production than is known about the other fundamental types of cognitive processes. For example, write “All for one, one for all” in capital letters. What processes are occurring during the ≈10 seconds during which the sentence is being transmitted from your mind to the paper? Do the processes occur in a largely serial fashion or do they partly overlap and even run substantially in parallel? How is the decomposition of the phrases, words, letters and graphical elements organised and are there measurable manifestations of this in observable behaviour?

It is known that temporal patterns in behaviour may reveal the structure of chunks in memory (e.g., McLean & Gregg, 1967; Egan & Schwartz, 1979; Chase & Simon, 1973). In particular, when a series of actions are executed the duration of the pause before a given action is typically taken to be indicative of the amount of processing required to produce the output, with longer pauses indicating boundaries between different chunks in memory. Particular thresholds have even been used to segment recorded behaviours into sub-sequences that are associated with particular chunks. However, a detailed processing account of how chunks are produced, particularly in the context of graphical production, is yet to be provided.

Research in human computation interaction and cognitive modelling has developed analysis methods and models of some forms interactive behaviour and skilled performance that can model sequences of perceptual, cognitive and motor actions (e.g., John & Kieras, 1996; Anderson, 1998). Recent developments include formal approaches to model the parallel execution of actions (Howes, Vera, Lewis & McCurdy, 2004). This work has focussed on keyboard tasks and direct manipulation interface environments and is it yet to be applied to graphical production. The nature of graphical production is an open research question and is a challenge because of the dominance of memory retrieval and internal cognitive processes.

Work in the area of motor behaviour has provided evidence that chunks are important in the programs that govern motor behaviour and that such programs appear to have a hierarchical structure (e.g., Rosenbaum, Hindorff & Munro, 1987). However, these studies use a response latency paradigm to focus on relatively simple chunks, with the processing of just one chunk per trial with the manipulation of the size or complexity of the chunk between trials these study, (e.g., Lochy, Pillon, Zesiger, & Seron, 2002; Hulstijn & van Galen, 1983). The effects of interest have been concerned with the immediate production of fully prepared motor program chunks, with response latencies under 200 ms. Obviously, much more is occurring during the production of the above saying that takes 50 times as long.

We have previously reported work on a research programme that has developed *Graphical Protocol Analysis* (GPA) as a method to identify the structure of chunks in an individual’s memory by analysing the processes of writing and drawing of complex stimulus over long(er) durations (≈10 s). GPA allows the extraction and interpretation of a strong and robust temporal chunking signal across a range of tasks, media, stimuli formats and levels of expertise, including: drawing simple geometric patterns using a pen or a computer mouse (Cheng, McFadzean & Copeland, 2001); recalling and writing number sequences (Cheng & Rojas-Anaya, 2005); writing familiar and unfamiliar phrases (Cheng & Rojas-Anaya, 2006); expert and novice copying of mathematical equations (Cheng & Rojas-Anaya, 2007). The potential benefits of GPA include: the use of modern, economical, simple to use graphics tablet technology; raw data that is rich (hi-frequency), accurate and precise; automatic initial extraction, analysis and coding of digital behaviour protocols by computer (although current tools are

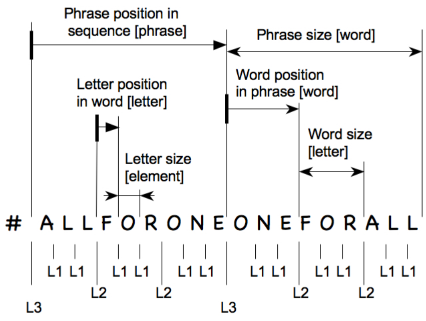

Fig.1. Definition of parts of stimulus sequence

research prototypes); the capture and analysis of continuous extended behaviour sequences encompassing multiple chunks; the use of relatively naturalistic tasks in an experimental context.

The primary aim of this paper is to introduce a model of the processes involved in the graphical production of complex stimuli that predicts the patterns found in the temporal chunk signal in the previous GPA experiments. The model provides an account of how complex sequences (e.g., the saying above) are processed and of the effects discovered in the previous studies. The model is introduced in the next section and its explanation of the previous experiments then follows. A second aim of this paper is to present a new experiment that is a further test of the graphical chunk production model.

Process model of graphical chunk production

First consider some terms to define hierarchical chunk structures. Fig. 1 shows the levels that constitute a given stimulus or *sequence* and identifies the position and size of constituents of the levels [with unit measures in brackets]. A sequence is composed of one or more *phrases* (level L3); a phrase is composed of one or more *words* (L2); a word is composed of one or more *letters* (L1); letters are composed of graphic *elements*. A *mark* is the action that makes a physical graphical element. A *pause* is the duration between placing the pen down on the paper to start a given mark and lifting the pen at the completion of the *previous* mark.

The proposed graphical chunk production model hypotheses 12 “constraints”, which are classified into three classes. There are five *global level* constraints that govern the overall form of the model: (A) The processes of graphical production operate in a serial fashion with no overlapping. (B) Graphical production takes chunks in the form of a hierarchical structure. (C) Processing occurs in a depth first manner with processing being pushed down to the next level at the earliest opportunity and returning up to lowest level at which chunks still remain to be processed. (D) The processing of chunks at a given level involves selecting a chunk and retrieving its sub-chunks on the level below from memory. (E) The pause before the production of an element (mark) is the sum of the durations needed to execute all of the processes over successive levels since the completion of the previous element. The four *local level* constraints refer to processes at individual levels: (F) The processing demands and time required for the selection of sub-chunks in working memory increase with the complexity of the sub-chunk structures. (G) Processing demands and time increase for retrieval from long-term memory with the complexity of the sub-chunk structures. (H) The processing of graphic elements, level L0, includes the programming of the movement of the pen to the starting position of the element and executing the movement. (I) Longer distances between the end of one mark and the beginning of the next mark will increase the duration of the associated pause (Fitt’s law might be used to compensate for this). The final set of *stimuli-structure* constraints concern the impact of the nature of the stimulus: (J) For linear sentence-like stimuli, the processing of chunks will follow the order of presentation in the sequence. (K) When the order is not given (e.g., elements within a letter), the order will follow an individ-

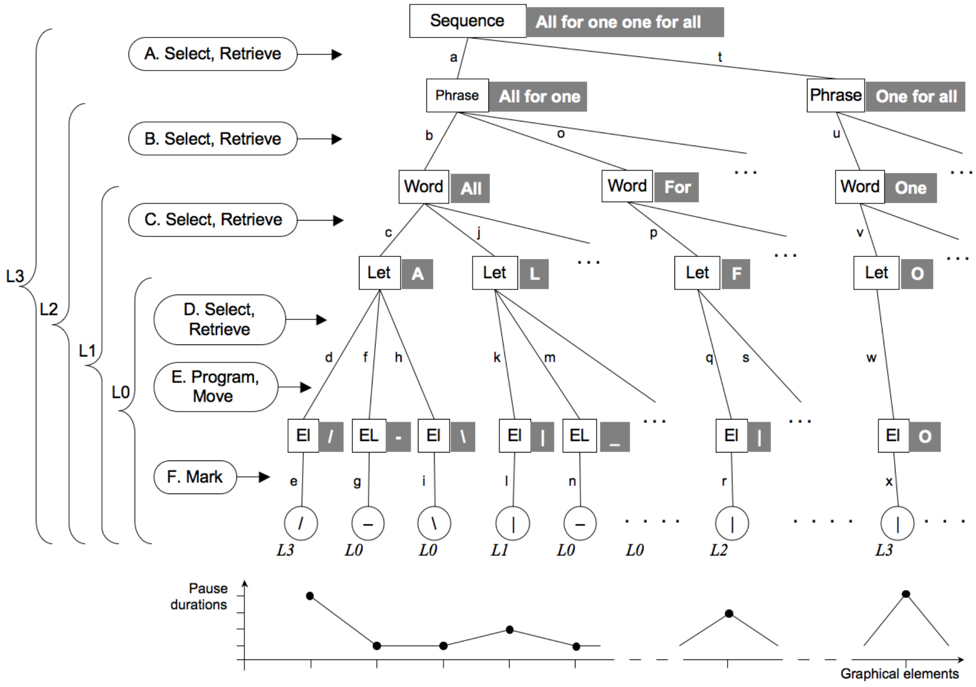

Fig. 2. Relations between chunk structure, processing steps and pause levels. (Letters a-x label processing steps.)

ual participant's typical pattern of production. (L) The position of the starting point of a mark for an element is determined by the stimulus structure, or by the individual's typical pattern of production, or both.

The temporal aspects of constraints E, F, G and I provide a basis for estimating the duration of individual pauses.

Clearly variants to the proposed constraints may be consistent with our current knowledge but for the sake of clarity in introducing the model the constraints have been stated in their simplest and strongest form. (See the discussion for one variant on the serial processing constraint-A).

Fig. 2 provides an example of how the hypothesised constraints determine the writing of our adopted saying – *All for one, one for all*. The overall chunk for this sequence is given so the first processing step (Fig. 2, process-a) is the selection of the first phrase sub-chunk and its retrieval – applying constraint-D. Processing then goes down one level – constraint-C – with the selection of the first word chunk and its retrieval (process-b) – constraint D. Next processing occurs at the letter chunk level – constraint-C – with the selection and retrieval of the first letter (process-c) – constraint-D. The particular sub-chunks selected by processes a, b, and c are determined by the stimulus – constraint J. The first element of the letter is selected and retrieved – constraints-C, D and K – and the pen moved to the start location for the element – constraints H and L (process d). The mark is then made (process e). Processing now returns to the letter level as elements still to be processes for a letter – constraint C. The next two elements are processed (processes f-g, and h-i), following which the processing returns to the word level for the processing of the next letter in a similar fashion (processes j-k-l-m-n etc.). The final letter is then produced (ellipsis in Fig. 2) and processing moves up to the phrase level – constraint-C – and the next word is processed (processes o-p-q-r-s etc.). Once all the words of the current phrase are produced, processing returns to the phrase level and the next phrase processed (processes t-u-v-w-x etc.).

On the basis of constraints E, F, G and I the (relative) duration of the pause prior to producing a given mark can be inferred from the sequence of serial processes involved. The pauses for the production of marks at the beginning of phrases (L3), words (L2), letters (L1) and elements (L0) will have the relation L3>L2>L1>L0, because the amount of processing needed just prior to making a mark is successively less at each level. The schematic graph at the bottom of Fig. 2 illustrates the general shape of predicted pattern of pauses associated with successive elements for this example. The actual durations of the pauses will be affected by the structure of the sequence and by the distance that the pen is moved between marks – constraints F, G and I.

Model Consistent With Previous Findings

Findings in our previous experiments using GPA with complex stimuli are consistent with the graphical chunk production model. In the Cheng & Rojas-Anaya (2005) experiment participants memorised and wrote sequences of numbers with different groupings (e.g., 303 5 404 5 … versus 30 35 40 45 …). In Cheng & Rojas-Anaya (2006) participants wrote word phrases, such as our exemplar saying. The stimuli in both the experiments thus had predetermined chunk structures with three levels: L0, L1 and L2. In order to minimise temporal differences in the pauses before letters within and between words all the letters were written individually in a series of regularly spaced rectangles printed on the response sheet. In all three experiments the L2 pauses durations were significantly greater than L1 pauses, and L1 pauses were significantly great than L0. For the writing of simple number sequences L1≈280 ms and L2≈440 ms (Cheng & Rojas-Anaya, 2005). For familiar and unfamiliar word phrases L1≈270 ms and L2≈400 ms (Cheng & Rojas-Anaya, 2006). Note the similarity between these pairs of times. It was found that the L2 pause for more complex words with more letters tended to be longer than for simpler words (e.g., 3035 & 4045 versus 30 & 35). Further, with the word phrases, it was also found that words processed later in a phrase tended to have shorter L2 pauses, which coincides with the decreasing complexity of remaining part of the phrase yet to be processed.

In the Cheng, McFadzean & Copeland (2001) experiment participants were given the names of previously memorized simple geometric patterns, which they reproduced. For the drawing of simple geometric objects the pauses were L1≈410 ms and L2≈620 ms. The longer pauses for drawing compared to writing might be attributed to some intrinsic differences between these two modes of graphical production. However, the graphical chunk production model provides a basis for more precise process explanations. One such is a difference due to the particular nature of the drawing task, in which target objects were given by name rather than being shown as diagrams. This requires an extra translation process from name to image that is in addition to the selection and retrieval processes (Fig. 2, labels b and c). This is a rational variant of constraint-D. Another contribution (and explanation) of the longer times is the greater distance that must be moved between successive "letters" and "words" in the drawing task compared to the smaller fixed distance for number sequence and word phrase tasks – constraint-L.

In contrast with the above experiments, in which chunk structures were induced by participants memorising predefined patterns, Cheng & Rojas-Anaya (2007) had participants with four substantially different levels of experience to copy mathematical expressions (with out restrictions on the position of the symbols). It was predicted that the more expert participants would parse the expression as a small number of chunks each encompassing many symbols whereas the novices would use many small chunks. The observed pattern of pauses was consistent with this prediction to a high degree of reliability at the level of individual trials.

In all the experiments, patterns in the graphs for individual participants doing single trials resemble the patterns shown in the ideal graph in Fig. 2 (examples are given in Cheng, McFadzean & Copeland (2001), Cheng & Rojas-

Anaya (2005, 2006, 2007)). The temporal chunk signal in GPA is strong and robust, which means that meaningful patterns can often be found in data without the need to aggregate over multiple participants, multiple trials or both.

Overall, the findings across the previous experiments are consistent with the proposed graphical chunk production model. They do not provide evidence of the sufficiency and necessity of the theoretical constraints (A-L), but the findings are consistent with the model taken as a whole. The remainder of this paper presents a somewhat more direct test of the model.

Artificial Sentences Experiment

This experiment used artificial sentences as the stimuli, so that the structure of the sequences could be systematically manipulated. Four types of letters were used, consisting of one to four marks: I, T, H and *, where * represents four crossing lines (i.e., '+' and 'x' superimposed). Each word consisted of one type of letter repeated one to four times. Phrases consisted of sets of three or four of these words and sequences (sentences) were made up of repeats of the phrases. Table 1 gives the number of letters in each word and the pattern of words in each repeating phrase.

The model is challenged in this experiment in two separate ways: (a) by increasing the number of levels in the hierarchical chunk structure to four, with the inclusion of L3 level phrases; (b) by not imposing a chunk structure using a plain sequence of letters with no word or phrase structure (S9, Table 1). The model predicts: (a) that the inclusion of phrase level structure will result in an addition L3 pause level that will be substantially greater than the L2 pauses; (b) that a lack of structure above the letter level will result in pauses being produced only at the L1 level and not above.

The experiment also included the systematic variation of the phrase pattern (S1-S8), word length (1 to 4 letters), and the letter size (1 to 4 elements) to allow second order effects of the complexity at different hierarchical levels in the model to be probed.

Table 1. Definition of artificial sentence words & phrases

S1 - Hash, One-Two-Three, One-Two-Three, …
S2 - Hash, Two-Three-Four, Two-Three-Four, …
S3 - Hash, Three-Two-One, Three-Two-One, …
S4 - Hash, Four-Three-Two, Four-Three-Two, …
S5 - Hash, Three-Three-Two-Two, Three-Three-Two-Two, …
S6 - Hash, Two-Two-Three-Three, Two-Two-Three-Three, …
S7 - Hash, One-One-Four-Four, One-One-Four-Four, …
S8 - Hash, Four-Four-One-One, Four-Four-One-One, …
S9 - No sequence

Method. The experimental procedure was similar to that used in the number sequence and word phrase experiments (Cheng & Rojas-Anaya, 2005, 2006). The participants were presented with a card showing one of patterns from Table 1 and told which letter to use. They were instructed always to begin a sequence by writing a hash (#), which ensured that writing was smoothly underway before the occurrence of the first element of interest. They wrote each letter in a horizontal set of 28 uniformly spaced rectangles printed on the response sheet, which maintained a largely constant pen movement distance between all letters irrespective of their level. Writing stopped when all the spaces were filled. Participants said the pattern out loud as they wrote each sequence so that the experimenter could verify they were using the correct pattern. They wrote sequences without thinking of a pattern when they saw the S9 'no sequence' card. S1-S8 were written once and S9 twice for each letter type. The order of the stimuli was randomised.

A standard graphics tablet (Wacom Intuos$^{2®}$) and specially designed drawing/writing analysis software, TRACE (Cheng & Rojas-Anaya, 2004), were used to record the writing actions, to extract the pen positions and times, and to compute the duration of pauses between drawn elements.

Participants. Our previous studies (Cheng & Rojas-Anaya, 2005, 2006) with comparable stimuli showed that significant patterns occur in data from individual participants in single trials. The design of the present experiment meant

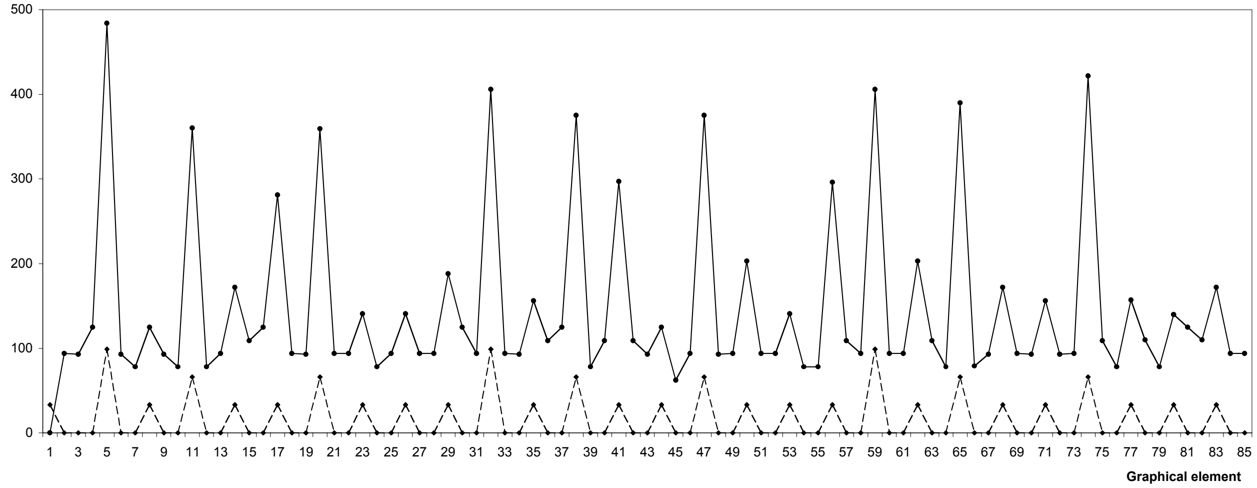

Fig. 3. Temporal chunk signal protocol graph for MT writing stimulus S2 with the letter H.

that there would be between three to five L3 pauses per trial, 28 L3 data points per letter type, and 112 L3 data points per participant. There were approximately three times the number of L2 data points. Therefore, just two participants, MT and NP, were recruited for the experiment. They were adults working at the University of Sussex.

Results

The overall pattern of results for MT and NP was comparable with a closely consistent pattern of effects found for both the participants.

Example protocol graph. Fig. 3 shows a typical example of the temporal chunk signal. MT is writing the S2 sequence with phrases consisting of three words with two, three and four repeats of the letter H. The solid line gives the measured pauses and the dashed line shows the expected chunk structure with the chunk level multiplied by an arbitrary 33 ms (e.g., 66 ms for L2). Each peak corresponds to the writing of the first element of each letter H and is followed by two points corresponding the second the third elements. The pattern of pause durations matches the predicted pattern well.

L3, L2 and L1 pauses across trials. For each trial with each participant (e.g., Fig. 3) the pauses corresponding to the four levels were coded and means computed. Fig. 4 shows all these value for the 32 trails by MT. The data for NP is comparable. Analysis with medians is equivalent to that with the means, except the absolute pause durations are slightly lower. The group of data to the right of Fig. 4 without stimuli numbers are the overall means across trials S1-S8. The magnitude of mean pauses for each trials is L3>L2>L1>L0, with three exceptions for L3-L2 and one for L2-L1. This pattern of values is obviously unlikely to be due to chance. (For example, if it is assumed that the chance of L3<L2 is 50%, then by the binomial distribution with the 32 trials, the chance that three or fewer trails have L3<L2 is p<.001.) The overall means are 542, 391, 241 and 91 ms, respectively. For NP the overall means are 649, 495, 251 and 75 ms, respectively.

The no sequence S9 trails are to the far right of Fig. 4. The magnitude of L0 pauses are comparable to those in the other trials. The pauses at the beginning of each letter have been labelled as L1 because these trails had no imposed word (or phrase) structure. The magnitudes of these pauses are quite comparable to the L1 durations for the other trials and clearly far from the L2 durations. The same pattern of results is found with NP's no sequence trials.

Effects of complexity and position. The data for both participants was analysed for potential second order effects, but none were found for both participants, with one exception. For completeness the analyses conducted and cases where significant effects for just one participant are presented.

There was no effect on L3 pause duration with complexity of the phrases (compared across letter types). For MT (but not NP) there was a significant effect on L3 pauses with increasing complexity of the first word of a phrase (compared across letter types), with an approximate increase of 50 ms per additional letter. There was no significant effect on the length of L3 with the letter type (compared across the size of the first word in a phrase). There were no effects on L2 pauses of: length of words (compared across letter types); type of letter (compared across position in words); position of words in a phrase (compared across word length).

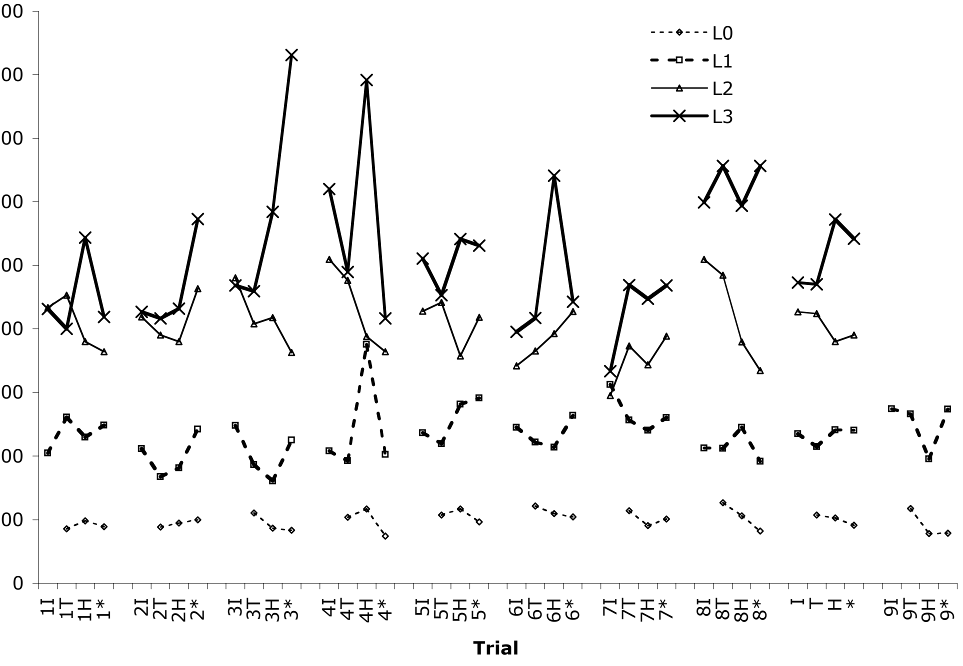

Fig. 4. Participant MT's Mean chunk level pause durations for each trial (sequence and letter)

The exceptional effect is an intriguing pattern. Compared across letter types, the pauses for the first word in a phrase is longer than the pauses for the rest of the words (i.e., L3>L2), but the L2 pause for the third word is significantly longer than the pauses for the second and fourth words (i.e., $L3 > L2_{word\text{-}3} > L2_{word\text{-}2} \approx L2_{word\text{-}4}$). The model predicts the pauses should decrease strictly in order.

Discussion

Overall the results are comparable to the effects and magnitudes of pauses found in the previous experiments, especially the number sequence and word phrase experiments (Cheng & Rojas-Anaya, 2005, 2006) that used the same task methodology. Hence, it is unlikely that some thing unique about the general nature of the artificial sentences is responsible for the overall pattern of results.

The overall pattern of results is consistent with the main predictions derived from the model. In particular there is a distinct L3 level of pauses corresponding to the new phrase level and the magnitude of the letter level pauses in no sequence trials is comparable to the L1 values in the other trials. This provides further evidence for the model's global constraints – graphical production proceeds largely through: (A) serial processes; (B) operating over a hierarchical structure of chunks; (C) in a depth first manner. Although the predictions and findings are consistent at this level, claims about the specific validity of these constraints should still be qualified, because the design of the experiment did not attempt to independently manipulate each constraint: if indeed such a design could be envisaged. This and the previous experiments do not rule out the possibility that other quite different combinations of constraints or particular variants of the posited constraints could be necessary and sufficient. For example, some systematic overlapping of the selection, retrieval, motor programming, moving and marking processes cannot be ruled out by the existing evidence.

The model posits second order effects regarding the impact of the complexity of the chunks to be processed at different levels (constraints F & G). No significant effects were found on the L3 and L2 pause level due to greater complexity at the phrase or word level. One possibility is that these constraints do not, in fact, impact on graphical production. Given the ubiquity of such effects in other tasks this seems unlikely. Another explanation is that there was insufficient statistical power in the present experiment to reveal such effects. Such effects were found in our previous experiment, which involved about one third the number of trials but five times the number of participants, so it is surprising that there were so few hints of trends in the data (with the exception of MT's effect of first word complexity on L3 and the intriguing pattern). A third and perhaps the most plausible explanation of the lack of chunk complexity effect trends comes from the recognition in hindsight that the design of the artificial sentence may not be having the expected impact, because of the small variety of words and letters included, and by the use of the same letter within each trial. Although logical complexity of the structure of the sentences did vary, this may not have translated into computational complexity because of priming effects. This contrasts markedly with the substantially greater variety of the stimuli in the previous experiments.

The intriguing $L3 > L2_{word\text{-}3} > L2_{word\text{-}2} \approx L2_{word\text{-}4}$ pattern presents an interesting challenge to the constraints posited by the model. A similar pattern was observed in the experiment with the simple geometric diagrams at the L1 letter level, but not reported in detail. We are currently designing experiments to probe this phenomenon and to further test the proposed graphical chunk production model.

References

Anderson, J. R. (1998). *The atomic components of thought*. Mahwah, N.J.: Lawrence Erlbaum Associates.

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 *Proceedings of the 23rd 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. (2005). Writing out a temporal signal of chunks: patterns of pauses reflect the induced structure of written number sequences. In *Proceedings of the 27th Annual Conference of the Cognitive Science Society* (pp. 424-429). Mahwah, NJ: L. Erlbaum.

Cheng, P. C.-H., & Rojas-Anaya, H. (2006). A temporal signal reveals chunk structure in the writing of word phrases. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: L. Erlbaum.

Cheng, P. C. H., & Rojas-Anaya, H. (2007). Measuring mathematic formula writing competence: An application of graphical protocol analysis. In *Proceedings of the 29th Annual Conference of the Cognitive Science Society* (pp. 869-874). Austin, TX: Cognitive Science Society.

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

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.

Howes, A., Vera, A., Lewis, R. L., & McCurdy, M. (2004). Cognitive constraint modeling: A formal approach to supporting reasoning about behavior. In *Proceedings of the 26th annual meeting of the cognitive science society* (pp. 595-600). Mahwah, NJ: Lawrence Earlbaum.

John, B. E., & Kieras, D. E. (1996). The goms family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer-Human Interaction, 3*(4), 320-351.

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

McLean, R., & Gregg, L. (1967). *Effects of Induced chunking on Temporal Aspects of Serial Recitation*. *JEP*, 74(4), 455-459.

Rosenbaum, D. A., Hindorff, V., & Munro, E. M. (1987). Scheduling and programming of rapid finger sequences: tests and elaborations of the hierarchical editor model. *JEP:HPP, 13*(2), 193-203.