Causes for segmentation of keying sequences: a predictive model

Merijn Bruijnes (s0097519)

July 12, 2010

The University of Twente, The Netherlands

Abstract

Literature has been studied to find potential causes of segmentation during motor learning of a keying sequence in an attempt to make it possible to predict the segmentation of any sequence for any person. A categorization of causes for segmentation (also patterns or salient points) of a keying sequence has been made; *procedural causes, sequence based causes,* and *participant related causes*. It was investigated whether participants will segment a sequence at a salient point in the sequence, indicating predictability in segmentation might be possible. Inter hand transitions (IHTs) and runs of consecutive keys were chosen as salient points. Also, it was investigated whether a sequence entirely made up of salient points with one key transition without such a salient point would segment at the 'dull point', it did not. Little evidence for a successful prediction of consistent segmentation using IHTs was found, none using runs. Interpretation of the data was hindered by finger effects due to incomplete balancing. Recommendations for future experiments are given.

Keywords: Motor learning, keying sequence, chunking, segmentation causes, inter hand transition.

Number of words: 12800.

Introduction

A sequence of movements that is practiced over and over again can group consistently into segments. For example, when typing my name I always have a small pause between two segments: "Mer" and "ijn". Thus, in this example a well practiced keying sequence groups into segments of several responses. This report will attempt to find potential causes of segmentation during motor learning of a keying sequence in an attempt to make it possible to predict the segmentation of any sequence for any person.

There are several sequence production paradigms that were used to investigate motor learning (and segmentation), three important and relevant paradigms will be discussed next, the serial reaction time task, the discrete production task, and the 2 x N paradigm. The Serial Reaction Time (SRT) is the most popular paradigm to investigate implicit learning, motor learning and chunking (Jiménez, 2008). An early SRT experiment, conducted by Nissen and Bullmer (1987) consisted of four lights with buttons directly below them and required participants to respond as quickly as possible to a light by pressing the button directly below it. After each response the next stimulus was presented after a response to stimulus interval (RSI) of about 200ms. The SRT usually repeats a (long) sequence over and over again. Another frequently used paradigm to investigate motor learning and chunking is the Discrete Sequence Production (DSP) task (e.g. Verwey, 1999). A DSP also task uses key specific cues which require an immediate response from the participant. However, a DSP generally does not contain a RSI (i.e. when the response is detected, the next stimulus is generated) which is the main difference between SRT and DSP. Sequences in the DSP are usually comprised of six or eight items (Rhodes, Bullock, Verwey, Averbeck, & Page, 2004). In the DSP sequences have limited length and the sequences are easily recognized from the first stimulus which encourages internal (e.g. long-term- or working-memory-guided) control (Rhodes et al., 2004). Finally, the 2 x N paradigm (e.g. Sakai, Kitaguchi, & Hikosaka, 2003) uses LED buttons that are usually arranged in 4x4 matrices. Two LED buttons are illuminated simultaneously, this is called a set, and participants have to sequentially press the illuminated buttons. The correct order of the set of button presses is predetermined by the computer, and the participants have to find it by trial and error. Hyper sets of N consecutive sets are presented in a fixed order. There are still other paradigms and many (small) variations of the paradigms discussed above, but they are not discussed here in detail.

A sequence of movements can be segmented consistently. It is suggested that segmentation occurs to prevent overloading the human information processing capacity and allows for more specific representations (i.e. motor chunks). A chunk is a group of several items that are represented in memory as one. According to Miller (1956) most mental (sub)systems seem to utilize chunking; among which are processing and memorization of complex stimulus patterns over all modalities. In motor learning chunking is also found to occur (e.g. Verwey, 1996). A motor chunk generally contains 3 to 5 individual motor memory unit and it appears that the exact size of a motor chunk varies between subjects and situations (Miller, 1956; Verwey, 2003b; Verwey & Eikelboom, 2003; Verwey, Abrahamse, & Jiménez, 2009). Chunking can be a strategy in which any multiple items can be used in a relational structure to produce elaborate sequences or representations. This means that several motor chunks can be concatenated by a cognitive processor to form a longer sequence of movements (Verwey, 1996). Also, motor chunks are found to be robust in that chunks are available for use in different situations than in the situation they were acquired (Verwey, 2001).

The location of a switch between two consecutive segments (or motor chunks) can be detected by a significantly longer pause. When the last key press of the first chunk is pressed the next chunk needs to be loaded in the motor memory, this takes some time. This means that the time between the last key press of the first chunk and the first key press of the next chunk is longer compared to the other pauses between key presses (e.g. Verwey & Eikelboom, 2003; Verwey et al., 2009). Further, whether a motor chunk has developed can be determined by testing whether participants can produce the sequence without information about the individual key presses within a chunk (Verwey, 1999).

Motor chunks appear to be forgotten or strengthened following the rules of associative memory, implying that anything that causes a sequence to be segmented in a consistent way might be a significant cause for motor chunks to form. This report will try to shed light on the causes of segmentation and motor chunking, seeking a model that might predict where and how motor sequences segment. After reviewing a wide array of literature I propose three categories of potential causes of consistent segmentation (i.e. motor chunking) in a motor sequence; *procedural related causes, sequence related causes,* and *participant related causes* for segmentation (see figure 1). First taxonomy will be established after which each category will be discussed in more detail. Finally, an experiment will be described in which an attempt was made to predict the segmentation of several sequences using the suggested causes for segmentation.

Procedural related segmentation causes are imposed on the sequence and the participant and are external factors such as instruction or stimulus presentation. For example the setup of an experiment or more concrete: the prestructured sequence (e.g. Verwey et al., 2009) or a RSI (e.g. Stadler, 1993). *Sequence related causes* for segmentation describe the specifics of the (to be) learned sequence that might cause segmentation. For example, a run (i.e. a run consists of keys next to each other) from left to right followed by a run from right to left might cause both runs to form individual motor chunks (e.g. Koch & Hoffmann, 2000a). *Participant related causes* for segmentation of a sequence have their root in the participant, i.e., the physical or mental constraints of the human body. For example, the maximum number of individual key presses that can be loaded in motor memory at once, up to 5 keys (Verwey, 2003b). As can be seen in figure 1 overlap between causes for motor chunking is assumed to be possible. When two inducers for motor chunking are present in one sequence presumably smaller motor chunks are formed and learning is hampered (as found in experiment 2 in Koch & Hoffmann, 2000a). Next, the three categories and their supporting literature will be elaborated on.

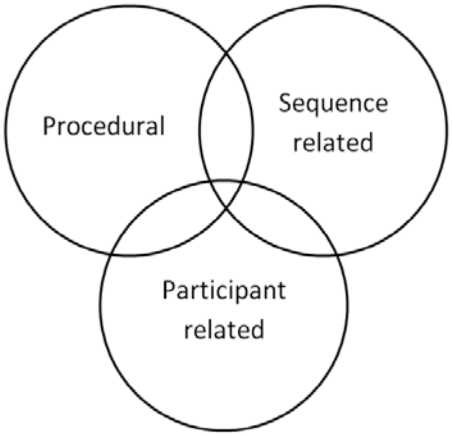

Figure 1: Three categories of potential causes of segmentation. Notice that categories might overlap causing potential difficulty in predicting sequence segmentation even when each separate category would be fully understood.

Procedural causes for segmentation

Procedural causes for segmentation are related to the experimental setup. Such factors imposed on the sequence and the participant as external factors such as instruction or stimulus presentation. It is well established that when pauses (i.e. RSIs) are introduced at some consistent locations in a sequence during the practice of a sequence (i.e. a prestructured sequence) these pauses are found to induce lasting segmentation that continues to be executed even without RSIs present in the stimuli and the induced segmentation is the same for all participants (e.g. Koch & Hoffmann, 2000a; Verwey, 1996, 2003b, in preparation; Verwey et al., 2009; Verwey & Dronkert, 1996). Stadler (1993) investigated RSIs by alternating a repeating and a random structure in a SRT task. When a RSI was presented at the end of each sequence during practice this improved learning compared to when no RSI was presented. When the RSI was presented at random locations during sequence practice this hampered learning, but did not completely prevent it. It suggested that insertion of RSIs induces parsing of the sequence into memory chunks (Stadler, 1993). In addition, Verwey, Abrahamse and

Jiménez (2009) recently found that the segmentation pattern of a sequence that was prestructured did not spontaneously transfer to other sequences (i.e. it did not induce similar segmentation in other sequences that were practiced simultaneous). Therefore there is evidence that there is no transfer of segmentation between sequences. This seems to imply that when investigating segmentation in a sequence, participants might be used in several (semi-)simultaneous sequence learning tasks without fear of transfer of segmentation.

When a series of movements is shown (e.g. a series of quasi-random hand and arm movements in a figure drawing task) to a participant, while the movements during the demonstration are segmented into sub-series the participant is likely to copy this segmentation when asked to repeat the movements (Agam, Galperin, Gold, & Sekuler, 2007). A similar finding was made for a series of movements shown and repeated on a screen (Rice & Sekuler, 2008). During a SRT when participants are merely looking at a sequence of stimuli during practice they are able to perform at a similar level as participants who actually practiced the sequence via key presses. This suggests that participants are able to learn a sequence solely by observation (Howard, Mutter, & Howard, 1992), at least for some paradigms. However, further proof for observational learning was found by Lee and White (1990). They found that when a participant (i.e. the observer) observed an unskilled participant (i.e. the model) improve on a task (e.g. a computer game that required timing) it generated more effective learning than observing a skilled model perform a task flawlessly (Lee, & White, 1990). In addition, it is found that even mental practice alone is sufficient to increase skill, though less efficient than actual physical practice (cf. for meta-analyses; Feltz & Landers, 1983; Driskell, Copper, & Moran, 1994).

The amount of practice has an effect on how the sequences are represented in motor memory. After extensive practice of a keying sequence two separate learning components seem to be present related to the fingers used; an effector-dependent and an effector-independent learning component. A learned sequence cannot transfer between effectors (e.g.

from left to right hand) when it is represented effector-dependent, and it can transfer when it is effector-independent. In SRT tasks evidence was found for effector-independent learning in sequential keying. For example, Cohen et al. (1990) found in a SRT study effector-independent learning in that learning transferred from a sequence that was practiced with three fingers to the same sequence executed with only one finger. Effector-dependent learning in the SRT task seemed elusive but was found by Verwey and Clegg (2005) however only after a much longer practice period (i.e. 1,300 sequence repetitions) than usual in SRT. Also with the DSP task evidence for effector-dependent and effector-independent learning was found. In Verwey and Wright (2004) a sequence was practiced extensively with either three fingers of the same hand or two fingers from one hand and one finger from the other hand. When in the test phase the groups from both hand conditions executed the same and different sequences an effector-dependent effect was found (Verwey & Wright, 2004). The difference between the findings of SRT and DSP might lay in the amount of practice common in both paradigms; limited practice in SRT tasks (i.e. usually around 100 sequence cycles) and extensive practice in the DSP task (i.e. usually more than 500 cycles). Further, practice appears to influence the effector-dependent component of learning as more practice of a sequence shows more effector-dependency (Park & Shea, 2003). However, the Park and Shea (2003) experiment did not use a sequential keying task and this might pose problems for generalization to the keying sequences paradigm, they used a pushing exercise. Additionally, effector-dependent effects might also be attributed to extensively practiced sequences relying on different processors operating in parallel. When different fingers than the ones used in the practice are required to execute the sequence the contribution by effector-specific processors might be eliminated (Verwey, 2003a; de Kleine & Verwey, 2009). Further, effector specific (i.e. in this case finger specific) effects might be strongest for the index finger (see figure 2) as suggested by Verwey, Abrahamse and Jiménez (2009), six out of eight segmentation point involved an index finger either before or after the slow IKI.

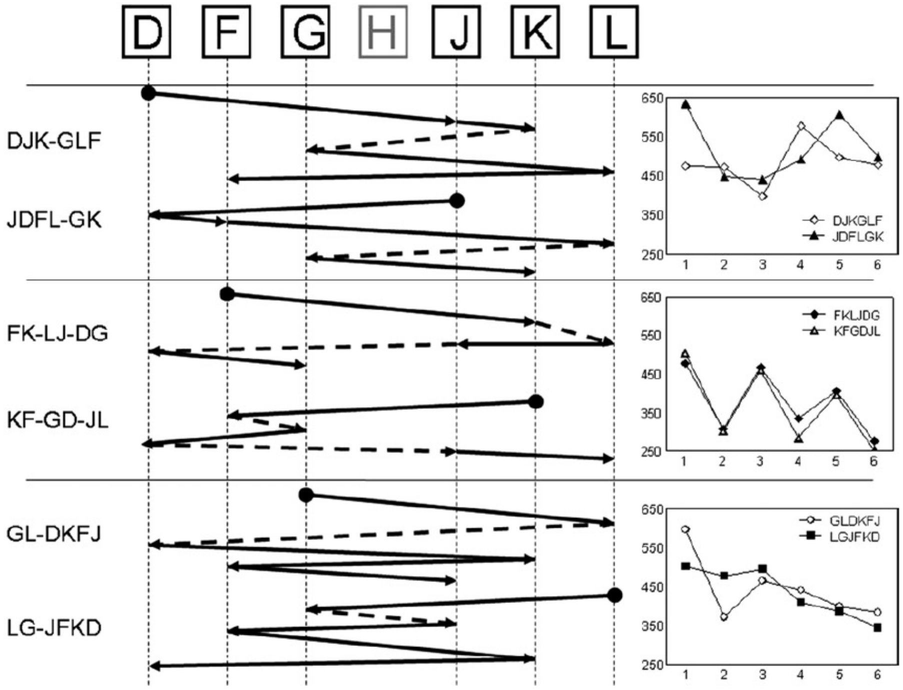

Figure 2: Average inter key intervals for each individual new sequence (right column), along with the segmentation of these sequences (left column), and a graphical overview of the temporal-spatial key press orders in each of three subgroups (middle column). The letters on top indicate the key positions on the keyboard. Keying order in the middle column for each sequence is from top to bottom. Dashed lines indicate relatively slow transitions. Eight slow transitions occurred of which six at a hand transitions (From Verwey et al., 2009).

In conclusion, results from observational learning (e.g. Agam et al., 2007; Howard et al., 1992) are likely similar to the prestructured sequence effects (e.g. Stadler, 1993; Verwey et al., 2009). The similarity suggests that external pacing factors might be generalized to all motor learning paradigms. Generalization would imply that any segmentation or regular discontinuous pacing during the training of a motor sequence might be a potential cause of the formation of motor chunks. Therefore, it seems that regular discontinuous pacing induces parsing of the sequence into chunks. However, continued research into the extent of generalization might be required, for example the amount of practice might have an effect on the extent to which parsing of a sequence into memory motor chunks works under different paradigms.

Summarizing, procedural segmentation causes are related to the procedural setup of an experiment. They are imposed on the sequence and the participant and consist of external factors: instruction, stimulus presentation, observational learning, and amount of practice.

Sequence based causes of segmentation

Sequence based causes for segmentation describe the specifics of the (to be) learned sequence that might cause segmentation. For example, a run (i.e. a run consists of keys next to each other) from left to right followed by a run from right to left might cause both runs to form individual motor chunks (e.g. Koch & Hoffmann, 2000a).

Many studies found that regularities in the sequence caused spontaneous segmentation. Koch and Hoffmann (2000a, 2000b) suggested that there can be two categories of regularities in a sequence that might be responsible for the spontaneous segmentation of a sequence: statistical and relational regularities. The statistical regularities that will be discussed next are; statistical redundancy, finite state grammar and uniqueness of transition between consecutive stimuli. After that the relational regularities will be discussed. Some of this research comes from language studies as language acquisition requires segmenting a continuous language stream into words, which makes that language studies might provide us with useful ideas concerning segmentation. Last, inter hand transitions (IHT) will be discussed. An IHT is a point in a sequence at which participants need to switch between hands (e.g. right pink to left middle finger).

On a side note, it is suggested that segmentation of longer sequences can be concealed by individual differences, however apparent regularities in the keying pattern can induce common segmentation (as can be seen in figure 3 taken from Verwey & Eikelboom, 2003). Why segmentation appears at a particular position in a sequence or why individuals often differ in their segmentation is not yet fully understood. However, several causes of segmentation, some of which post hoc, have been suggested in the literature. Also, some motor learning literature does not mention segmentation or chunking, but does mention a difference in mean reaction times between sequences or conditions. This might warrant further investigation, if possible, into the causes of this difference in RTs. The difference in RTs might be due to an unnoticed segmentation of a sequence that was averaged out over the entire trial or averaged between subjects, thus causing an increase in the mean RT (Verwey et al., 2002; Verwey & Eikelboom, 2003). The well known finding that the mean execution time increases with sequence length supports this notion. This effect is known as the sequence length effect on (mean production) rate (SLER) (Sternberg, Monsell, Knoll, & Wright, 1978; Rhodes et al., 2004). A longer sequence might be represented by more chunks in motor memory and as every chunk has a significant onset delay the mean RT increases. This notion was supported by a detailed analysis of 6-key sequences by Verwey (2003b) showing that SLER increases were due to a few slow elements in each 6-key sequence, while the fastest elements were similar to corresponding items in shorter (2-key and 4-key) sequences.

A potentially useful way of quantifying a sequential structure is the information metric structure in which information can be expressed in percent *redundancy*. As a sequence becomes increasingly random, it decreases in redundancy. A completely random sequence has 0% redundancy and a completely predictable sequence has 100% redundancy (Stadler, 1992; Stadler & Neely, 1997). For example, in Nissen and Bullemer's (1987) sequence DBCACBDCBA, the level can be one trial (D, B, C, etc.), pairs of trials (DB, BC, CA, etc.), triplets (DBC, BCA, CAC, etc.), and so on. If each level occurs in a sequence with equal frequency the redundancy is 0%. Redundancy is an useful way to express that relative to a random sequence condition, participants in a repeating sequence condition have fewer unique runs of trials to practice, and they practice them more often (i.e. they are practicing a high redundant sequence). Therefore, it seems that participants exhibit greater implicit learning when a sequence is higher in redundancy, which was confirmed in SRT studies by Stadler (1992) and Stadler and Neely (1997). The relevance of these studies for the investigation of segmentation of sequences lies in the possible predictions redundancy may provide for the induced location of segmentation points. For example, redundant levels (e.g. repetitions of a triplet ABC) might evolve into chunks.

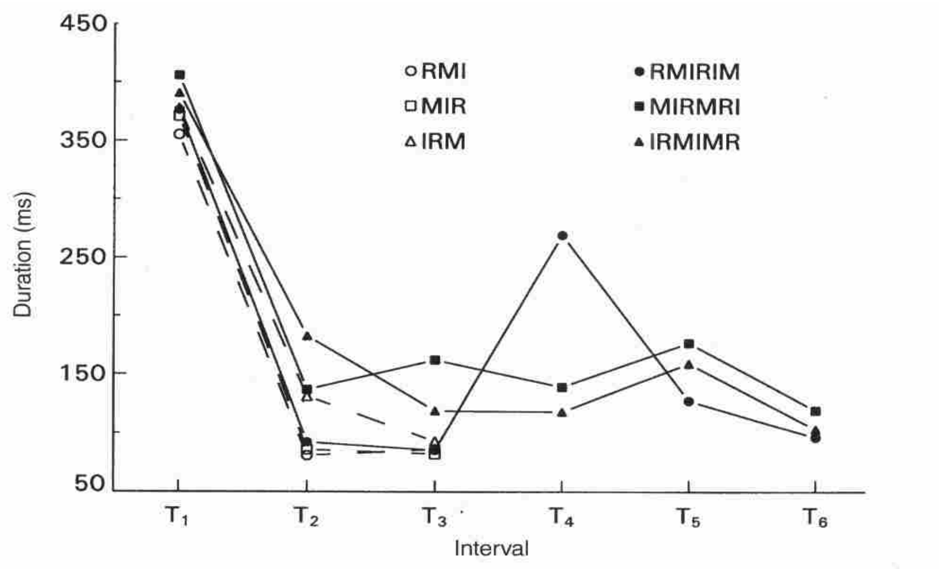

Figure 3: Three different versions of three- and six-key sequences. I = index finger press, M = middle finger press, and R = ring finger press. Sequence IRMIMR involved a five-element right-to-left-to-right regularity at intervals T_2-T_3 (i.e., RMIMR). Sequence RMIRIM included a three-element right-to-left regularity at intervals T_1-T_3 (i.e., RMI). Sequence MIRMRI had no apparent regularity. The RMIRIM sequence appears to have been segmented in two parts (i.e., RMI and RIM) as can be determined from the long T_4 interval which is the onset for the second segment (from Verwey & Eikelboom, 2003).

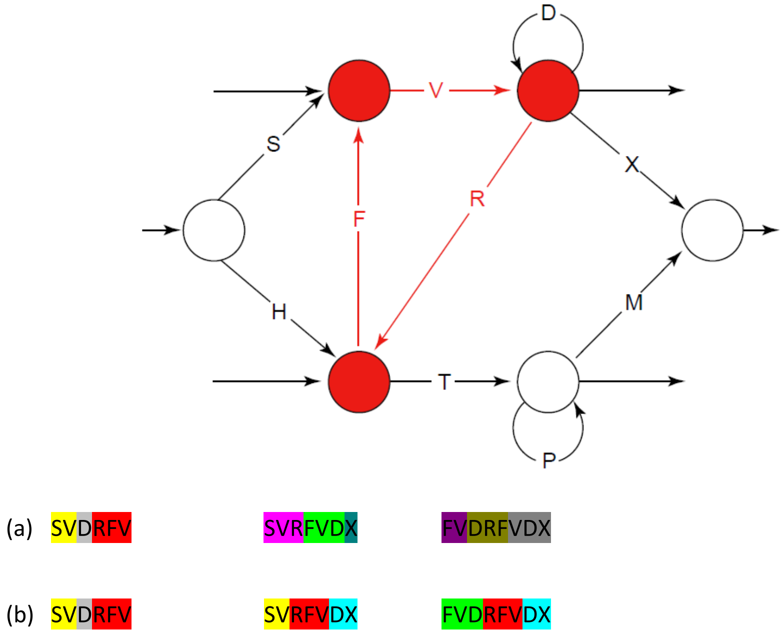

Figure 4: Top: A typical finite-state grammar, with the recursive loop RFV highlighted. Below: Three sequences that are generated by the grammar. Chunks (denoted by different colors, with a maximum size of three letters) in line (a) were chosen random, while chunks for line (b) were not chosen random. Notice that one chunk can now implicitly encode for the recursive loop RFV that is present in all three sequences. This is more efficient, as less unique chunks are needed. (Perruchet & Pacton, 2006)

Some researchers have used sequences created by a *finite state grammar* to investigate implicit learning (e.g. Jiménez, Mendéz, & Cleeremans, 1996). Finite state grammar might provide an explanation of how participants are able to chunk efficiently. A finite state grammar consists of a set of rules that can be represented in a flow-chart, following the flow-chart from the *IN*- to the *OUT-node* traversing several nodes with every path containing a symbol that will make up the sequence as each node in the grammar is passed (see figure 4). A finite state grammar is capable of producing several sequences that share the same grammar, yet may be very different. Interestingly, when participants memorized a set of sequences which they knew were generated by a finite state grammar they were able to determine significantly better than chance whether novel sequences followed the grammar or not. However, participants were unable to explicitly describe the grammar or rules (Cleeremans, Destrebecqz, & Boyer, 1998). Perruchet and Pacteau (1990) showed that grammaticality judgments by participants who initially studied grammatical letter strings did not differ from judgments by participants that learned a list of the bigrams making up these strings. These results indicate a fragment-based or chunking approach to exploit the redundancy of the training material by decomposing it into short chunks such as bigrams or trigrams. New sequences are classified according to how many chunks they share with the training material. Interestingly, segment selection does not appear to be random (Perruchet & Pacteau, 1990; Perruchet, Vinter, Pacteau, & Callego, 2002) (see bottom figure 4, for an illustration). In a study by Perruchet and others (2002) participants were instructed to mark the 'natural segmentation points' in strings generated by a finite state grammar. Over practice participants noted segmentation that was increasingly in line with the structure of the grammar, a result that was implemented in a computational model; PARSER (which is discussed in more detail in the appendix). According to the PARSER model chunks are formed on a random basis as the attentional processing of the information is capacity limited. The formed chunks are then forgotten or strengthened following the rules of associative memory (Perruchet et al., 2002).

In addition, Servan-Schreiber and Anderson (1990) found that participants who studied a finite state grammar via generated sequences had difficulty detecting violations of the grammar if the chunks they had formed were respected. However, the use of chunking strategies in grammar learning was challenged in another study. It was stated that a chunking strategy is used when little information about the grammar is available, but when more information is available chunking strategies seem to disappear and more abstract representations of the grammar rules are used (Meulemans & van der Linden, 1997). Unfortunately, these studies might not generalize well to motor learning as they used recognition of sequences and the discrimination of higher order representations of sequences instead of the production of sequences that are the interest of this article. However, it does show that participants tend to use chunking strategies when little information is available about a sequence. Little (explicit) information is usually available to the participant in most SRT or DSP tasks. Therefore, (chunking) strategies used by participants in finite state grammar research might share some similarities to strategies used in SRTs and DSP tasks.

A predictable structure in a sequence seems to enhance (implicit) learning; for example *uniqueness of transition* between consecutive stimuli (Cohen, Irvy, & Keele, 1990). In their experiment a distracter task was inserted in the intervals between responses and subsequent stimuli of the primary task, which appeared to block attention to the relation between successive events. A sequence that has unique transitions between consecutive stimuli (e.g. 15243, after a 1 always a 5) could be learned during the distracter task. A sequence with transitions between consecutive stimuli of which none are unique (e.g. 132312, a 1 is followed by a 2 or a 3, etc.) was much harder to learn during the distracter task and seemed to require more attention for learning. Interesting note, a chunking strategy (132 – 312) might have solved this. Unfortunately, due to the nature of their experiment (i.e. a SRT with a 200ms RSI) delays in certain inter key intervals (IKIs) revealing a chunking strategy would have been obscured by the RSI and no such analysis was made (Cohen et al., 1990).

Concluding the statistical patterns, even in a sequence with no apparent pattern, segments might be chosen randomly or not. However, statistical pattern research seems to indicate that segments are not chosen at random even if the sequence has no apparent structure (Perruchet & Pacteau, 1990).

Relational structures in sequences consist of systematic relations in the successive stimuli or responses (Restle, 1970), which Koch and Hoffmann (2000a) named “relational patterns”. Such patterns are beneficial to learning. An example of a relational pattern would be the sequence 123321, which can be split into two triplets. In both triplets the individual numbers are related by transpositions; the numbers are ascending and descending *runs* of three consecutive keys. Further, the triplets are related by an inversion, the first triplet is inverted by the second. Relational patterns can occur at different levels and can be similar or different for each level in a sequence, i.e., levels similar to the levels described by Stadler (1992) (e.g. a level can be one stimulus, or pairs of stimuli, or triplets of stimuli, etc.). Several relational patterns or classes of transitions are distinguished; repetitions, transpositions, alterations or inversions (Restle, 1970; Koch & Hoffmann, 2000a) (for more details on these transitions, see the appendix). Participants in SRT studies sometimes report noticing certain relational structures, for example a right to left run in Hoffmann and Koch (1997). In an SRT study Koch and Hoffmann (2000a) compared two statistical identical sequences (of 24 items each) while varying the relational pattern structure (i.e. triplet runs). It appeared that sequences containing runs (i.e. relational patterns) showed a significant learning effect and sequences without any runs did not. Examining the mean RT-pattern after some practice blocks revealed that the sequence appeared to have been chunked into triplets (which were all runs of three consecutive keys). It was suggested that relational patterns facilitate the formation of consistent chunks which improves learning (Koch & Hoffmann, 2000a). Note that a run (as I define it) does not necessarily consist of three consecutive keys; more or less keys are possible as long as the keys are consecutive. An informal analysis I conducted (visual inspection of RTs) on data from de Kleine (2009) seemed to indicate that the boundary of a run of two keys was likely to have a slow RT (i.e. five out of six possible occurrences). A run of three keys seemed even more likely to have a slow RT at its boundary (i.e. two out of two).

Another sequence based potential cause for segmentation is an *inter hand transition* (IHT), this is a point in a sequence at which participants need to switch between hands (e.g. right pink to left middle finger). The opposite of an IHT is a within hand transition (WHT) (e.g. right pink to right middle finger). Participants performing a sequence in a DSP task that requires an IHT seem to form segmentation points at a transition between the hands, see figure 2. An IHT might cause the effector-specific processor to load another reference frame; the spatial reference frame from the other hand (Verwey et al., 2009). However, the amount of practice might be an interacting factor as the earlier described effector-specific representation might only form after (extensive) practice. Koch and Hoffmann (2000a) found that participants divided a sequence that contained relational patterns (triplets) into chunks in a predictable way; segmentation occurred at the triplets which contained relational patterns. However, they did report some irregularities in the triplet segmentation pattern and two out of three reported irregularities could be ascribed to IHTs (Koch & Hoffmann, 2000a). Thus, with extensive practice effector-specific representations of keying sequences can be formed which might be caused by a sequence segmenting at an IHT.

In conclusion, motor sequence learning seems to benefit from patterns in a sequence. Three types of patterns can occur; statistical, relational, and inter hand transitional patterns. Statistical patterns can be statistical redundancy, finite state grammar, and uniqueness of transition between consecutive stimuli. Relational patterns can be relational structures in sequences. IHTs are the points in a sequence at which participants need to switch between hands. All patterns in a sequence might facilitate the parsing of the sequence into manageable sized segments (motor chunks), which can be loaded into the motor memory buffer as proposed by Verwey (1996).

Participant related causes for segmentation

Large individual differences can be found in the segmentation of a sequence. There are factors that have their root in the participant and can have an influence on the segmentation of a sequence, i.e., the physical or mental constraints of the human body.

Working memory size seems finite, something Miller (1956) already proposed and this might be an indication that the maximum size of a chunk that can be loaded into motor working memory is also finite. Much (implicit) evidence for the finite size of motor memory has been found. Most sequences seem to segment in smaller parts even if a single larger segment would be more economical (e.g. Verwey, 1996, 2001). Also, most sequences seem to have some slow elements and which element is slow might differ between participants (e.g. Verwey

& Eikelboom, 2003; Sakai et al., 2009). One might argue that segmentation is always due to some pattern or regularity; however a 2003 study seems to indicate that lasting segmentation does occur even without a pattern or regularity (Verwey & Eikelboom, 2003). Thus, the capacity limitation of the working memory might determine the boundaries of motor chunks. This is exactly what Bo and Seidler (2009) found, visuospatial working memory could predict the learning and chunking pattern. A larger visuospatial working memory leads to faster learning and bigger motor chunks (Bo & Seidler, 2009).

Individual errors at the beginning of learning were suggested by Sakai, Kitaguchi, and Hikosaka (2003) to be an important factor in the formation of chunks. In their 2 x n study they found that interruption of sequence production caused by errors may have induced a breakpoint in sequence representation which persisted even after extensive learning and during error-free performance (Sakai et al., 2003).

Strategy during the practice phase of a DSP task experiment might be an important factor in the eventual formation of chunks (Verwey, 1996). A strategy relying on chunks was found in most participants as indicated by high start-to-within segment ratios, while some participants used a different representation not indicated by sequence grouping (Verwey, 1996).

Concluding, motor sequence learning seems to be influenced by three participant related factors. The size capacity of the motor memory seems to determine the maximum boundaries of chunks. Errors made early during practice might evolve into chunks (as found in 2xn studies). And finally, participant strategy might facilitate or impede formation of chunks.

Summarizing the model

Recapitulating the large amount of discussed literature, *procedural segmentation* causes are related to the procedural setup of an experiment. They are imposed on the sequence and the participant and consist of external factors: instruction, stimulus presentation, observational learning, and amount of practice.

Three types of patterns can be found in *sequence related causes*; statistical, relational, and inter hand transitional patterns. Statistical patterns can be; statistical redundancy, finite state grammar, and uniqueness of transition between consecutive stimuli. Relational patterns can be relational structures in sequences; runs and key transition classes. IHTs are the points in a sequence at which participants need to switch between hands (the opposite is WHT).

Three *participant related factors* seems to influence motor sequence learning. The capacity of the motor memory seems to determine the maximum boundaries of chunks for an individual. Errors made early during practice might evolve into chunks (in 2xn studies). And finally, participant strategy might influence the formation of chunks.

As seen above many factors can induce segmentation. It seems that all factors that are beneficial to segmentation introduce a (persistent) pattern in a sequence. Each beneficial pattern is a salient point in a sequence as it provides structure and stands out from the rest of the sequence. The rest of the sequence is made up of *dull points* (which are the opposite of salient points), dull points are points in a sequence where there is not a pattern and which do not stand out. Compare this to a row of light bulbs; if a few are switched on they are clearly salient. However, when all but a few are switched on the active bulbs are no longer salient. In this case it is even possible that the bulbs that are off (i.e. the dull points in a sequence) become salient. A salient point might facilitate the parsing of the sequence into manageable sized segments which might be consolidated into a motor chunk over practice. Motor chunks can then be loaded into the motor memory buffer as proposed by Verwey (1996).

Further, I propose that participants will attempt to make their segments as large as possible, due to a larger chunk being more efficient than a smaller one (i.e. in a sequence of 8 keys 2 chunks of 4 keys seems more efficient than 8 chunks of 1 key). Note that this will not be tested directly in this experiment. Choosing the segmentation point might not be a conscious process, it is more likely that when segmentation occurs at a next salient point the segmentation might not consolidate as the segment is unable to fit in the motor memory at once, resulting in segmentation at a previous salient point. It should be noted that at or after every chunk the motor memory is supposedly empty (cf. CQ models, see the appendix) and the search is restarted for a salient point that is closest before the limit of the motor memory is reached. Also, the size of the motor memory is not the same for everyone, likely resulting in different segmentation for different participants. This might become especially obvious in an irregular sequence (i.e. a sequence without regularities) as an irregular sequence is expected to be segmented at the maximum motor memory size and not at a salient point as it contains none. A similar statement might be made about a sequence that contains a salient point at every key transition (e.g. a sequence with an IHT at every key). Such a sequence might also be expected to segment at the maximum motor memory size.

Summarizing, I propose that a participant will segment a sequence before the limit of the memory size is reached, making segmentation more and more urgent as the limit is approaching. In a sequence that contains salient points the last salient point before the limit of the motor memory is reached is likely to be chosen as a segmentation point, as the salient point facilitates the parsing of the sequence into segments.

Hypotheses

This paper will use the causes for segmentation found in the literature in an attempt to make predictions about the segmentation of a sequence. It will be tested whether the proposed model is able to predict the segmentation of a sequence. Unfortunately it is not easy to make segmentation predictions for any sequence for every participant, due to the individual differences in motor memory capacity. If the sequence contains salient points that are scattered too far, the segment between two salient points might exceed the motor memory capacity of some participants and unpredictable segmentation might ensue. Fortunately it is easy to develop sequences with salient points at all but a few positions. This provided us with an interesting notion that was tested in the experiment: where will a sequence segment that is entirely made up of salient points with one key transition without a salient point (the dull point)? It is expected such a sequence might segment at the dull point.

Unfortunately the scope of this experiment does not allow for thorough experimental investigation of all possible causes for segmentation. A choice had to be made which potential causes are tested. Constructing sequences in which salient points can be manipulated is difficult. Therefore, two potential causes for segmentation that are relatively easy to manipulate were chosen; IHT and run. A run of two keys can contain a dull point at the transition between the two keys. In addition, runs of two keys and IHTs are easy to incorporate in a sequence. Also, the location in a sequence of IHTs and runs can be manipulated, they can occur at different locations in a sequence or at the same location (e.g. a run and IHT at the same location: left index finger to right index finger).

This experiment used the DSP paradigm as it seems best suited to investigate the segmentation of a sequence (i.e. the formation of chunks) as it encourages internal control and as each key press takes so little time that individual inter key intervals (IKIs) are likely only caused by the action of the underlying mechanisms such as the motor buffer loading. The internal control is encouraged by the limited sequence length and because the sequences are recognized easily because there are only a limited number of sequences that might be composed (Rhodes et al., 2004). Other paradigms seem less suitable, for example the SRT paradigm might hide segmentation caused by chunking in the sequence because of the commonly used interval between the response and the next stimulus. Further, the 2 x N paradigm might inherently contain segments of 2-key length. However, it is recognized that in most serial order paradigms some form of segmentation has been reported. Also, spontaneous appearance of chunks in a 2 x N task (in a 2 x 10 hyper set) was reported by Sakai, Kitaguchi, and Hikosaka (2003) suggesting that the 2 x N paradigm might be applicable when investigating segmentation.

In this experiment it was investigated whether a sequence would segment at a run. It was expected that a run would cause segmentation (at the run boundary). Additionally, it was investigated whether a sequence would segment at an IHT. It was not expected that the IHTs would result in a consistent segmentation as all but one key-transfer would be an IHT. Also, it was investigated whether a sequence would segment at the dull point (i.e. no IHT and no run (boundary)), it was expected it would as the dull point might 'stand out'. Finally, it was investigated which of these potential causes for segmentation would be more likely to cause segmentation. This might be indicated by a more pronounced difference in RTs, i.e., a potential cause for segmentation that is more likely to cause segmentation will cause consistent segmentation in more participants. Therefore, this open question was answered by comparing the difference between the differences: run versus no-run, and IHT versus WHT. The size of the difference indicates the potency of a pattern (i.e. potential cause for segmentation). For example, the difference between run and no-run might be 60ms and the difference between IHT and WHT might be 10ms. This would indicate that a run is a more potent cause for segmentation.

Methods

Participants

Psychology students from the University of Twente (n=56; 17 male and 39 female) participated after informed consent. Ages ranged from 18 to 30 and averaged at 21.2 (SD=2.35). Participation was rewarded by a credit required for an obligatory subject pool. Participants were randomly assigned to one of the 7 sequence conditions (Table 1).

Task and procedure

The task consisted of a DSP task with sequences of 7 keys. Four fingers (not using thumbs) of each hand were placed on a qwerty-keyboard at keys **a s d f** for the left hand and **j k l ;** for the right hand. A trial began with the presentation of an 8-key representation. The representation consisted of eight horizontally aligned black outlined squares of white on a white background which was filled to indicate that the corresponding key had to be pressed. When the correct key was depressed the box turned white again and the next box was filled and so forth. When all keys from the 7-key sequence were pressed, which concluded the trial, the next trial was presented after a 1200 ms non aging interval. To reduce differences between sequences and participants one finger was never used (i.e. the right pink) since a 7-key sequence cannot use 8 fingers. Each participant performed two sequences, a sequence from each condition. The order of the two sequences was selected randomly. Participants performed four blocks (three practice and one test block) of 40 trials per sequence per block, resulting in each sequence being performed 160 times. An informal analysis of results from de Kleine (2009) showed clear segmentation of sequences from 150 repetitions. When an error was made an error message appeared and a pause of 1000 ms was inserted to irritate the participants so that they would prevent further errors after which the sequence was continued starting at the erroneous key.

Two possible causes for segmentation were investigated, runs and IHTs. The sequences used attempted to only manipulate the sequence related potential causes for segmentation that were relevant for this study (i.e. run and IHTs) and keep other sequence related causes constant or absent. As mentioned earlier, it was impossible to create a sequence with just one IHT that met the criteria: sufficient length to expect segmentation, and not repeating any keys. Therefore, all key transitions in the sequence were IHTs except for one, i.e., a within hand transition (WHT).

Two sequences of 7 keys each were used in this experiment: called separated and combined. In the separated sequence the run and the IHTs occur at separate locations in the sequence (i.e. a run and a WHT occur at the same location: left pink to left ring). In the combined sequence the run and an IHT occur in the same location (i.e. a run and IHT at the same location: left index finger to right index finger). The separated sequence used was

2351647 (represented on the keyboard as SDJAKFL) and the combined sequence was 5462713 (JFKSLAD).

In the sequence from the separated condition, 2351647 (SDJAKFL), the first key transition (2-3) is a run as the keys are next to each other and it is not an IHT. The second key transition (3-5) is the end of a run and an IHT. The rest of the key transitions are also IHTs. Thus, in the separated condition the dull point (no run boundary and no IHT) is located in the middle of the run at the second key.

In the sequence from the condition combined, 5462713 (or SDJAKFL), the first key transition (5-4) is a run as the keys are next to each other but the first transition also contains an IHT. The last key transition (1-3) is not a run (not adjacent keys) and not an IHT. The rest of the key transitions are all IHTs. Thus, in the combined condition the dull point (no run boundary and no IHT) is located at the end of the sequence at key 7 and most importantly, it is not located within a run.

Additionally, in an attempt to cancel fingers effects, both sequence conditions were balanced across fingers. It is possible that some fingers are faster or slower than others and this might interfere with the results. Therefore, it is important to counter this possible effect by counterbalancing the sequences, so finger effects will average out. Balancing is done by rotating the sequence one key at a time (i.e. placing the last key in front of the next sequence). This yields 7 versions of each sequence, see table 1. The created sequences are balanced across participants. For example, participant one (and participant 8, and 15, ..., etc) would execute SDJAKFL, while participant two would execute sequence LSDJAKF, and so forth.

Table 1

The sequences used in the experiment as represented by the corresponding keys on the keyboard. A run (i.e. two consecutive keys next to each other) is enclosed by dashes, the dashes represent the run boundaries (-). An IHT is denoted by a vertical line (|). The dull point is denoted by the dash (*), not that these are always WHTs. Each sequence was balanced across all the keys, see text for more details, resulting in 7 sets of sequences.

Sequences for participant	**Condition separated: Run separated from IHT**	**Condition combined: Run and IHT combined**
1	-S*D-|J|A|K|F|L	-J|F-|K|S|L|A*D
2	L|-S*D-|J|A|K|F	D|-J|F-|K|S|L|A
3	F|L|-S*D-|J|A|K	A*D|-J|F-|K|S|L
4	K|F|L|-S*D-|J|A	L|A*D|-J|F-|K|S
5	A|K|F|L|-S*D-|J	S|L|A*D|-J|F-|K
6	J|A|K|F|L|-S*D-	K|S|L|A*D|-J|F-
7	D|J|A|K|F|L|S	F|K|S|L|A*D|J

Design and analysis

Inter key interval (IKI) was defined as the time between the onset of the stimulus and its consecutive key press (stimulus onset co-occurred with depression of the previous key). All IKIs were recorded and saved for analysis to determine if and where consistent segmentation (implicating chunking) occurred. T1 indicated the time between the first stimulus onset and depression of the first key. The IKIs T2-T7 preceded key presses 2 through 7. T1s were not included in the analysis as they are always slow due to the onset delay (i.e. they are always the first key of the first chunk). The first two trials of every block and trials in which one or more errors had been made were excluded from analysis. Participants performed four blocks of trials: the first three blocks were learning blocks, the last block was the test block. All analyses will be based on the test block.

Whether segmentation was present was calculated by comparing the response times that were associated with the first key of a segment with those of later keys (i.e. not the first key in an alleged segment). Further, if some participants exhibit consistent segmentation (i.e. at the same key) the RT at this key will be slower. This effect will increase when more participants exhibit consistent segmentation.

To test the prediction that a run would cause segmentation, all run boundary IKIs were averaged and compared against the average of all other IKIs, and checked for significance using a paired-sample t-test. If the run boundaries are found to be significantly slower, run boundaries are likely the location of the start of a chuck.

The prediction that the IHTs would not cause segmentation was tested together with the prediction that a dull point would cause segmentation. All IHT IKIs were averaged and compared to the average of all WHT IKIs (which are the dull points), and checked for significant difference using a paired-sample t-test. If significant difference is found, the direction of the difference indicates which prediction is confirmed. If the average of the IHTs is faster than the WHT (or dull points) it would appear that the sequences are segmented consistently at the dull points and not at an IHT.

To ascertain which of the tested potential causes for segmentation would be more likely to cause segmentation, a comparison was made of the difference between the differences: run versus no-run, and IHT versus WHT. Differences were calculated by subtracting the salient point from the other IKIs (i.e. run - no-run, and IHT - WHT). The size of the differences indicates the potency of the patterns (i.e. potential causes for segmentation). The difference between the differences was checked for significance with a paired-sample t-test.

Finally, error analysis was conducted to ascertain whether there was a speed accuracy trade-off. To conduct an error analysis arcsine transformations were performed on mean error rates before the data was subjected to ANOVAs.

Apparatus

The experiment was programmed in E-prime 2.0 and ran on a PC with a CRT monitor. The responses were inputted by the participant on a standard qwerty-keyboard. The participants were seated in front of a 17 inch CRT monitor at an approximate viewing distance of 60 centimeters however the distance was not strictly controlled.

Results

T1s were not included in the analysis as they are always slow due to the onset delay. The first two trials of every block and trials in which one or more errors had been made were excluded from analysis. Participants performed four blocks of trials: the first three blocks were learning blocks, the last block was the test block. All analyses will be based on the test block (block 4).

To test the prediction that a run would cause segmentation, all run boundary IKIs were averaged and compared against the average of all other IKIs, and checked for significance using a paired-sample t-test. Keys at run boundary (M=297.2, SD=87.3) and no run (M=311.5, SD=85.9) had a difference in RTs that approached significance; t(47)=-1.82, p=0.075. Keys that were run boundaries were executed faster than no run keys, and this difference was approaching significance.

The prediction that the IHTs would not cause segmentation was tested together with the prediction that a dull point would cause segmentation. All IHT IKIs were averaged and compared to the average of all WHT IKIs (which are the dull points), and checked for significant difference using a paired-sample t-test. IHTs (M=314.4, SD=84.9) and WHTs (M=311.5, SD=105.4) had a difference that was not significant; t(55)=.31, p>.20. IHTs were marginally slower than WHTs but the difference was small and not significant.

To ascertain which of the tested potential causes for segmentation would be more likely to cause segmentation, a comparison was made of the difference between the differences: run versus no-run, and IHT versus WHT. The size of the differences indicates the potency of the patterns (i.e. potential causes for segmentation). The difference between the differences was checked for significance with a paired sample t-test. The differences run versus no-run (M=-14.1, SD=62.5) and IHT versus WHT (M=7.3, SD=53.9) was approaching significance; t(47)=1.78, p=.08. Note the negative difference between run and no-run, showing that no-run boundaries were slower than run boundaries, see figure 5.

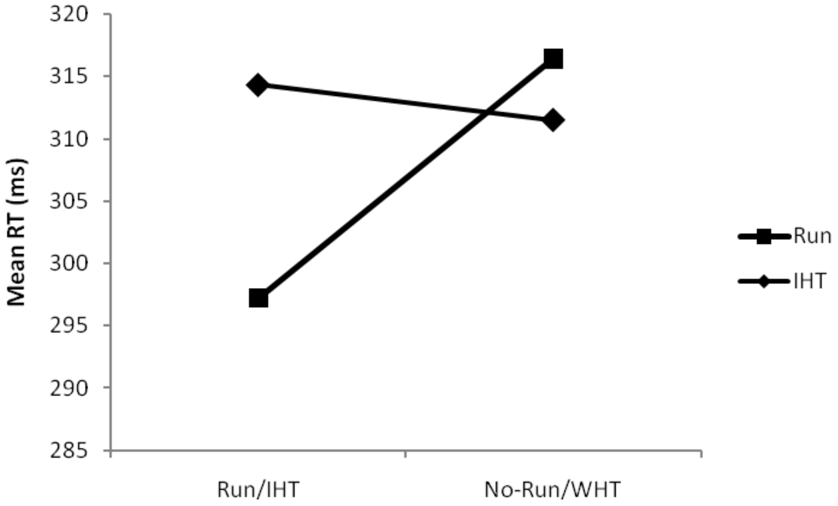

Figure 5: The mean response times for run/no-run (squares) and IHT/WHT (spades). Segmentation was expected to start with a Run or IHT. Thus Run and ITH were expected to be slower. The no-Run and WHT IKIs were expected to be faster as they would be the (faster) execution of the segment.

Table 2

Error percentages (%) per Key as a function of Block with their respective means. The error percentages from both sequences (separated and combined) were averaged for this table as the sequences did not differ significantly.

	Key	1	2	3	4	5	6	7	Mean
Block	1	2,6	3,2	2,2	3,0	3,6	3,0	3,1	2,9
	2	2,2	2,6	2,2	3,2	2,6	2,3	2,7	2,5
	3	1,8	2,5	2,4	2,7	3,2	2,6	2,2	2,5
	4	1,7	2,6	2,4	3,3	3,1	2,8	2,3	2,6
	Mean	2,1	2,7	2,3	3,0	3,1	2,7	2,6	

Errors were analyzed with a 2 (sequences separated vs. combined) x 4 (block) x 7 (key) ANOVA (note that T1s were included for the error analysis). Error rates did not differ significantly between sequences separated and combined. Error rates between blocks approached significance but were not significant; F(3, 3132)=2.17, p>.05. Error rates between keys did differ significantly; F(6, 3129)=4.48, p<.001. This could be attributed mostly to a lower percentage of errors at the first key, see table 2.

Discussion

Runs of consecutive keys

Recapitulating, in this experiment it was determined whether a sequence would segment at runs. It was expected that a run of two consecutive keys would cause segmentation. This would mean that the IKIs at the boundaries of the run would be executed slower on average than other IKIs (i.e. the start of a chunk would lay at a run boundary). However, the results from this experiment seemed to indicate that run boundary-keys were executed faster than no run keys. This is the opposite of what was expected on the basis of findings by Koch and

Hoffmann (2000a). They found consistent segmentation at runs of three consecutive keys (i.e. the first IKI of a run was slower) and they concluded that runs were the cause of the segmentation. Runs of two keys as used in this experiment might be a less potent cause for segmentation than runs of more keys.

Additionally, in the Koch and Hoffmann (2000a) study the sequences consisted of more runs (e.g. 123 234 345). Perhaps a repetition of salient patterns (such as runs) makes them more prone to consistent segmentation. This makes sense when the motor program for a sequence might become more generic or abstract when there is a clear repetition of the same salient pattern. For example, when a sequence of three runs would have to be stored in a low abstract level it would be stored in individual key presses (e.g. 123789456). When this same sequence is stored in chunks of three (e.g. 123 - 789 - 456) it might become more economical. Each chunk might be loaded in the motor memory buffer as it were one piece of information instead of three pieces of information (e.g. Verwey, 1996). Perhaps the repetition of runs as a salient pattern provides the possibility of a more generalized recall (e.g. 'start with' 1 'count three keys' - 'start with' 7 'count three keys' -'start with' 4 'count three keys'). In this generalized model of a sequence all that needs to be stored would be the start point of the runs (i.e. 1 - 7 - 4).

Inter hand transitions and dull points

In this experiment it was also investigated whether a sequence would segment at IHTs. It was not possible to create a sequence that met the requirements with only one IHT. Therefore, it was determined that sequences be used with IHTs at all but a few positions. This also provided the interesting notion: where will a sequence segment that is entirely made up of salient points with one key transition without a salient point (the dull point)? It was expected such a sequence might segment at the dull point, as the dull point might 'stand out' since there is only one such dull point. It was not expected that any single IHT would result in a consistent segmentation. Therefore, all IHT IKIs were averaged and compared to the average of all WHT

IKIs (which are the dull points). It was found that IHTs were marginally slower than WHTs, indicating that chunk boundaries (which are expected to be slower) might lay at IHTs. Chunk boundaries are not likely located at the WHTs. However, the difference was small and not significant.

Thus, there was a difference found between IHTs (slower) and WHTs (faster) but it was found to be small and not significant. In total (over all used sequences) there were more IHTs (n=72) than WHTs (n=12) in the sequences used. Even if segmentation occurred at IHTs and never at WHTs, it seems logic that not all IHTs were chosen as a chunking point. It is more likely that only a few IHTs (perhaps just one per sequence) would be chosen as a chunking point. Also it seems likely different participants chose different IHTs as their chunking point as different participants might have different motor memory capacities (Bo & Seidler, 2009). Considering this it seems a few slow IHTs (chunk starting points) are averaged with the other IHTs (faster as they are not chunking points) leading to an overall mean for IHTs that is only slightly slower than the mean for WHT (with no chunking points). This might possibly explain the somewhat ambiguous results.

Pattern potency

Finally, it was determined which of these potential causes for segmentation would be more likely to cause segmentation. This was an open question and it was investigated by comparing the difference between the differences: run-boundary versus no-run, and IHT versus WHT. The size of the differences indicates the potency of the patterns. This means that a larger difference between IKIs with the salient point and IKIs without the salient point indicates that the pattern is more potent.

As discussed above runs did not induce consistent segmentation, on the contrary run boundaries were executed faster than no-run boundaries. Additionally, it was discussed that IHTs might have induced segmentation albeit not consistent. Therefore, I feel I must conclude that IHTs are more potent causes for segmentation than runs (of two consecutive keys).

Concluding while noting that there was little definitive evidence, the results from this experiment seem to indicate that a sequence will not segment at a run or at a dull point. Additionally, it is likely that a sequence will segment at an IHT but not at all IHTs. Finally, an IHT seem to be a more potent cause for segmentation than a run.

Difficulties with the design

Interpretation of the results should be viewed with suspicion due to a possible problem with finger effects. A finger effect is the premise that some fingers might be faster or slower than others and this might interfere with the results. Therefore, it is important to balance a sequence across fingers. Balancing means that the sequence is rotated one key at a time (i.e. placing the last key in front of the next sequence). This yielded 7 versions of each sequence, see table 2. However, there is a problem with the balancing over fingers. The WHT and the run boundaries are always performed by the same few fingers. The WHT is always performed by the left hand and differs between S-D (left ring - left middle) and A-D (left pink - left middle). Run boundaries differ between D-J (left middle - right index), L-S (right ring - left ring), and F-K (left index - right middle). Therefore possible finger effects might have contaminated the results.

These difficulties show the importance of properly counterbalancing fingers. Unfortunately balancing makes for a cubic growth of sequences in the design used here as each key in a sequence requires another sequence to balance for it. Further, not all sequences can be balanced across fingers without compromising some properties of the sequence itself (e.g. a property such as just one run).

Another possible way to combat the finger effects might be using longer sequences made up of triplets (or other number of keys). In such a sequence each triplet is a new condition with properties such as a run or an IHT. This resembles a design used by Koch & Hoffmann (2000a). Even though Koch and Hoffmann did not use finger balancing, this approach would allow for easy finger balancing as only three keys in the triplets have to be balanced. However, it might be wise to balance the order of the triplets themselves and then this approach might even require more sequences: $(n_{triplets} \times 3) \times n_{triplets} = n_{sequences}$.

Finally, the amount of practice of sequences in this experiment might have obscured segmentation. In this experiment participants performed each sequence 160 times. An informal analysis of results from de Kleine (2009) showed clear segmentation of sequences from 150 repetitions. However, most DSP tasks use over 500 sequence repetitions (e.g. Rhodes et al., 2004). Perhaps more sequence repetitions might consolidate the sequence segmentation in consistent sequence or chunks.

Future directions

This paper gives a fairly exhausting view of causes for segmentation in a motor sequence. When investigating chunking, it seems important to consider all the causes for segmentation discussed here. The review of the literature and the discussed causes for segmentation might provide other motor sequence researchers with a hint of what might happen with their sequence before they put the sequence to the test. However, as illustrated by the experiment in this paper consistent segmentation might be quite elusive.

The most important questions that remain are: What is the potency of the other causes for segmentation? And what (if any) is their (inter)relation? Many causes for segmentation were not investigated in this paper. However, it is still interesting if any or all of the suggested causes for segmentation are equally powerful in causing segmentation, i.e. will some causes be preferred over other causes. The literature provided in the introduction does seem to hint in such a direction. Most, if not all, procedural causes are likely to largely influence segmentation such as the prestructured sequence. Participant related causes such as WM size might also be very important. Sequence related causes such as relational patterns in the sequence seem less potent. However, more research is needed to determine their potency and the relation amongst the causes.

In this experiment it seemed that IHTs were more potent than runs. Runs did not seem to induce segmentation. However, it is possible this was an interaction effect (i.e. this combination of causes for segmentation caused no segmentation at run boundaries). In other words, it might be possible that when a run is combined with another cause for segmentation a run will cause segmentation. Also, combined with another cause for segmentation an IHT might possibly not cause segmentation at all. Therefore, more research seems possible to determine such interaction effects between causes for segmentation.

References

Abbott, L.F., & Blum, K.I. (1996). Functional Significance of Long-Term Potentiation for Sequence Learning and Prediction. *Cerebral Cortex, 6,* 406-416.

Agam, Y., Galperin, H., Gold B.J., & Sekuler, R. (2007). Learning to imitate novel motion sequences. *Journal of Vision, 7,* 1-17.

Bo, J. & Seidler, R.D. (2009). Visuospatial Working Memory Capacity Predicts the Organization of Acquired Explici Motor Sequences. *Journal of Neurophysiology, 101,* 3116-3125.

Bullock, D. (2005). Modeling Cortico-Subcortical Interactions During Planning, Learning, and Voluntary Control of Actions. *Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 31.*

Carpenter, G.A. (2001). Neural network models of learning and memory: leading questions and an emerging framework. *Trends in Cognitive Sciences, 5,* 114-118.

Carpenter, G.A., & Grossberg, S. (1987). A Massively Parallel Architecture for a Self-Organizing Neural Pattern Recognition Machine. *Computer vision, graphics, and image processing, 37,* 54-115.

Cleeremans, A., Destrebecqz, A., & Boyer, M. (1998). Implicit learning: news from the front. *Trends in Cognitive Sciences, 2 (10),* 406-416.

Cohen, A., Irvy, R.I., & Keele, S.W. (1990). Attention in Structure in Sequence Learning. *Journal of Experimental Psychology: Learning, Memory, and Cogntion, 16 (1),* 17-30.

De Kleine, E. (2009). *Cognitive control of sequential behavior: Preparing mirrored motor sequences* (Doctoral dissertation, Universiteit Twente, 2009).

De Kleine, E., & Verwey, W.B. (2009). Representations underlying skill in the discrete sequence production task: effect of hand used and hand position. *Psychological Research, 73,* 685–694.

Driskell, J.E., Copper, C., & Moran, A. (1994). Does Mental Practice Enhance Performance? *Journal of Applied Psychology, 79 (4),* 481-492.

Feltz, D.L., & Landers, D.M. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. *Journal of Sport Psychology, 5,* 25-57.

Grossberg, S. (1978). Behavioral contrast in short term memory: Serial binary memory models or parallel continuous memory models? *Journal of Mathematical Psychology, 17,* 199–219.

Howard, J.H., Jr., Mutter, S.A., & Howard, D.V. (1992). Serial Pattern Learning by Event Observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18 (5),* 1029-1039.

Jiménez, L. (2008). Taking patterns for chunks: is there any evidence of chunk learning in continuous serial reaction-time tasks? *Psychological Research, 72,* 387–396.

Jiménez, L., Mendéz, C., & Cleeremans, A. (1996). Comparing Direct and Indirect Measures of Sequence Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22 (4),* 948-969.

Koch, I., & Hoffmann, J. (2000a). Patterns, chunks, and hierarchies in serial reaction-time tasks. *Psychological Research, 63,* 22-35.

Koch, I., & Hoffmann, J. (2000b). The Role of Stimulus-Based Spatial Information in Sequence Learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26 (4),* 863-882.

Lashley, K.S. (1951). The problem of serial order in behavior. In: L.A. Jeffress, Editor, *Cerebral mechanisms in behavior*, Wiley, New York, 112–131.

Meulemans, T., & Van Der Linden, M. (1997). Associative chunk strength in artificial grammar learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 1007–1028

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits in our capacity for processing information. *Psychological Review, 63,* 81-97.

Nissen, M.J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology, 19,* 1-32.

Park, J, & Shea, C.H. (2003). Effect of Practice on Effector independence. *Journal of Motor Behavior, 35 (1),* 33-40.

Perruchet, P., & Pacteau, C. (1990). Synthetic grammar learning: implicit rule abstraction or explicit fragmentary knowledge? *Journal of Experimental Psychology: General, 119,* 264–275.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning: one phenomenon, two approaches. *TRENDS in Cognitive Sciences, 10 (5),* 233-238.

Perruchet, P., & Vinter, A. (1998). PARSER: A Model for Word Segmentation. *Journal of Memory and Language, 39,* 246–263.

Perruchet, P., Vinter, A., Pacteau, C., & Callego, J. (2002). The formation of structurally relevant units in artificial grammar learning. *The Quarterly Journal of Experimental Psychology, 55A (2),* 485–503.

Povel, D., & Collard, R. (1982). Structural factors in patterned finger tapping. *Acta Psychologica, 52,* 107-123.

Rhodes, B.J., Bullock, D., Verwey, W.B., Averbeck, B.B., & Page, M.P.A. (2004). Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science, 23,* 699-746.

Restle, F. (1970). Theory of serial pattern learning: Structural trees. *Psychological Review, 77,* 481–495.

Rice, N.J., & Sekuler, R. (*Under review*). Chunking and compound cueing of movement sequences: Learning, retention and transfer.

Rosenbaum, D. A. (1987). Successive approximations to a model of human motor programming. In G. H. Bower (Ed.). *The psychology of learning and motivation: Advances in research and theory* (Vol. 21, pp. 153–182). San Diego, CA: Academic Press.

Rosenbaum, D.A., Cohen, R.G., Jax, S.A., Weiss, D.J., & van der Wel, R. (2007). The problem of serial order in behavior: Lashley's legacy. *Human Movement Science, 26,* 525–554.

Rosenbaum, D. A., Weber, R. J., Hazelett, W. M., & Hindorff, V. (1986). The parameter remapping effect in human performance: Evidence from tongue twisters and finger fumblers. *Journal of Memory and Language, 25,* 710–725.

Sakai, K., Kitaguchi, K., & Hikosaka, O. (2003). Chunking during visuomotor sequence learning. *Experimental Brain Research, 152,* 229–242.

Salthouse, T. A. (1986). Perceptual, cognitive, and motoric aspects of transcription typing. *Psychological Bulletin, 99,* 303–319.

Servan-Schreiber, E., & Anderson, J.R. (1990). Learning artificial grammars with competitive chunking. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16 (4),* 592-608.

Sternberg, S., Monsell, S., Knoll, R. L., & Wright, C. E. (1978). The latency and duration of rapid movement sequences: Comparisons of speech and typewriting. In G. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 117–152). New York: Academic Press (Reprinted in *Perception and production of fluent speech*, pp. 469–505, by Cole, R.A. (Ed.), 1980, Hillsdale, NJ: Erlbaum).

Stadler, M.A. (1992). Statistical Structure and Implicit Serial Learning. *Journal of Experimental Psychology: Learning. Memory, and Cognition, 18 (2),* 318-327.

Stadler, M.A., & Neely, C.B. (1997). Effects of sequence length and structure on implicit serial learning. *Psychological Research, 60,* 14-23.

Verwey, W.B. (1996). Buffer loading and chunking in sequential keypressing. *Journal of Experimental Psychology: Human Perception & Performance, 22,* 544-562.

Verwey, W.B. (1999). Evidence for a Multistage Model of Practice in a Sequenctial Movement Task. *Journal of Experimental Psychology: Human Perception and Performance, 25 (6),* 1693-1708.

Verwey, W.B. (2001). Concatenating familiar movement sequences: the versatile cognitive processor. *Acta Psychologica, 106,* 69-95.

Verwey, W.B. (2003a). Processing modes and parallel processors in producing familiar keying sequences. *Psychological Research, 67,* 106–122.

Verwey, W.B. (2003b). Effect of Sequence Length on the Execution of Familiar Keying Sequences: Lasting Segmentation and Preparation? *Journal of Motor Behavior, 35 (4),* 343-354.

Verwey, W.B. (in preparation). Probing keying sequences with a tone counting task: preliminaries to a processing framework of movement skill.

Verwey, W.B., Abrahamse, E.L., & Jiménez, L. (2009). Segmentation of short keying sequences does not spontaneously transfer to other sequences. *Human Movement Science.* doi:10.1016/j.humov.2008.10.004

Verwey, W.B., & Clegg, B.A. (2005). Effector dependent sequence learning in the serial RT task. *Psychological Research, 69,* 242–251.

Verwey, W.B., & Eikelboom, T. (2003). Evidence for Lasting Sequence Segmentation in the Discrete Sequence-Production Task. *Journal of Motor Behavior, 35 (2),* 171-183.

Verwey, W.B., Lammens, R., & van Honk, J. (2002). On the role of the SMA in the discrete sequence production task: a TMS study. *Neuropsychologica, 40,* 1268-1276.

Verwey, W.B., & Wright, D.L. (2004). Effector-independent and effector-dependent learning in the discrete sequence production task. *Psychological Research, 68,* 64-70.

Appendix

Internal models

Several internal models will be discussed here; some models are capable of capturing a theory of how a sequence might be chunked (e.g. RTM-model and PARSER), other models give insight into the workings of the sequence learning and production system (e.g. N-STREAMS).

Hierarchical models provide a way to represent how a sequence might be represented or constructed in the brain as demonstrated by Restle (1970) with his RTM-model. The hierarchical model provides the operators *transpose* (T), *repeat* (R) and *mirror* (M), note that the operators are somewhat similar to the relational patterns described above. Given the alphabet (1 2 3 4), the hierarchical model might represent the sequence (1 2) by T(1), sequence (2 2) by R(2) and sequence (1 2 4 3) by M(1 2). Operations can also be applied hierarchically and can be represented by a structural tree (see figure A1). The time it takes to produce the individual responses might be computed by assuming that traversing branches in the tree takes an operation which takes a fixed amount of time. Resulting in traversing 3 operators (T, R and T) for the first item and 1 operator (T) for the second item in the T(R(T(1))) example from figure A1. The timing for the keys of this sequence might be (3x 1x 2x 1x 3x 1x 2x 1x) with x being the time it takes to complete an operation (Povel & Collard, 1982). This hierarchical model requires movement n to trigger movement n + 1, which in turn is necessary for triggering movement n + 2, and so on; a reflex chain (Rosenbaum et al., 2007). A reflex chaining hierarchical model does predict that some items might have longer RTs than others. However, it does not provide insight into the motor chunking of a sequence and is unable to explain why a well learned sequence that consists of several concatenated chunks can be scrambled and still be produced better than a random sequence as long as chunks are respected.

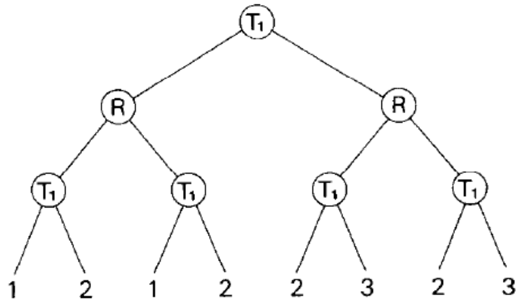

Figure A1: Structural tree for RTM-code T(R(T(1))) producing sequence (1 2 1 2 2 3 2 3) by traversing from left to right through the branches (from Povel & Collard, 1982).

Fortunately for our quest for the causes of segmentation reflex chaining has had an alternative since the influential Lashley (1951) paper; “The problem of serial order in behavior”. Lashley (1951) proposed that the solution to the serial order problem (i.e., the solution to the problem of how behaviors are sequenced) lies in that the nervous system can “predispose the actor to behave in particular ways in the future” (i.e., make plans for behavior), therefore, eliminating the need to ascribe triggers to sensory feedback. Lashley (1951) stressed that plans for behavior are organized hierarchical and since then much evidence for this notion has been found (see for reviews Rhodes et al., 2004; Rosenbaum et al., 2007). One piece of evidence for hierarchical plans that is particularly interesting for us pertains to the fact that the long-term learning of (motor) skills is naturally characterized by the process of forming ever larger hierarchical units or chunks (Miller, 1956; Rhodes et al., 2004; Rosenbaum et al., 2007). In other words, chunks might be seen as hierarchical plans for motor behavior, i.e., chunks are hierarchical motor plans.

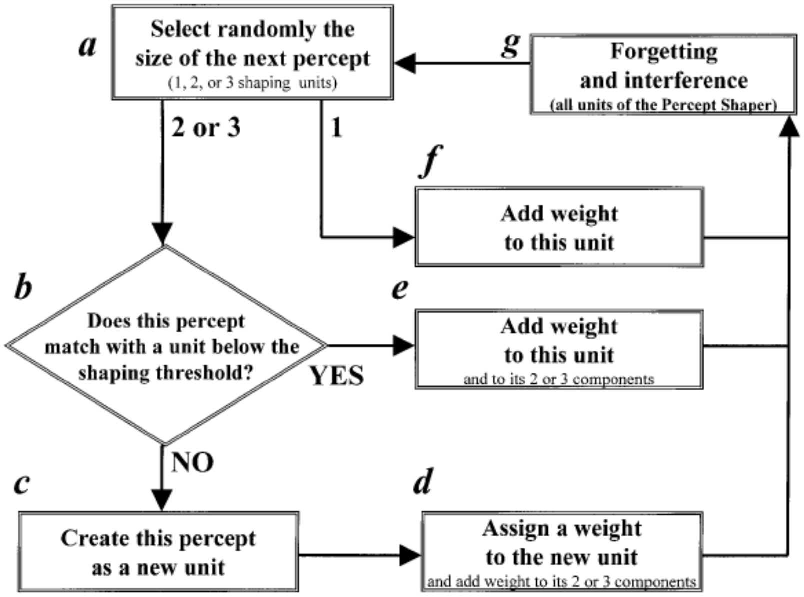

Figure A2: Operations performed by the PARSER model.

Rosenbaum et al. (1986) describe a sequential effect called the *parameter remapping* effect which suggests that hierarchical motor plans are updated rather than replaced when the following motor plan needs to be executed. Parameters that were mapped into a plan for a response persist after the plan has been executed and extra processing is required if a parameter has to be replaced should the following plan require a different parameter. This was supported by their finding that in keyboard sequence typing performance, repetitions of finger tapping sequences suffer if the number of consecutive taps by the same finger changes from cycle to cycle (Rosenbaum et al., 1986). Later Rosenbaum et al. (2007) added that the plan-change method of plan formation is computationally economical. It is presumably easier just to deal with differences than always to build new structures. Finally, the parameter remapping effect might cause the often occurrence of segmentation at irregularities in a sequence since an irregularity such as an inversion causes a thorough revision of the action plan which makes a longer delay seem inevitable. This longer delay might be solidified into a memory chunk with practice in a similar way as external pacing in the practice phase can induce segmentation as discussed earlier (e.g. Verwey, 1996).

A possible way in which chunks might be formed is described by the PARSER model (see figure A2), a model proposed to explain segmentation in artificial language studies (Perruchet & Vinter, 1998). The frequency of stimuli (individual items or subsequences) in the formation of chunks has had attention as many chunking models implement associative learning principles. However, forgetting is due in large part to the interference generated by the prior or subsequent events that are related in some way to the target event (Perruchet & Pacton, 2006). The PARSER model is centered around a single vector called the percept shaper (PS). PS is the internal representation of the displayed material and each representation can carry a weight that can be manipulated. The weight represents the familiarity of the person with the percept. When a new sequence in analyzed percepts are randomly created (step a in figure A2). The percept is checked for familiarity (step b in figure A2) and when it is not recognized it is created (step c in figure A2) or else its weight is increased (step e in figure A2). Eventually this results in frequently co-occurring segments becoming heavy weights (Perruchet & Vinter, 1998). Notice that the PARSER model has some similarities to the statistical redundancy proposed by Stadler (1992) discussed earlier.

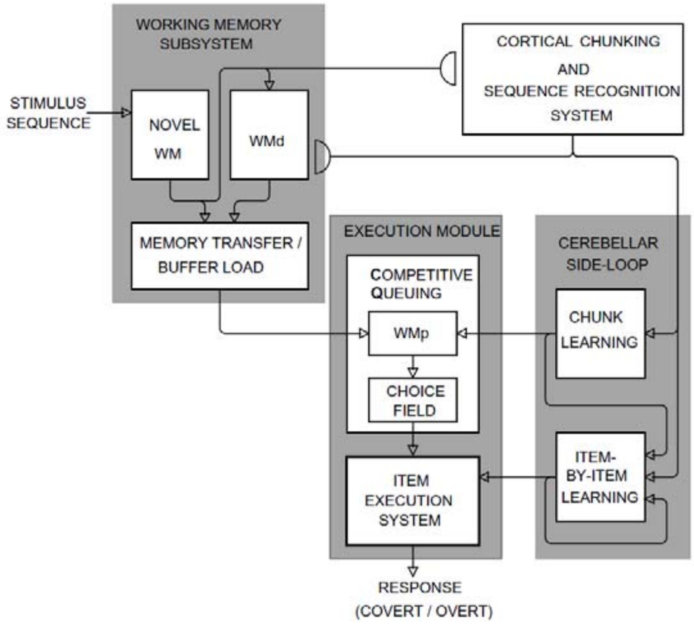

Figure A3: Macrocircuit illustrating the global architecture of the N-STREAMS model. Only major components and links between them are depicted here, but the associated simulation model shows how these major parts can be fully implemented as neural networks. At the core of N-STREAMS is a competitive queuing (CQ) system. This includes a plan working memory (WMp) and a choice field capable of choosing (for performance) the most active remaining plan in WMp. The CQ core is augmented by a declarative WM (WMd) and a closely associated adaptive cortical chunking system, which learns to recognize and recall the parallel sequence representations that pass through the WMd. Sites of learning between WMd and the cortical chunking subsystem are shown as semi-circles. Additional sites of learning are located in the other major augmentation of the CQ core, namely the cerebellar side loop. The upper cerebellar side loop generates a parallel output to the CQ plan layer (WMp). The lower cerebellar side loop illustrates the cerebellum's output-to-input recurrence, which allows the N-STREAMS model to utilize item-to-item associative links as a subsidiary sequencing mechanism (From Rhodes et al., 2004).

As pointed out by Rhodes et al. (2004) the class of parallel sequence production models (e.g. Grossberg, 1978) that are known as competitive queuing (CQ) may inherently describe motor chunking. In a CQ model there are two layers; a parallel planning layer in which more than one plan representation (for example an unpacked motor chunk consisting of several keys) can be active at one time and a competitive choice layer in which the most-active plan representation is chosen by a competition run, a maximum activation finding which is a winner takes all (WTA) process. WTA coding can be produced by a competitive neural network with strong inhibitory connections in which the node receiving the largest total signal suppresses all other activation. When an action is chosen by this competitive choice layer, a strong output signal is generated and the corresponding plan in the planning layer is deleted, leaving the remaining most-active plan to be selected till no plans remain and the complete action plan is produced (Carpenter, 2001; Rhodes et al., 2004).

The name N-STREAMS is an acronym for Neural Substrates That Rehearse, Encode, And Memorize Sequences (see figure A3). It captures the hypothesis that several substrates capable of sequence representation combine their outputs to compose a stream of behavioral outputs. Perceived sequential stimuli are represented in a declarative working memory (WMd) that uses the standard CQ format, similar to the CQ model described above. The order of the stimuli determines the activation of the corresponding activation. Upon preparation to produce an already represented sequence the WMd loads the representation via a buffer loader into the plan working memory (WMp) which serves as the plan layer of the CQ model. Transfer between the two WMs can also be voluntary and serves four functions. First, it transforms the declarative representation into a plan representation. Secondly, after the CQ generated the complete sequence it is empty. The WMd is still loaded enabling the transfer and re-initialization of the WMp with the sequence. Each new transfer affords another rehearsal by the CQ system of whatever sequence is stored in WMd, including any novel sequence not yet stored in long term memory (LTM). Third, each WM transfer generates signals that facilitate the consolidation of the chunk into an unsupervised LTM in a cortical circuit. Finally, every WM transfer also generates signals that guide LTM (supervised) learning in a cerebellar circuit (cf. Carpenter & Grossberg, 1987 for a review of unsupervised and supervised learning). These two types of LTM included in N-STREAMS are complementary. The cortical chunk-learning circuit recruits a new node to serve as a compressed LTM representation, and recognizer, of each novel sequence registered in the WMd. The cerebellar circuit learns both sequence chunks and individual inter-response transitions from the WMp and the item execution system respectively. After learning occurs, the cerebellar mechanism becomes able to anticipate and preempt slower cortical loading of the appropriate gradient into the production buffer as well as to speed up the execution of individual responses within the sequence (Bullock, 2005; Rhodes et al., 2004).

The fact that mental practice appears to be able to increase motor skill (e.g. Driskell et al., 1994) might be explained by N-STREAMS as the transfer of representations between the WMs which is voluntary and the transfer provides a learning signal for the cerebellar module.