Spike-Timing Neuronal Modelling of Forgetting in Immediate Serial Recall

Panagiotis Ioannou
Department of Computing
University of Surrey
Guildford, Surrey
GU2 7XH, UK

Matthew Casey
Pervasive Intelligence Ltd
Stoneleigh, Frimley Road
Ash Vale, Surrey
GU12 5PN, UK

André Grüning
Department of Computing
University of Surrey
Guildford, Surrey
GU2 7XH, UK

Abstract—**Three major and often contradicting hypotheses regarding forgetting in instant sequential recall include the Time-based decay, Temporal distinctiveness and Interference. Here we will explore these hypotheses using the Spike-timing theory of working memory. We represent memory traces as spatio-temporal patterns of spikes called polychronous groups which behave according to short-term and long-term synaptic dynamics. In spite of their contradictions, we show that all three forgetting hypotheses can co-exist - yet each one can have a different degree of influence in the process of forgetting with interference having an immediate impact.**

I. Introduction

Working memory includes a short-term mechanism for temporary storage and manipulation of information, by which memory traces decay over time [1]. There is some debate as to why these memory traces decay [1]. Time-based theories support that memory traces disintegrate due to the passage of time [1], [3], [5], [25], whilst others support the temporal distinctiveness and interference theory, with which it is proposed that forgetting does not depend on the passage of time per se, but on their temporal isolation, and new memory traces that interfere with existing ones [4], [14], [15], [18], [19]. Here we will examine these theories under the spike-timing paradigm of working memory, in which memory traces are represented by spatiotemporal patterns of spike activity, called polychronous groups (PNGs) [12]. This is a suitable technique to use due to the spike-timing aspect of PNGs, as we can have multiple PNGs simultaneously loaded without resulting in an epileptic-like excitation in the network, thus having an advantage over existing models. Our aim is to investigate whether such an approach can shed light over these forgetting hypotheses.

In the spike-timing theory of working memory [24], sensory input is represented by selectively activating a PNG, that is, by stimulating its comprising neurons and thus generating spikes. After the stimulation and the initial spikes, the PNG gets spontaneously reactivated due to the strengthening of their short-term synaptic dynamics. This implementation was based on experimental findings that showed an elevated firing activity at the neurons of macaque monkeys, while conducting working memory tasks [6], [7], [17]. One might argue that this is evidence of time-based decay in forgetting in short-term memory. In this paper we explore forgetting using the spike-timing theory of working memory both in a temporal and non temporal way, and demonstrate how time-based decay, temporal distinctiveness, and interference (definitions addressed later on) can influence the retention of memory traces in short-term memory.

II. Decay, temporal distinctiveness and interference

The hypothesis that short-term memory content decays over time is strongly supported by its limited duration [1]. The most notable theory in support of the decay paradigm is presented through the multiple component model of working memory [1]. This temporal decay theory is often correlated with the notion of quick repetition, that is, to extend the duration of the information that is sustained and overcome its temporal limitations, memory traces are being periodically rehearsed by verbal or mental articulation. This articulatory mechanism forces memories to loop back in the short-term memory inventory and to be sustained for longer.

Moreover, the hypothesis that memories decay rapidly through time, as well as results favouring the rehearsal process, derive from the *word length effect*. This word length effect [1], comes from the observation of the immediate recall of long words, e.g. "amplification", are a lot less correctly recalled than short words e.g. "strong". Another striking phenomenon is the *phonological similarity effect*, by which individuals are less correct in recalling lists of similar sounding words for example "stop", "tall", "crop". In addition, words with similar meaning ("huge", "large", "big", "wide", "tall") do not affect the accuracy of recall. This rehearsal mechanism takes place in the multiple component model of working memory at the phonological loop and visuospatial sketchpad components of the Baddeley and Hitch's model of working memory [1].

In 2004, Lewandowsky, Duncan and Brown proposed a mechanism that stops this repetition process we mentioned above [14]. In their studies, participants undertook a recall task in which they had to recall a sequence of letters, and they had to use the keyboard for their responses while at the same time repeating an irrelevant word ("super") loudly in order to stop the repetition. This mechanism is known as articulatory suppression and stops the phonological loop from the repetition, thus stopping memories from reactivating. They also showed that the articulatory suppression length, that is the

number of repetitions of the irrelevant word, did not influence the recall, supporting that time per se is not a critical factor in forgetting in serial recall.

The temporal distinctiveness theory [8] supports that the temporal separation of memory items at encoding affects the performance of its retrieval. These distinctiveness models support that memories and their recall depend on their distance from all other memories across a time space. More specifically, they show that memory recall strengthens according to their time distance from adjacent memories. For example temporal distinctiveness supports that memories are better remembered when they are very distant than memories that are temporally closer, or, in the context of items in a list, the more distinct the easier they are remembered than the ones which are temporally close [8]. Temporal distinctiveness theory is also in contrast with decay theory, as it does not support the notion of memories that fade because of the passing of time. Furthermore, temporal distinctiveness approaches do not support the repetition system as the one responsible for the sustainability of memory items through a period of time.

In our experiments, we address these temporal and non-temporal approaches to forgetting in short-term memory with the spike-timing theory of working memory. As we mentioned earlier, polychronous groups can be used as a representation of memory traces [12]. When synapses between neurons are subject to short-term STDP dynamics, then they can be sustained, once selectively activated, representing in this way sustainable memory traces that decay through time [24]. These specific experiments are in accordance to the decay paradigm, by which it is hypothesised that memory traces disintegrate as a passage of absolute time [1]. This postulate is supported by experimental findings in which neurons in macaque monkeys had a persistent elevated firing activity during working memory tasks [6], [7], [17]. The sustained reactivation can be presented as the rehearsal, i.e. the verbal reverberation at the phonological loop, whereas the stoppage of the rehearsal process can be achieved through an articulatory suppression mechanism.

Is the decay approach really in contrast with the other theories, i.e. the temporal distinctiveness or interference? How can the spike-timing theory of working memory be put in context with the serial recall in an item list? Here we view the theories from a neuronal level, and show that these theories do not necessarily contradict each other but can actually co-exist, each having a role to play in forgetting in immediate serial recall.

III. METHODOLOGY

We follow the methodology used in [24] and [10]. The network consists of 1000 neurons, 800 excitatory (pyramidal neurons exhibiting regular spiking), 200 inhibitory (GABAergic interneurons exhibiting fast spiking) representing standard neocortical parameters. Excitatory connections have random delays up to 20ms, whereas inhibitory connections have 1ms delays. The connectivity probability is 0.1, and neurons are connected randomly. We use the Izhikevich's simple model of spiking neurons [11].

Like in [24], excitatory towards inhibitory and all the inhibitory connections do not follow plasticity rules (have static strengths). The strengths of the synapses of the input connections to a neuron change with regards to the STDP rule [23]. The synaptic strengthening depends on the arrival of spikes to the postsynaptic neuron: It strengthens if the spike from the presynaptic neuron has arrived at the postsynaptic target before the firing of the postsynaptic neuron, while it is weakened if the spike from the presynaptic neuron has arrived after the postsynaptic target fired. The level of strengthening equals $A_+ e^{-\frac{\Delta t}{\tau}}$ and on the other hand for weakening is $A_- e^{-\frac{\Delta t}{\tau}}$, where Δt is the time difference between the arrival of the presynaptic spike to the postsynaptic neuron and the postsynaptic spike, $\tau = 20$ms, $A_+ = 0.1$, and $A_- = 0.12$. The weights (synaptic strengths) are between $[0 \dots 8]$mV.

As in [24], the synaptic strengthening of excitatory neurons also change according to a baseline within a short timeframe. According to short-term STDP, input to neuron i at time t, $I_i(t)$, equals $\sum_{j \in J} s_{ij}(1 + sd_{ij})$, where s_{ij} is the synaptic weight of the analogous connection between neuron j and i, and J refers to the group of presynaptic neurons whose spike arrived at neuron i at time t. The weight increases or decreases according to the factor sd, a variable which varies for each synapse according to the STDP rule with the parameters A_+, A_- as above and without synaptic input it decays back to 0 within 5 seconds. Consequently, when there is no input the synaptic efficacy stays the same, pre-then-post spikes short-term strengthens the synapses, and post-then-pre spikes short-term weakens them.

To begin with, we run the simulations for a period of time to find emerging PNGs as described in [12]. To achieve persistent reactivation (representing the reverberation of the memory trace), we select one PNG, and we activate the neurons of a PNG in order according to their polychronous (spike-timing) format at 100ms intervals during a 1 second interval to short-term strengthen their synaptic weights. We measure the quality of a PNG when interpreted as an item in working memory by: 1. their *strength*, and 2. their *duration*, as defined in [10]. We measure strength as the area under curve (AUC) of a PNG activation rate plot, that is, the percentage of PNG neurons (per second) which are activated over the total number of neurons in a PNG through time. We measure the duration of a PNG as the time difference between the start of the stimulation where the activation frequency increases until drops it back down to a baseline (2Hz).

PNG quantity (how many they can emerge) in a network with regard to various network parameters was explored in [9] and the factors influencing PNG retention by time-based decay was explored in [10]. Here we further explore time-based decay as we run different simulations to see how a PNG size can affect its sustainability. Then we explore temporal distinctiveness theory by increasing the temporal distance between the stimulation of 2 PNGs. Finally we explore interference theory by stimulating again 2 PNGs at fixed temporal
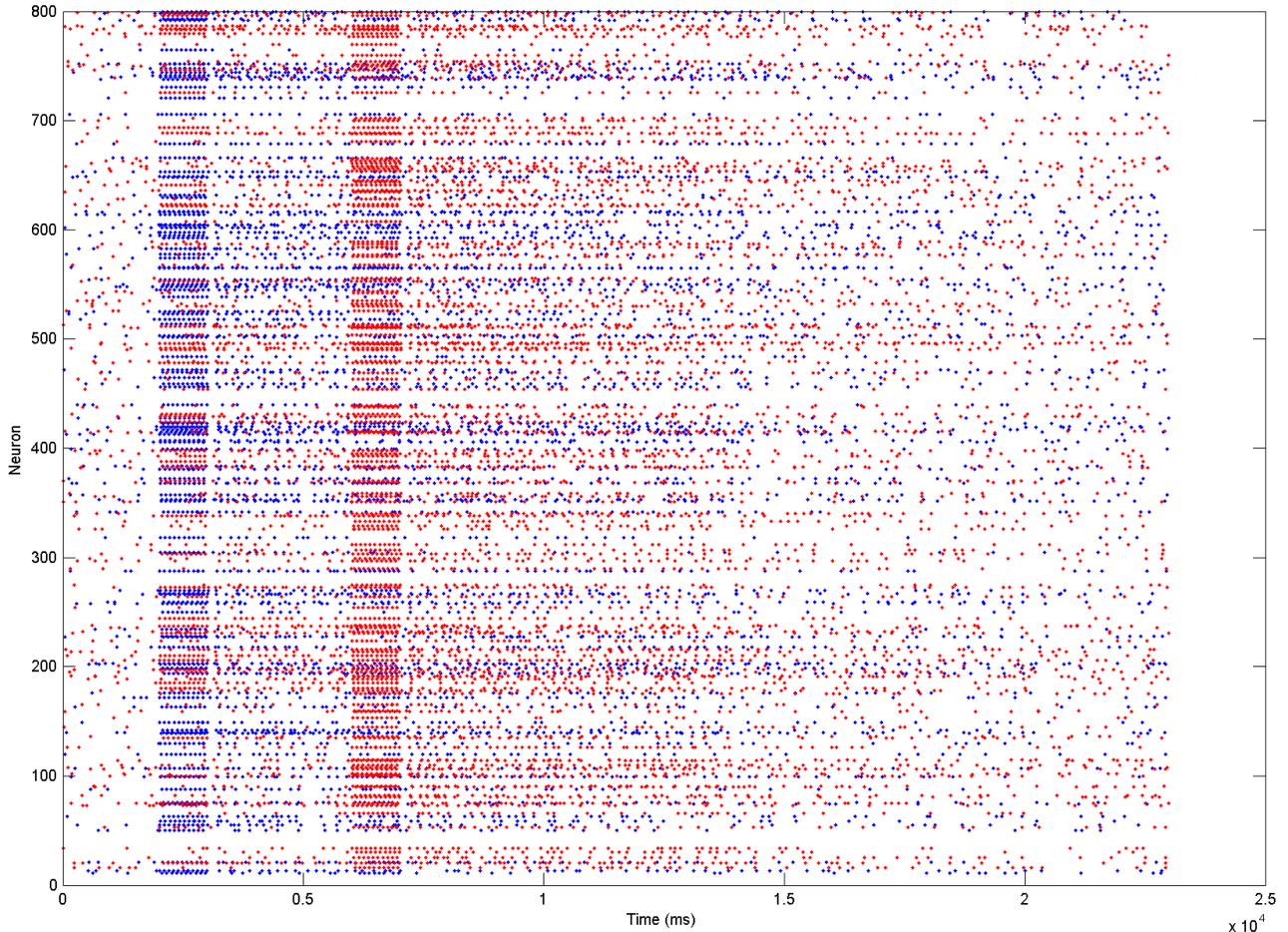
Fig. 1: Firing plot of a simulation in which we stimulate 2 PNGs. Blue denotes PNG (A), which is stimulated at the 2nd second, and red denotes PNG (B) which is stimulated at the 6th second. The 2 groups share 40 neurons. The shared neurons of PNG (B) interfere with the sustained reactivation of PNG (A) in two ways: 1) PNG (B)'s shared neurons fire at a different temporal pattern than the one that constitutes PNG (A). 2) They share neurons and in addition exhibit the same spatiotemporal pattern.

differences, only this time increase their percentage of their shared neurons. A raster plot of 2 PNGs activated in the same network can be seen in figure 1 and their firing rate plot in figure 2. The parameter table can be viewed in I.

Network description	1000 neurons, 800 excitatory (regular spiking), 200 inhibitory (fast spiking). 100 connections per neuron. 20ms maximum axonal conduction delays. Short term and long term STDP. 100 neuron PNGs.
Time based decay	Stimulated PNG size: 10-150 step of 10
Temporal distinctiveness	Time difference between stimulation of 2nd PNG: 3-10 step of 1
Interference	Percentage of shared neurons of 2nd PNG: 20-40 step of 5

TABLE I: Model parameters

IV. EXPERIMENTS AND EVALUATION

A. Time based decay

The word-length effect supports that series of long words are more difficult to be remembered than those with smaller words [1] and in extent, series of polysyllabic words are more difficult to be remembered than monosyllabic words. However, it is argued that the reason behind this is the phonological complexity of those words [22]. In [22] it was proposed that phonological complexity and articulation duration are the critical variables when the two factors are disentangled.

Another word-length effect in support of the decay theory is derived when repeating words with the same syllables and phonemes. It was shown that short phoneme words were better remembered than the long phoneme words [2]. On the other hand, others argue that these observations especially in disyllable words could have been created by chance because of coincidental word selection [16].
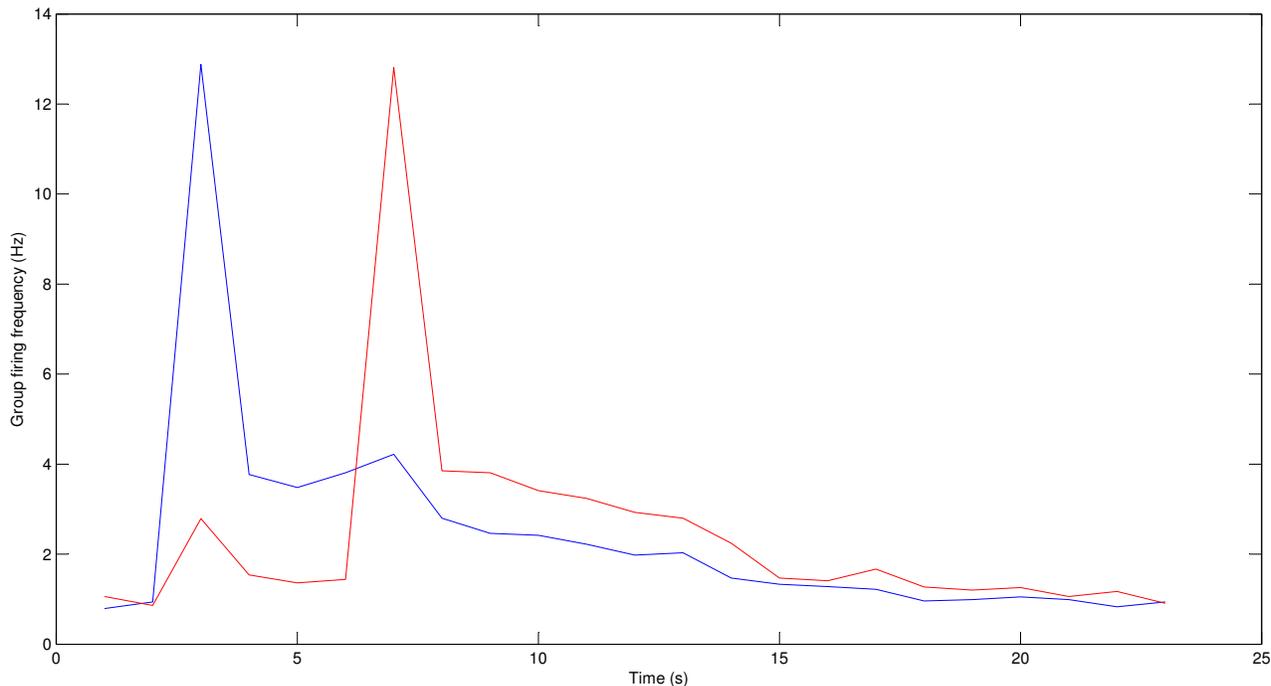
Fig. 2: Activation rate of two polychronous groups (PNGs). Group A (blue) gets selectively activated 10 times per 100ms during the 2nd second. Group B (red) gets selectively activated 10 times per 100ms intervals during the 6th second. Due to the overlap of the groups (40 out of 100 neurons), the stimulation of group A affects the activation rate of group B and vice versa. This is shown at the 2nd second where activation rate of group B increases as a result of stimulation of group A, and at the 6th second stimulation of group B increases the activation rate of group A.

The above contradicting findings have motivated us to explore time-based decay and the word-length effect from a microscopic point of view in the form of neuronal simulations consisting of biologically realistic spiking models. According to short-term dynamics implemented in this theory, these PNGs can be stimulated and their firing activity spontaneously sustained even after the stimulation offset. This sustained activity disintegrates back to its baseline in a few seconds.

Clearly this theory is in support of the time-based decay theory. We will further explore this theory in relation to the word-length effect. Since PNGs represented memory cues/items in the original model, here we represent the length of words and length differences as neuronal size differences of PNGs. For instance short words are represented by PNGs consisting of fewer neurons, versus long words represented by PNGs with more neurons.

For our first set of experiments, in order to explore the word-length effect we selectively activate groups of different sizes (10 to 150 in a step of 10) keeping all other parameters fixed and following the methodology in [24] and summarized in our method section. We observe (see Fig. 3) that the group's size greatly affects the sustainability of their firing rate, both in terms of its percentage of activation as well its duration. More specifically, we get very small values for small groups of size 10, and then their values increase linearly until group size 60, after which their sustainability surprisingly reaches a plateau and stabilizes for group size 60 and above. This can also be quantified with an analysis of variance (ANOVA) resulting in a very small p value for strength ($p = 0.05568e^{-11}$) and duration ($p = 1.13926e^{-11}$), indicating that the differences between column means are significant.

Our results show that larger PNGs are more efficiently sustained both in terms of strength and duration. In the word-length effect and specifically in the case of a long word, each syllable might be represented by a finite subset of neurons out of a total population. We note the possibility that the more
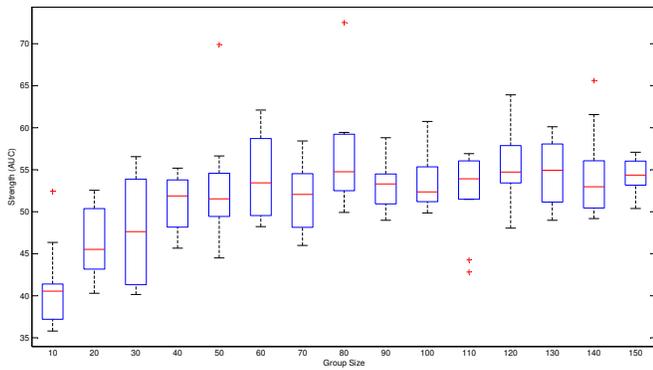
(a) Strength
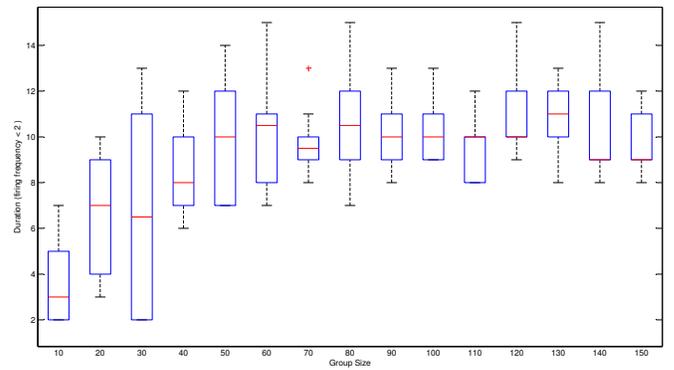
(b) Duration

Fig. 3: Sustainability of polychronous groups of various sizes. We notice that size affects sustainability, both in strength and duration, as we have very small values for size 10 groups, then gradually increasing for size 20, 30, 40, and 50, reaching to a plateau for groups size 60 and above.

syllables the larger subset of neurons will be required in order to represent them and thus less neurons can be assigned to each population. What is evident though, is that if one has to deal with a small word, then one can use more of the available resources to represent it. In this case more neurons and larger PNGs, thus leading to a more efficient sustainability. Or from a different perspective, more syllables can produce larger and more complicated patterns that can produce more interference between them confusing their spatiotemporal sequence as we will explore next. In either case, our results indicate that larger PNGs are better sustained.

B. Temporal distinctiveness

Temporal distinctiveness theory was explored in [8] in the form of retrieving cued memory items. According to this theory, during the presentation of a list, the memories that are more time distinct from their neighbouring items are remembered better. On the other hand however, investigations in [13] showed that time does not influence encoding and showed that temporally isolating memories does not affect retrieval, although they mention that this could increase the consolidation, rehearsal or grouping or the articulation reverberation mechanism.

Here we will explore the temporal distinctiveness again in the form of PNGs (representing memory items) and the spike-timing theory of working memory. More specifically, at first instance we selectively activate two different PNGs of the same size and in the same network but at a specific temporal distance (3 seconds). That is we activate group A at time $t = 2$ seconds and then activate group B at time $t = 6$ and we measure the sustainability of group A as we described its quantification earlier. Then we run a series of simulations, and each time we gradually increase the gap by adding 1 second

in their temporal distance. Thus the temporal distance changes to $t = 3, t = 4. \ldots t = 10$ seconds.

Results indicate (see Fig. 4) that in our simulation setup, increasing the temporal distance does not influence sustainability decay. This can be also be quantified after an ANOVA resulting in moderate p values for strength ($p = 0.8987$) and duration ($p = 0.6169$), indicating that the differences between column means are not significant.

In our model we explore decay per se, as we limit the selective activation into a 1 second interval, and yes, our results are in conjunction with articulatory suppression experiments [14] which suggested that time is not a critical factor on item recall. Nonetheless, we need to keep in mind that increasing the temporal distance can increase the stimulation time induced by the articulatory rehearsal (i.e. in these experiments expanding the selective activation beyond the 1 second interval), and thus further increasing the intra PNG's synaptic dynamics.

C. Interference

Authors favouring time-based decay as a means of forgetting in working memory include [3], [5], [21], [25]. However, [4], [14], [15], [18], [19] doubt time-based decay and support some form of interference to forgetting. When a series of memory items are stored in short-term memory at the same time, they might compete or degrade each other. So in a way new memory items push the old ones, unless the old ones are specifically focused upon.

Here we will explore interference from a neuronal perspective again, using PNGs as memory item representations that are subject to short-term synaptic dynamics as proposed in the spike-timing theory of working memory. The central assumption in our neuronal model is that memory item A, can be represented by a PNG (A), and can be selectively
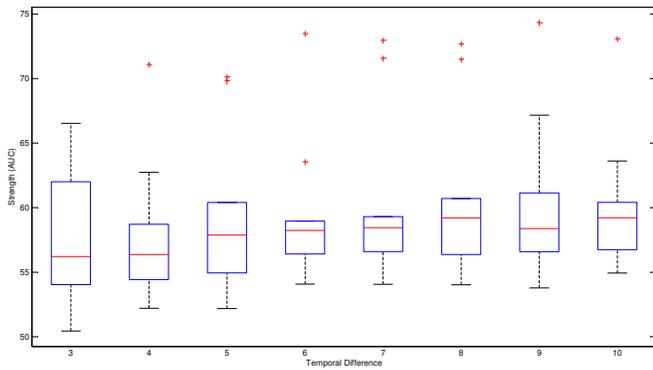
(a) Strength
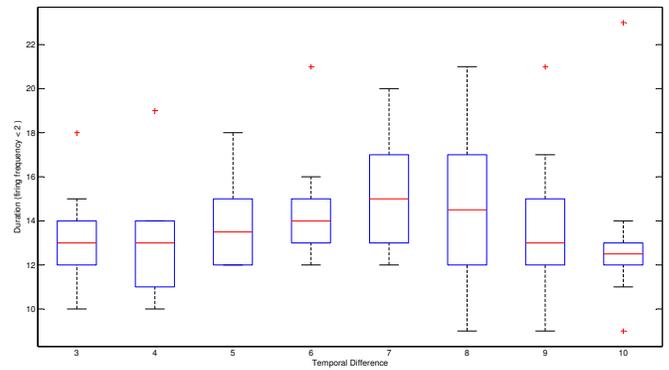
(b) Duration

Fig. 4: Sustainability of polychronous group A when we vary the temporal distinctiveness between the activation of group A and group B. We notice that sustainability strength slightly increases as the temporal window increases. The duration increases for distances 5 to 7 seconds reaching maximum at 7 seconds, then gradually decreases.
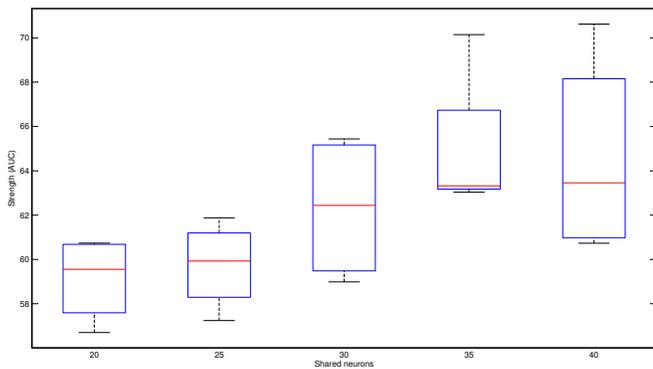

(a) Strength
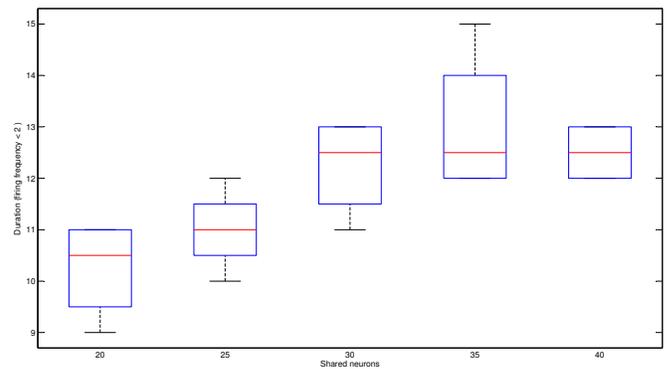
(b) Duration

Fig. 5: Sustainability of polychronous group A when we vary the percentage of the shared neurons between group A and group B. We notice that the percentage of shared neurons has a great effect on the sustainability strength as well as the sustainability duration. Strength and duration gradually increase from neurons 20% to 30%, reaching a plateau from 35% to 40%. However, since this sustainability only refers to an activation rate, the quality of the group might degrade as a result of the different temporal pattern of group B's neurons (see figure 1.)

activated and sustained using the method described in [24]. Then a second item (B) represented by a PNG (B) can be activated at a specific temporal distance from PNG (A). To measure interference, we quantify PNG (A)'s sustainability as an expression of its persistent activity as we described earlier. As we mentioned above, dissimilarity between items in a serial recall list favours their recall, so we move one step further and quantify this similarity as the percentage of shared neurons between PNG (A) and PNG (B). Our assumption is that the more neurons they share the more similar the items. Our assumption is in line with the shared features interference model used in [19].

Our results (see Fig. 5) indicate that interference is a critical factor to the sustainability of a PNG. This can also be quantified with an analysis of variance (ANOVA) resulting in a very small p value for strength ($p = 0.0666$) and duration

($p = 0.007$), indicating that the differences between column means are significant.

This is easily observed, as the activation of PNG (B) has a concurrent effect to the synaptic dynamics of its shared neurons with PNG (A), directly altering their firing activity. What is interesting however, is that while it seems that it strengthens/increases the activity of the neurons that belong to PNG (A), they may not follow the same time pattern of spike firing, and thus altering what constitutes a PNG (a PNG is not a set of neuronal firings per se, but a specific spatiotemporal pattern of spikes) (see Fig. 1). For instance a shared neuron (number 388 on figure 1) belongs to PNG (A) if it fires at time 35ms and to PNG (B) if it fires at 93ms. So activation of PNG (B) changes the firing time of that neuron. On the other hand, if the shared neurons have the same spatiotemporal pattern, then activation of PNG (B) can concurrently strengthen the shared spatiotemporal pattern of PNG (A) (see Fig. 1). Again this does not mean that it contradicts the interference and dissimilarity phenomena, because at the point where the shared spatiotemporal pattern ends, the pattern might not follow the pattern of PNG (A), but will follow instead the spatiotemporal pattern of PNG (B) whose synaptic dynamics were more recently altered. This could explain the dissimilarity phenomenon: let us say that the PNG (A) and PNG (B) do not share neurons at all, then when one attempts to retrieve PNG (A), its spatiotemporal pattern will be activated easily, even by stimulating just some anchor neurons of that pattern, as that pattern would not have been altered by the activation of PNG (B).

V. Conclusion

The exact mechanism that causes forgetting stills remains elusive [20]. The three fundamental hypotheses of time-based decay, temporal distinctiveness and interference have been studied in a number of experiments and models, where some results argue the validity of others. Here we explored the concept of forgetting in immediate serial recall from a neuronal point of view, in the form of neural network simulations where items are represented by specific spatiotemporal patterns of spikes (PNGs). Our results show that activation of a PNG can strengthen its synaptic dynamics and retain its firing pattern and disintegrate back to its baseline in full conjunction with the time-based approach. In addition we have shown that our model can give an explanation to the word-length effect, as small words can be represented by larger PNGs, which can then be more efficiently sustained. Our temporal distinctiveness models have not favoured the homonymous forgetting hypothesis, as an activation of a second PNG (B) did not impact the sustainability of PNG (A) enough to suggest criticality. We do however note that temporal difference between representations can affect the phonological loop/reverberation mechanism and thus the selective reactivation of a PNG, this way further strengthening its synaptic dynamics. Our interference models have shown how critical the interference of a second PNG is on the sustainability of the first, and how this criticality can be expressed as the interference between

their spatiotemporal patterns of firing activity. Our results suggest that interference can have an immediate and dominant effect on the recall mechanism.

We conclude that according to our models, both temporal and non-temporal approaches to forgetting and their associated phenomena can co-exist in a microscopic neuronal perspective and the existing spike-timing theory of WM. However, it is a matter of degree of influence that each approach can have on the retention and recall of items in a list, and according to our results, interference has an immediate and direct influence.

References

[1] Alan Baddeley and Graham J. Hitch. Working memory. *Scholarpedia*, 5(2):3015, 2010.
[2] Alan D Baddeley, Neil Thomson, and Mary Buchanan. Word length and the structure of short-term memory. *Journal of verbal learning and verbal behavior*, 14(6):575–589, 1975.
[3] Pierre Barrouillet, Sophie Bernardin, and Valérie Camos. Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1):83, 2004.
[4] Matthew M Botvinick and David C Plaut. Short-term memory for serial order: a recurrent neural network model. *Psychological review*, 113(2):201, 2006.
[5] Neil Burgess and Graham J Hitch. A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, 55(4):627–652, 2006.
[6] S. Funahashi, C. J. Bruce, and P. S. Goldman-Rakic. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *Journal of Neurophysiology*, 61(2):331–349, 1989.
[7] Joaquin M Fuster, Garrett E Alexander, et al. Neuron activity related to short-term memory. *Science*, 173(3997):652–654, 1971.
[8] Arthur M Glenberg and Naomi G Swanson. A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1):3, 1986.
[9] Panagiotis Ioannou, Matthew Casey, and André Grüning. Evaluating the effect of spiking network parameters on polychronization. In *Artificial Neural Networks and Machine Learning–ICANN 2012*, pages 255–263. Springer, 2012.
[10] Panagiotis Ioannou, Matthew Casey, and André Grüning. Factors influencing polychronous group sustainability as a model of working memory. In *Artificial Neural Networks and Machine Learning–ICANN 2014*, pages 723–731. Springer, 2014.
[11] Eugene M Izhikevich. Simple model of spiking neurons. *IEEE Transactions on neural networks*, 14(6):1569–1572, 2003.
[12] Eugene M Izhikevich. Polychronization: computation with spikes. *Neural computation*, 18(2):245–282, 2006.
[13] Stephan Lewandowsky, Gordon DA Brown, Tarryn Wright, and Lisa M Nimmo. Timeless memory: Evidence against temporal distinctiveness models of short-term memory for serial order. *Journal of Memory and Language*, 54(1):20–38, 2006.
[14] Stephan Lewandowsky, Matthew Duncan, and Gordon DA Brown. Time does not cause forgetting in short-term serial recall. *Psychonomic Bulletin & Review*, 11(5):771–790, 2004.
[15] Stephan Lewandowsky, Klaus Oberauer, and Gordon DA Brown. No temporal decay in verbal short-term memory. *Trends in cognitive sciences*, 13(3):120–126, 2009.
[16] Peter Lovatt, Steve E Avons, and Jackie Masterson. The word-length effect and disyllabic words. *The Quarterly Journal of Experimental Psychology: Section A*, 53(1):1–22, 2000.
[17] Yasushi Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335(6193):817–820, 1988.
[18] James S Nairne. Remembering over the short-term: The case against the standard model. *Annual review of psychology*, 53(1):53–81, 2002.

[19] Klaus Oberauer and Reinhold Kliegl. A formal model of capacity limits in working memory. *Journal of Memory and Language*, 55(4):601–626, 2006.

[20] Klaus Oberauer and Stephan Lewandowsky. Forgetting in immediate serial recall: Decay, temporal distinctiveness, or interference? *Psychological Review*, 115(3):544, 2008.

[21] Michael Page and Dennis Norris. The primacy model: a new model of immediate serial recall. *Psychological review*, 105(4):761, 1998.

[22] Elisabet Service. The effect of word length on immediate serial recall depends on phonological complexity, not articulatory duration. *The Quarterly Journal of Experimental Psychology: Section A*, 51(2):283–304, 1998.

[23] Jesper Sjöström and Wulfram Gerstner. Spike-timing dependent plasticity. *Scholarpedia*, 5(2):1362, 2010.

[24] Botond Szatmáry and Eugene M Izhikevich. Spike-timing theory of working memory. *PLoS computational biology*, 6(8):e1000879, 2010.

[25] John N Towse, Graham J Hitch, and Una Hutton. On the interpretation of working memory span in adults. *Memory & Cognition*, 28(3):341–348, 2000.