Dendrites support formation and reactivation of sequential memories through Hebbian plasticity

Alessio Quaresima[1,*], Hartmut Fitz[1,2], Renato Duarte[2], Peter Hagoort[1,2], and Karl Magnus Petersson[1,2,3]

[1] *Neurobiology of Language Department, Max Planck Institute for Psycholinguistics, 6565 XD Nijmegen, the Netherlands*

[2] *Donders Institute for Brain, Cognition and Behaviour, 6500 HE Nijmegen, the Netherlands*

[3] *Faculty of Medicine and Biomedical Sciences, University of Algarve, 8000–810 Faro, Portugal*

Keywords: dendritic computation, sequence recognition, network with dendrites, NMDA spikes, word recognition

Abstract

Storage and retrieval of sequences require memory that is sensitive to the temporal order of features. For example, in human language, words that are stored in long-term memory are retrieved based on the order of phonemes ("/b/æ/t/" versus "/t/æ/b/"). It is unknown whether Hebbian learning supports the formation of memories structured in time. Here, we investigated whether word-like memories can emerge in a network of neurons with dendritic structures. Dendrites provide neuronal processing memory on the order of 100 ms and have been implicated in structured memory formation. We compared a network of neurons with dendrites and two networks of point neurons that has previously been shown to acquire stable long-term memories and process sequential information. The networks implemented voltage-based spike-timing dependent plasticity (STDP) and were homeostatically balanced with inhibitory STDP. In the learning phase, networks were exposed to phonemes and words represented as overlapping cell populations. In the retrieval phase, networks only received phoneme sequences as input, and we measured the reactivation of the corresponding word populations. The dendritic network correctly reactivated the word populations with a success rate of 90%, including words composed of the same phonemes in different order. The networks of point neurons reactivated only words that contained phonemes unique to that word and confused words with shared phonemes (success rate below 20%). These results indicate that the slow timescale and non-linearity of dendritic depolarization allowed dendrites toX establish connections between neural groups that were sensitive to serial order. In addition, inhibitory STDP prevented the potentiation of connections between unrelated neuronal groups during learning. During retrieval it reduced activation of words with the same phonemes in a different order. Thus,

the model shows that the inclusion of dendrites enables the encoding of temporal relations into associative memories.

*Corresponding author: alessio.quaresima@mpi.nl

Introduction

Speech perception results from the integration of continuous streams of acoustic stimuli within the language knowledge of the listener. Explaining how it emerges from the neural activity underlying it remains a challenge in brain sciences. The difficulties arise from the lack of a robust theory of information storage that accounts for the formation and retrieval of word memories on the behavioral timescale of language processing (Poeppel, 2012; Poeppel & Idsardi, 2022). Following Pulvermuller (1999), we propose that the neural substrate of word memories is the cell assembly, and the association between phonemes and word forms emerges through Hebbian plasticity.

Formation and retrieval of word memories

A cell assembly is a functional neuronal population emerging from the cells' repeated co-activation (Hebb, 1949). The assembly, usually intended to be composed of excitatory neurons, has spatiotemporal dynamics, with the cells being part of it firing at specific intervals in time (Buzsáki, 2010). Cell assemblies are the core unit of memory (Poo *et al.*, 2016). Their reciprocal synaptic strengthening leads to the formation of synaptic engrams, interconnected subnetworks. Such engrams are carved in the neuronal tissue through experience, following associative plasticity (Langille & Brown, 2018; Dringenberg, 2020).

According to the cell assembly view, during sound perception, the recognition of a familiar stimulus prompts the mutual reactivation of groups of auditory neurons and, possibly, generates a burst of activity (Sakurai *et al.*, 2018). Reliable cell assembly activity is widely observed in the experimental literature on human and non-human animals (Almeida-Filho *et al.*, 2014; Cohen *et al.*, 2020; Hemberger *et al.*, 2019); a critical review of the phenomenon is presented in Langille & Gallistel (2020). Among the evidence on assemblies activity, we include two studies achieved with intracortical recordings in humans. First, Chan *et al.* (2014) observed bursts of neural activity in localized populations of the temporal cortex upon acoustic presentation of single words; the same neural markers were missing when the stimulus was played backward. Second, Vaz *et al.* (2020) isolated neuronal bursts in the middle temporal gyrus during recall of spoken word associations; the neurons reactivated in sequence during correct recollection, not otherwise. Taken together, the studies suggest that (1) access to word memories entails coordinated cell assemblies firing in speech areas, (2) the neuronal ensembles reactivated are sensitive to the order of the presented stimulus, on the time scale of $100\,\mathrm{ms}$ to $300\,\mathrm{ms}$.

However, the robust experimental observations on sequential firing do not match a sound theoretical understanding of how assemblies become sensitive to time-structured stimuli. On the

one hand, biologically constrained models (BCMs) of cortical networks confirm that robust engram memories can be instilled in the network connectivity via an associative protocol consisting of repeated presentation of the stimuli (Litwin-Kumar & Doiron, 2014; Zenke *et al.*, 2015; Tomasello *et al.*, 2018), which provides adequate causal explanations for the evidence in (1). On the other hand, whether cell assemblies can support the recognition of structured temporal patterns, as in (2), remains unanswered and computational evidence is contradictory or partial; *the role played by excitatory plasticity only provides a marginal advantage compared to static networks [...]* in the recognition of sequential stimuli (Duarte & Morrison, 2014). It remains unclear if the shortcomings of BCMs concerning sequence recognition are due to the limits in the underlying hypothesis of associative memories (Gallistel, 2021), or rather to ineffective models of synaptic plasticity, which prevent the network from reaching the proper engrams configuration for sequence detection (Shouval, 2011; Scott & Frank, 2023). As a matter of fact, the STDP rule has the potential to encode temporal structure (Clopath *et al.*, 2010; Fiete *et al.*, 2010), but its implementation in concert with cell assemblies has been only demonstrated to produce sequences rather than recognize them (Maes *et al.*, 2020; Cone & Shouval, 2021). In other words, network models can speak but are hard of hearing.

A third option is that the models used to investigate the task were missing some necessary computational primitives that are present in cortical networks. For example, a mechanism to carry information forward in time on the time scale of the stimulus to be recognized (Chaudhuri & Fiete, 2016). There are three main proposals for short-term memory in computational literature; persistent activity (Wang, 1999; Papoutsi *et al.*, 2014; Wang, 2021), synaptic short-term memory (Mongillo *et al.*, 2008), and neuronal memory (Fitz *et al.*, 2020; Salaj *et al.*). None of these memory mechanisms have been shown to interact with the assembly structure proficiently nor to satisfy the requirements of sequence recognition on the time scale of human words.

We hypothesize that the problem of memory is the central one in the word recognition task at hand; with appropriate short-term memory, the STDP rule should be sufficient to encode time-structure sensitivity in the network's engrams. Thus, we propose to leverage the dendritic memory mechanism offered by Tripod neurons (Quaresima *et al.*, 2022) to maintain depolarized states that bind the sequences of incoming phonemes. The idea that dendrites may determine the computational capabilities of neuronal networks has been increasingly explored in the last decades (Poirazi *et al.*, 2003; Kastellakis *et al.*, 2016; Ujfalussy *et al.*, 2018; Jones & Kording, 2021) however, it is not understood if explicit modeling of dendritic processes can enrich the computational capabilities of neural networks (*are dendrites conceptually useful?*, Larkum (2022)). The present study sought an answer in the domain of sequence recognition and proved that minimal reduction of the dendritic arborization deeply impacts the dynamics of the network.

The Tripod network model and the sequence recognition task

Dendritic neurons were modeled as Tripod neurons (Quaresima *et al.*, 2022), three-compartment models with two dendritic compartments and a soma. The dendritic compartments are endowed with NMDA receptors and voltage-gated channels that allow for dendritic memory in the order of hundred milliseconds (Fig.1*A*).

The network is composed of 2000 Tripod excitatory neurons, 175 fast, and 325 slow inhibitory single-compartment neurons (I_1, I_2). Within the network, each neuron was connected to 20% of the remaining neurons, independently from the population. The connections between Tripod neurons were located on the dendritic compartments (Fig.1*B*) and subject to voltage-dependent spike-time-dependent plasticity (vSTDP). Excitatory plasticity supports the formation of engrams of cell assemblies by strengthening the post-synaptic connections of neurons that fire on depolarized compartments. Because of the formation of cell assemblies, the network is prone to run-away activity. To counteract the destabilization caused by v-STDP we introduced a homeostatic mechanism that maintains the total incoming synaptic weight constant and an additional rapid compensatory mechanism through inhibitory spike-timing-dependent plasticity (iSTDP). The latter is implemented via plastic connections between I_1 neurons and the soma of the excitatory cells. In addition, to leverage dendritic computation and control the non-linearity of the dendritic compartment, we introduced voltage-dependent iSTDP between I_2 and the targeted dendritic compartments (Fig.1*C*). The network is stimulated with background noise, which maintains the cells depolarized and favors the onset of spontaneous activity. In addition, we introduce external projections to represent linguistic signs. Phonemes and word projections connect pre-synaptically to a subset of the excitatory neurons and stimulate the dendritic compartments with intense pre-synaptic firing (Fig.1*D*).

The goal of the network model is to create stable associations between phonemes and words that allow the retrieval of the correct word population upon the phonemic stimuli. The associations are formed during the early phase of the network simulation when words are presented in sequence, interleaved by a brief silence. For each word, the network receives strong stimulation on the word projections and the corresponding phonemes in order. The projections cause simultaneous firing in the assemblies and, through vSTDP, lead to the formation of overlapping engrams among the co-active cells. After, during the recall phase, the network receives stimulation only on the phonemes population while the word projections are silenced. We measure the re-activation of the neuronal population associated to word memories, and compare their dynamics in the dendritic network with the equivalent in two point-neuron models. The present work argues that for the formation and maintenance of phonemes-word associations, the presence of dendrites and local regenerative, non-linear events (NMDA spikes) is helpful, if not necessary.

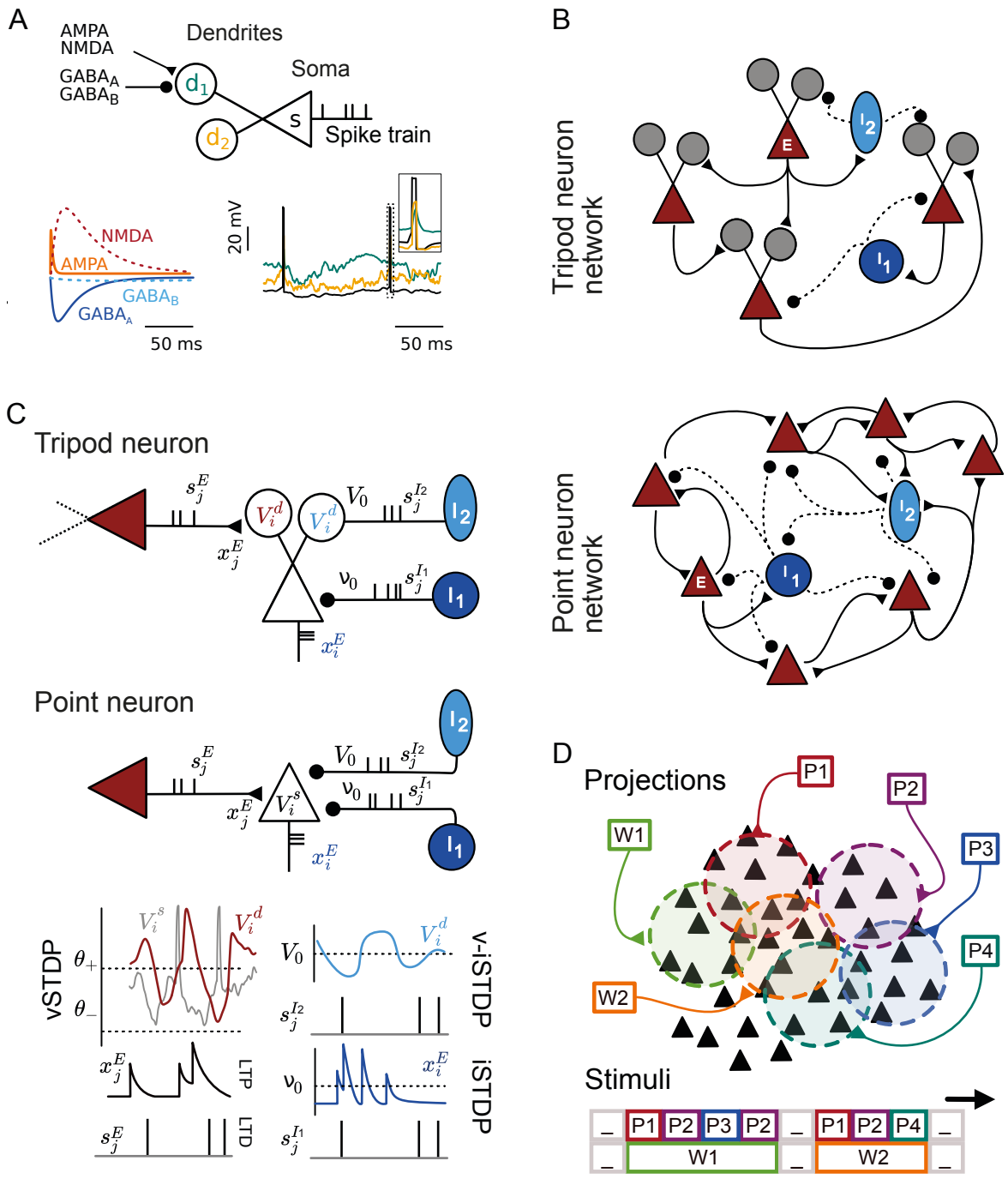

Figure 1: **Tripod network: neuron models, connectivity, and plasticity rules.**

Figure 1: **Tripod network: neuron models, connectivity, and plasticity rules.**

(**A**) Tripod neuron model with synaptic timescales and membrane dynamics. The Tripod neuron has two dendritic compartments (circles) and a soma (triangle). The dendritic length determines the electrical properties of the compartment. The four synapses contain glutamatergic (AMPA, NMDA) and gabaergic ($GABA_A$, $GABA_B$) receptors, whose timescales are illustrated in the bottom left panel (Table3). The membrane potential dynamics of the Tripod evolves accordingly to the AdEx model (bottom right panel) and the somatic spikes are backpropagated to dendrites, as shown in the zoom-in.

(**B**) Within the Tripod network recurrent connections, the Tripod's axon targets other excitatory neurons' dendritic compartments and inhibitory neurons' soma. The inhibitory neurons are single-compartments and belong to two classes, I_1(fast-spiking) and I_2(adaptive interneuron). They target respectively the soma and the dendrites of the Tripod neuron. They also form recurrent connections within the inhibitory populations. In the case of the soma-only neuron, all synapses connect to the somatic compartment.

(**C**) Schematics of excitatory and inhibitory synapses plasticity. Left side; the glutamatergic synapses between excitatory neurons undergo voltage-based STDP. The plasticity rules are the same in the dendritic and somatic compartments, except for the thresholds of long-term potentiation (LTP) and depression (LTD), respectively θ^+ and θ^-. Right side; the inhibitory synapses onto the excitatory cells are subject to iSTDP. The equilibrium values depend on the neuron types. I_2cells aim to stabilize the dendritic potential at $V_0 = -70\,\mathrm{mV}$ (v-iSTDP) and I_1cells regulate the firing rate of the soma, $x^E = 10\,\mathrm{Hz}$ (iSTDP). Both excitatory and inhibitory plasticity are driven by the pre-synaptic activity, in the form of filtered spike train (x_j^E) or sum of delta functions ($s_j^E, s_j^{I_1}, s_j^{I_2}$)

(**D**) External projections target a subset of excitatory neurons. The phoneme and word projections are fixed and target 5% of the excitatory population; 100 in the Tripod and DM networks, and 200 neurons in the LKD network. The phonemes are stimulated in sequence, and the corresponding word is stimulated throughout the interval.

Results

We investigate the capacity of a dendritic network to form stable and time structure-dependent memories by testing word recognition on a lexicon composed of words with shared phonemes; the phonemic overlap is such that correct word selection requires phonemes-word associations based on the phoneme orders rather than their identity. In the following sections, we analyze the network activity during the two phases and compare the network performance on a set of lexica with increasing phonemic overlap for neuron models with and without dendrites. Later on, we investigate the network structure emerging from unsupervised plasticity and analyze the interplay of recurrent connections and NMDA receptors in determining the sequence-recognition ability of the network.

Network activity in the associative and recall phases

We start the study of the Tripod network activity by tuning background noise and synaptic strengths such that the network is in a regime of sparse firing. When the three synaptic learning rules apply (v-STDP, iSTDP, v-iSTDP), the network's baseline activity has a low firing rate and tends to synchronize in slow bursts of activity. Conversely, when prompted with external stimuli, the network retrieves a low degree of synchrony and a sparse firing rate (Fig.2*A*). Compared to classical point-neuron spiking neural networks, the network of dendritic neurons appears more robust to variations in parameters. The improved network stability is due to the location of the recurrent connections on the dendrites; because the maximum current flowing from the dendrites to the soma is capped by the axial conductance, the network is refractory to epileptic firing regimes.

We characterize the firing activity of the network by measuring the average firing rate and the interspike interval coefficient of variation (ISI CV) per neuron (Fig.2*B*). To factor out the bias caused by the difference in rate, we measure the coefficient of variation by comparing only adjacent spikes, as suggested in Holt *et al.* (1996), referred to as ISI-CV2. In the first place, we observe that despite the somatic plasticity steers each neuron to fire at $10\,\mathrm{Hz}$, the network has a spread firing rate profile, not centered around the somatic iSTDP target. The lower and heterogeneous rate is due to the presence of the additional inhibitory rule on the dendrites. The firing profile of excitatory (Tripod) neurons divides into two classes. One group comprises neurons that receive external stimuli and thus belong to an assembly (*Exc. with projections*); these cells fire at various rates in the range $5\,\mathrm{Hz}$ to $15\,\mathrm{Hz}$. The second class is composed of neurons that do not receive external stimuli; they fire at a lower rate, in the range $0\,\mathrm{Hz}$ to $10\,\mathrm{Hz}$. The differences between stimulated and non-stimulated neurons are less visible in the ISI CV. For the excitatory neurons, the measure falls in the range 0.5 to 1.3, and it is

slightly lower for the neurons that receive external stimuli. The lower ISI CV2 indicates that the firing is regular within the intervals of elevated firing rate, which is consistent with a robust external drive that forces the neurons to fire. The robust firing activity of the excitatory cells associated with phonemes and words is necessary for vSTDP to form the cell assemblies associated with word memories. Interneurons of groups I_1 and I_2 fire at heterogeneous rates in the interval $5\,\mathrm{Hz}$ to $50\,\mathrm{Hz}$, their coefficient of variation remains around 1, consistent with Poisson input they receive from the remainder of the network.

Further analysis of the spectral and temporal correlation profile of the network is offered in Appendix A. The measures indicate that the activity of cell assemblies prompted by the word projections can be traced in the spectrogram and their activity has long autocorrelation. This is interesting from an experimental perspective because it indicates that lexical access has macroscopic signatures in the network.

Word assembly reactivates upon ordered phonemes

We follow the analysis of network activity by taking a closer look at the coordinated firing of word and phonemes population during the two phases. Fig.2*C* portrays the network dynamics during the associative and recall phases; it compares the raster plots of a subset of the phonemes (top) and word (bottom) populations during the associative (left) and recall (right) phases. The lexicon used in the experiment comprises ten words; some are contained within others, and some are anagrams or reversed. For example *poll, doll, lop, doll,* and *pollen.* The panels dissect an interval of the simulation during which the phonemes *P, O, L, L* are presented and evidence the co-activation (associative phase) or re-activation (recall phase) of the word *poll*, along with four competitor words. Both panels show the activity at a late stage of the simulation, which means that the formation of the assemblies has already taken place in the associative phase. Concerning the phonemes, we notice that despite the external input being the same for all populations, not all cell assemblies are reactivated with the same strength. The phonemes *p* and *l* respond with slightly stronger response than the phoneme *o* to the external stimuli. The difference is due to weaker recurrent connections of the phoneme *o*, and it is because the phoneme is shared among many words. The synaptic scaling applied to glutamatergic synapses tend to reduce the strength of less informative populations because they are more often activated and will share their glutamatergic resources among many pre-synaptic pathways. The illustration reveals the delayed and sustained reactivation of the networks' assemblies. The phonemes are stimulated for $50\,\mathrm{ms}$ only, but their firing persists for a longer time window ($100\,\mathrm{ms}$ or more). We deem such activity trace to be the result of dendritic depolarization and necessary for the formation of phonemes-to-words associations that are sensitive to the ordering of

the phonemes. The overlap of the phonemes' activity allows the formation of word memories that bind to the superposition of two phonemes. In the recall phase, the word populations do not receive inputs and are led to firing by the conjunct activation of the phoneme populations. All the words containing the input phonemes are partially reactivated, independently of their order. In the raster plot Fig.2*D*, upon the phonemes (*P, O, L, L*) both *lop* and *poll* populations fire; the latter, however, matches both the phonemes' identity and their order and it re-activate more intensely.

Similarly to the delayed activation of the phonemes, the word population activity peaks with a temporal offset to the stimulus. Comparing several word presentations, as in Fig.2*C*, we observe that the strongest response of the target word population happens with a delay of $100\,\mathrm{ms}$ to the onset of the first phoneme. i.e. the phoneme populations *P, O, L, L* are stimulated between $0\,\mathrm{ms}$ to $200\,\mathrm{ms}$ and the corresponding word assembly (*poll*) fire maximally between $100\,\mathrm{s}$ to $300\,\mathrm{s}$. Such delayed response is a structural property of the assembly reactivation, and an in-depth analysis of its time course is the focus of the following section. Now, in the interest of clarity, we assume that the optimal delay to measure the reactivation of the word population is exactly $100\,\mathrm{ms}$ after the end of the stimulus. Once the expected delay is fixed, the phonemic inputs determine unequivocally a time interval in which the corresponding word should be active (e.g., $100\,\mathrm{ms}$ to $300\,\mathrm{ms}$ *poll*; $350\,\mathrm{ms}$ to $500\,\mathrm{ms}$ *log*; etc.). Each of such intervals is considered a trial. Accordingly, we characterize the state of the network in each trial by computing the firing rate of all the assemblies, that is, the average among the neurons in the population. An illustration of such firing rate measure is expressed in Fig.2*E*; the four panels show the average firing rate of five assemblies during the interval associated with the sequence of phonemes (*P, O, L, L*). The population *poll* is repeated in the vertical axis of each panel; the horizontal axes report the activity of the competitor populations. The black circles indicate different trials where the same phoneme sequence was presented (approximately 100 trials), and the orange dots describe the average across all trials. For the phoneme sequence examined, the firing rate of the target word population is significantly higher than the other cell assemblies ($p < 0.01$); it can be used to discriminate the correct word from the competitors for almost every trial. Consistently with the associative mechanisms supporting word reactivation, word labels that share more phonemes with the target words are reactivated more robustly than those that share fewer or no phonemes (e.g., *doll* and *lop* versus *goal* and *log*). However, also the temporal component plays a role; the correct word *poll* is reactivated more than its anagram *lop*.

When compared against all the remaining words in the lexicon, we can estimate the average of correct trials for the word, e.g., $90\,\%$ for *poll*. The rate measures portrayed in Fig.2*D* are used to evaluate the network in the word recognition task in the remainder of this study.

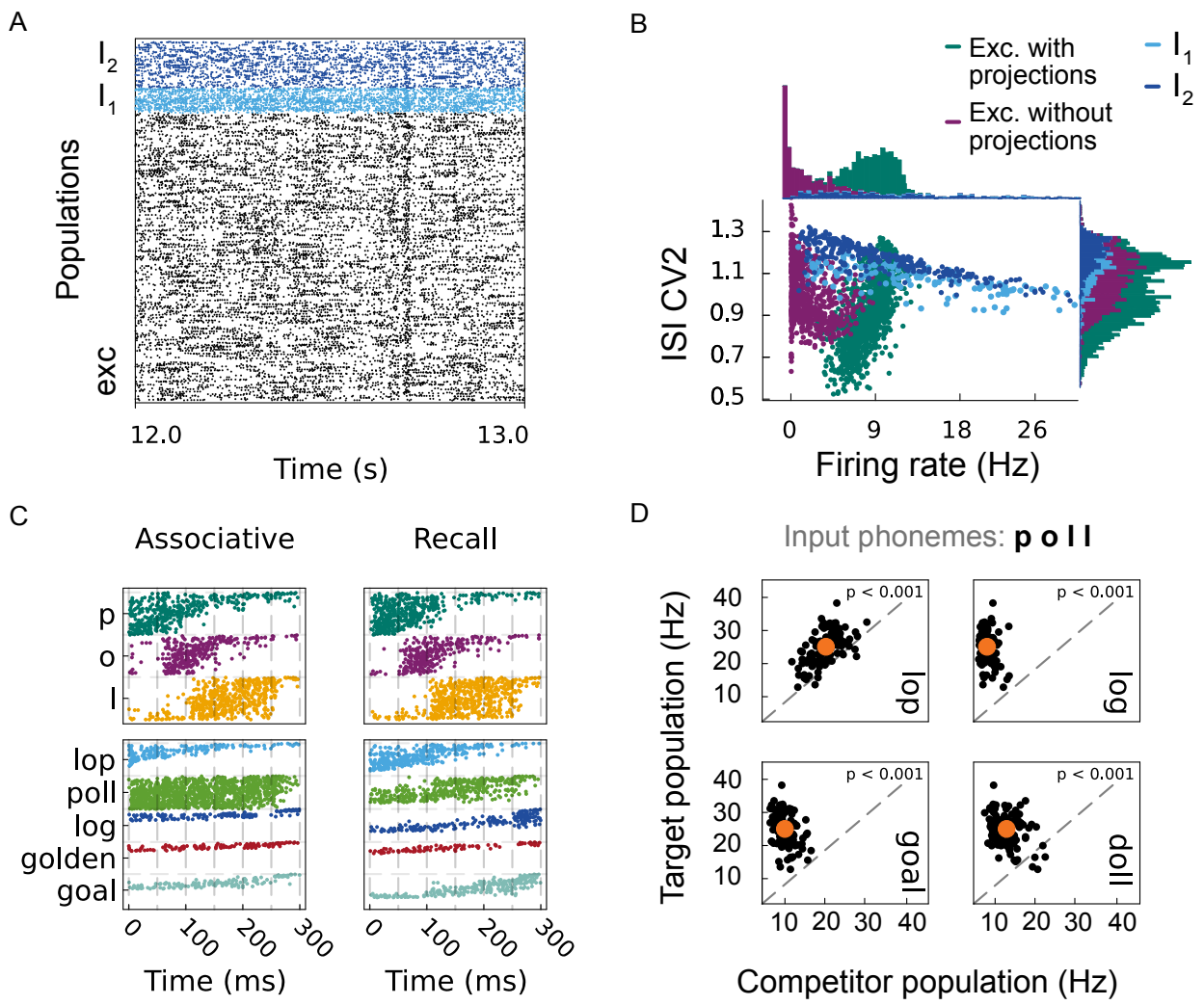

Figure 2: **Network activity and signatures of word recognition** (**A**) Raster plot of the entire Tripod network during an interval of 1 s. The excitatory neurons (Tripod) are portrayed in black, the fast-spiking interneurons (I_1) in light blue, and the slow-spiking interneurons (I_2) in darker blue. (**B**) Average firing rate and ISI CV2 for the excitatory and inhibitory populations. The excitatory neurons are divided into two groups, with (green) and without (purple) external projections. The neurons that receive external projections have a higher firing rate and a lower ISI CV2. (**C**) Raster plot of the phonemes and word populations during the presentation of the word *poll*. The phonemes are stimulated for 50 ms between 0 ms to 200 ms. The word assembly is externally stimulated in the associative phase (left) and re-activated by the recurrent connections in the recall phase (right). In the recall phase, it reaches the activity peak 50 ms to 100 ms after the end of the stimulus. (**D**) Firing rates of target versus competitor populations during the presentation of the phonemes *P, O, L, L*; the rates are measured over the word interval, shifted 100 ms forward. Each dot represents a phoneme sequence; when above the diagonal, it indicates a trial in which the target population fired more than the competitor assembly. The red dots show the average over trials.

Word recognition is successful after delay period

The preliminary analyses suggest that the Tripod network can achieve word recognition based on associative memories. However, phonemic overlap hinders correct reactivation, and the interference between cell assemblies could also cause mistakes. To characterize the model's capacity, we systematically compare the recognition accuracy for seven different lexica that vary in size, number of shared phonemes (word overlap), average word length, and onset overlap (sequence overlap). The seven lexica are borrowed from previous studies in word recognition or introduced for the present study. They are listed in SI Fig.3 . For the first two lexica, each word can be distinguished based on the first phoneme because they have no phonemic overlap, we refer to these as memory-less lexica. The remaining five require the integration of the symbols in a sequence.

To characterize the network word recognition capacity, we measure the firing rate of each word population during two intervals; the duration of the word and the duration of a phoneme. The word interval is measured at the end of the word (onset point) and the phoneme interval is measured at the phoneme that makes the word unique in the lexicon (uniqueness point). In addition, we shift the interval forward or backward to test the timing of word re-activation, we refer to this shift as the interval delay. For each measured interval, we select the population with the highest firing rate as the network output. An illustration of the measure is depicted in Fig.3 and fully detailed in the Method section. From the trial-by-trial recognition measures, we compute the confusion matrix and the average word recognition score.

Visual inspection of the confusion matrices indicates that word recognition grows steadily when the intervals are shifted $50\,\mathrm{ms}$ to $100\,\mathrm{ms}$ forward in time, both for the offset and uniqueness point measures. An example of the effect of the delay on the confusion matrix is illustrated in Fig.3*B*; the stacked heatmap shows the evolution of the confusion matrix for a sample lexicon (*5. Overlap*) for the offset point measure. The first confusion matrix corresponds to $0\,\mathrm{ms}$ of delay; the pairing between target and reactivated words is almost random. The confusion matrix aligns more and more on the diagonal as the delay increases till a maximum around $100\,\mathrm{ms}$ delay. This result is somewhat surprising because such an interval overlaps with the presentation of the following word. It is also remarkable that all the words reach their optimal recognition with similar delays (i.e., $\sim 100\,\mathrm{ms}$), because the lexicon probed has words with 3 to 7 phonemes, that is $150\,\mathrm{ms}$ to $350\,\mathrm{ms}$.

The scoring algorithm associates every delay shift with an average recognition score. Thus it offers a delay-dependent measure of word recognition; in addition, we rescale it against chance to obtain the Cohen's κ-score (Methods). We compute the measure for the seven lexica for the offset and uniqueness point measure. The recognition κ-score, plotted against the delay interval, is presented in Fig.3*C*. In agreement with the results on the confusion matrix, the κ-score increased

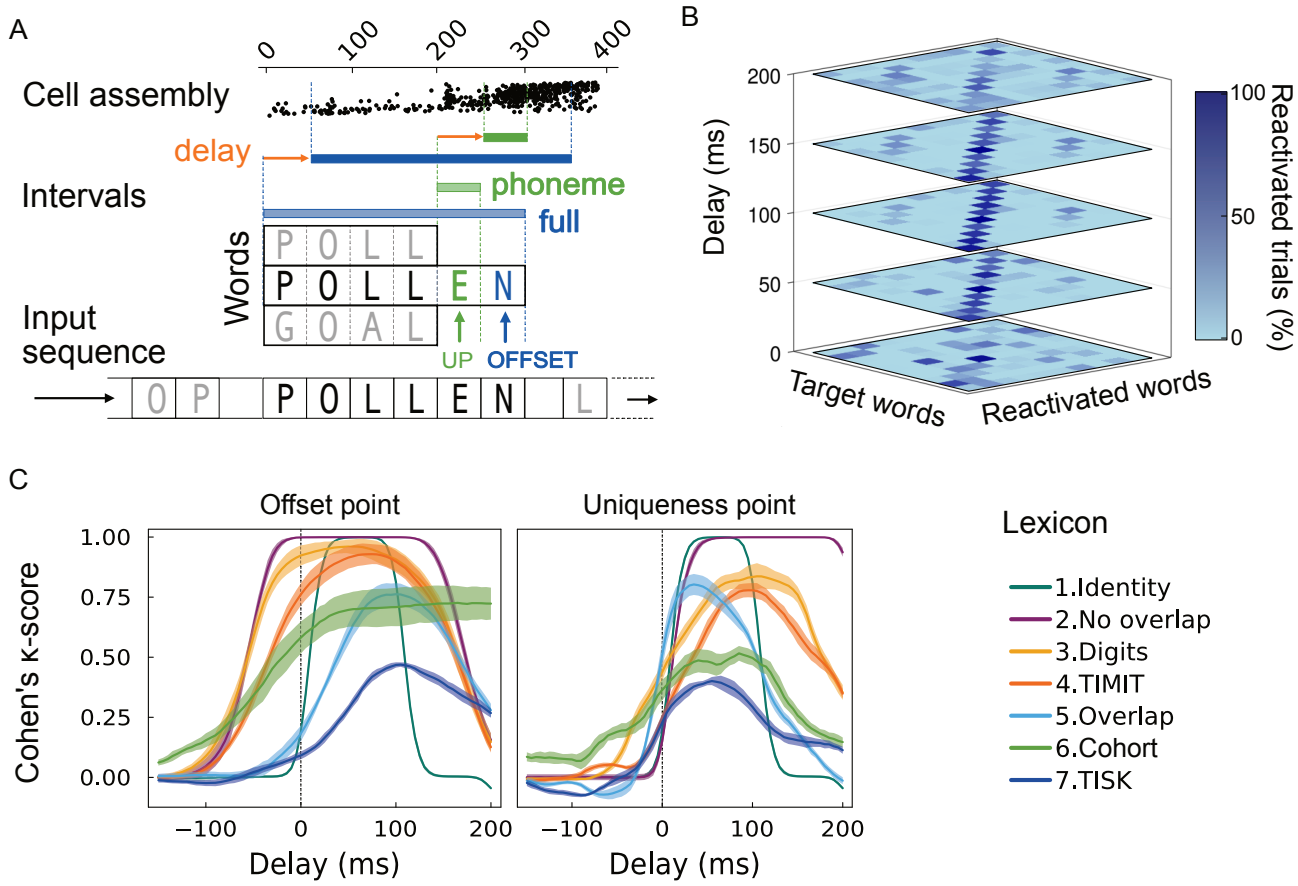

Figure 3: **Word recognition in the offset and uniqueness point measures** (**A**) Word recognition is tested for each input sequence at the word's offset and uniqueness points (blue and green arrows). The assembly activity is measured as the average firing rate of the assembly population in the interval between the word onset and the offset (full interval, blue bar) or the interval of the phoneme at the uniqueness point (single phoneme interval, green bar). The measured intervals can be delayed arbitrarily (orange arrow). The word with the largest firing rate is chosen as the network output. (**B**) The network responses are summarized in the confusion matrix; the matrix indicates the percentage of trials in which a word was reactivated for each target word. The recognition score is higher when the matrix aligns on the diagonal, indicating that the correct word was reactivated for each presentation. Off-diagonal values indicate mistaken reactivation. The vertical axis indicates the interval delay at which the matrix is computed. The matrix corresponds to the *5. Overlap* lexicon, with the offset point measure condition. (**C**) Cohen's κ-scores for the seven lexica averaged across the ten equivalent samples for both the offset (left panel) and uniqueness (right panel) point measures. The chance-corrected recognition scores are plotted as a function of the interval delay. All lexica reach a maximum for delays between $50\,\mathrm{ms}$ to $150\,\mathrm{ms}$. The sharp rise of the κ-score in the UP panel highlights that correct word recognition cannot be achieved before the uniqueness point.

when shifting the measured interval forward in time. For both the OP and UP measures (left and right panels), recognition peaked in the range $50\,\mathrm{ms}$ to $150\,\mathrm{ms}$. The offset point, or full interval measure, has a gentler trend than the UP because the measured interval is longer, and it smoothens the firing rate measure.

The recognition score was generally higher for the long intervals measured at the OP rather than the one-phoneme intervals measured at the UP. For the OP, the best accuracy was beyond $70\,\%$ for all but the *7. TIMIT* lexicon. For the UP, it was proportional but slightly lower (the Pearson correlation between the two scores is 0.9). The score difference is due to the correct assembly reactivating over a time course longer than one phoneme. However, despite the UP interval having a narrow time scope to pick on the network activity, the similarity between the two measures is remarkable. It indicates that the contribution of the activity before and after the UP is marginal or, at least, redundant. Further insights into the assembly activity before the uniqueness point are offered by the up-stroke of the score-delay curve in the UP panel (right). The measure indicates that correct recognition is impossible ahead of the UP (delays below $0\,\mathrm{ms}$); most lexica have a recognition score close to zero, which suggests that the correct word assembly is not yet active or that the network is in a state of co-activation, with not enough information to discern between words (flat CM, $\kappa = 0$). The measure moves away from zero for the lexica with longer words because some of their words can be discarded before the UP, resulting in a higher chance of correct recognition (skewed CM, $\kappa > 0$); for example, for the *6. Cohort* lexicon, the sequence *c, a, p, i, t* will exclude *capias, capillary, capistrate* despite the last phoneme in *c, a, p, i, t* is before the uniqueness point. Notably, the ten realizations of the network reached a similar score on the same lexicon; the standard deviation of the recognition score was below $6\,\%$ for all the dictionaries and conditions, except the *6. Cohort* dictionary in the offset measure, which reached $10\,\%$.

Dendrites are necessary to process lexica with phonemic overlap

Aiming to unveil the computational role of dendrites in the formation and reactivation of word assemblies, we compare the performance of the Tripod network with other models without dendrites. The two point-neuron network models were chosen for their biological plausibility and relevance to the task at hand. One of the models is the LKD network; it is a spiking neural network model that leverages excitatory and inhibitory STDP to form and maintain cell assemblies (Litwin-Kumar & Doiron, 2014). In the original work, the assemblies are formed upon presenting 20 weakly overlapping stimuli in the associative phase. During the recall phase, the network was shown to spontaneously and robustly reactivate the assemblies and to achieve pattern completion when the assemblies were partially stimulated. The second model, the DM network, emulates the network

of the L2/3 circuit of the mouse cortex. The study aims to probe heterogeneity's impact on the network's computational capabilities, such as temporal integration and delayed decision task(Duarte & Morrison, 2019). The network has no plasticity in the connections, and its recurrent weights are calibrated to achieve a balanced excitatory-inhibitory state. Because the word recognition task requires strengthening phonemes-to-words associations, we endowed the DM network with STDP on the excitatory-excitatory synapses. Preliminary analysis showed that the DM network was unstable and prone to epileptic oscillations. Hence, to prevent the network's instability, we added iSTDP and v-iSTDP homeostatic inhibition on the neurons' soma in the same form as the Tripod network. Both the LKD and the DM networks implement conductance-based synapses. However, only the DM network considers NMDARs, which are known to play a role in the maintenance of working memory, favoring the onset of persistent activity in networks' cliques (Wang, 1999; Papoutsi *et al.*, 2014; Wang, 2021). An illustration of the point-neuron models and their synaptic learning rules is presented in Fig.1*B*-*C*; the equations and parameters are detailed in the Methods. Additionally, under the hypothesis that the Tripod model configuration and its connectivity patterns are necessary to sequence recognition, we compared the vanilla Tripod model with a model with one single dendritic compartment and one with symmetrical projections and recurrent connections, all the other parameters being the same. The two variations are illustrated in Fig.4*A* and further described in the Methods.

The five models were trained on the seven lexica introduced in the previous section. For each pair of models and lexica, we computed the recognition score for the offset and uniqueness point measures, with delays in the range $-50\,\mathrm{ms}$ to $150\,\mathrm{ms}$. The comparison between the models' accuracy and average recognition delays (ARD) in the word recognition task is presented in Fig.4*B* and *C*. The panels refer to the scores and ARDs obtained for the offset measure; the results for the uniqueness point were qualitatively similar and are reported in SI Fig.4 . In the panels, we use a greyscale color code to indicate the two lexicons that do not require memory to be correctly recognized and colors to indicate those that require temporal integration. The second group of lexica scores is averaged together for each model and compared between each model in the right panel of Fig.4*B*. The color codes for variations within $+5$ percentage points between the model on the y-axis and the model on the x-axis. The panel also illustrates the significance of the relative differences between pairs of models.

The point-neuron models perform well on the memory-less tasks, but their accuracy drops on the lexica that require temporal integration. A closer look at their confusion matrices reveals that the networks indiscriminately reactivate only the two or three assemblies upon presenting any stimuli. Thus, the point-neuron models are missing the sequential structure of the word memories.

In contrast, the dendritic models (Tripod asymmetrical, symmetrical, and single dendrite) recognize most of the words in all the lexica presented; the errors are distributed evenly within their confusion matrices (SI Fig.3). Among the dendritic configurations, the differences in recognition are minor (within 5 percentage points); although the asymmetric Tripod has higher scores than the other models, the improvements are barely significant. The average delay measure, portrayed in the bottom panel in Fig.4*B*, confirms that the memory-requiring lexica can be correctly processed only when recognition is delayed upon the input presentation. The point-neuron models have negative average delays, indicating that the assembly reactivates immediately upon the phonemic stimulus.

Immediate memory access is a viable strategy if the identity of the early phonemes is sufficient to recollect the correct memory, as it happens for the first two lexica. Otherwise, if the early phonemes contain insufficient information, the recognition will be, at best, based on the marginal distribution of the phonemes. Quite the opposite, the dendritic model leverages the temporal integration of the inputs to reactivate the correct associations. The recognition happens with $50\,\mathrm{ms}$ to $150\,\mathrm{ms}$ delay; in addition, the ARDs follow the same trends evidenced in the previous section, i.e., governed by the statistics of the lexicon (word overlap, lexicon size, words length, and sequence overlap). Remarkably, the ARDs of the dendritic models resemble each other in the offset and uniqueness point measures. The model's similarities suggest that the networks deploy similar mechanisms to carry information over time and reactivate the words independently of their connectivity and configuration. Crucially, the difference in accuracy does not depend on the lexicon, suggesting that neither asymmetry nor the number of dendrites provides an additional and necessary mechanism to solve the more complex problems. On the other hand, the lack of significant differences between the point neuron models indicates that voltage-dependent inhibitory plasticity in the DM network is insufficient to solve the task.

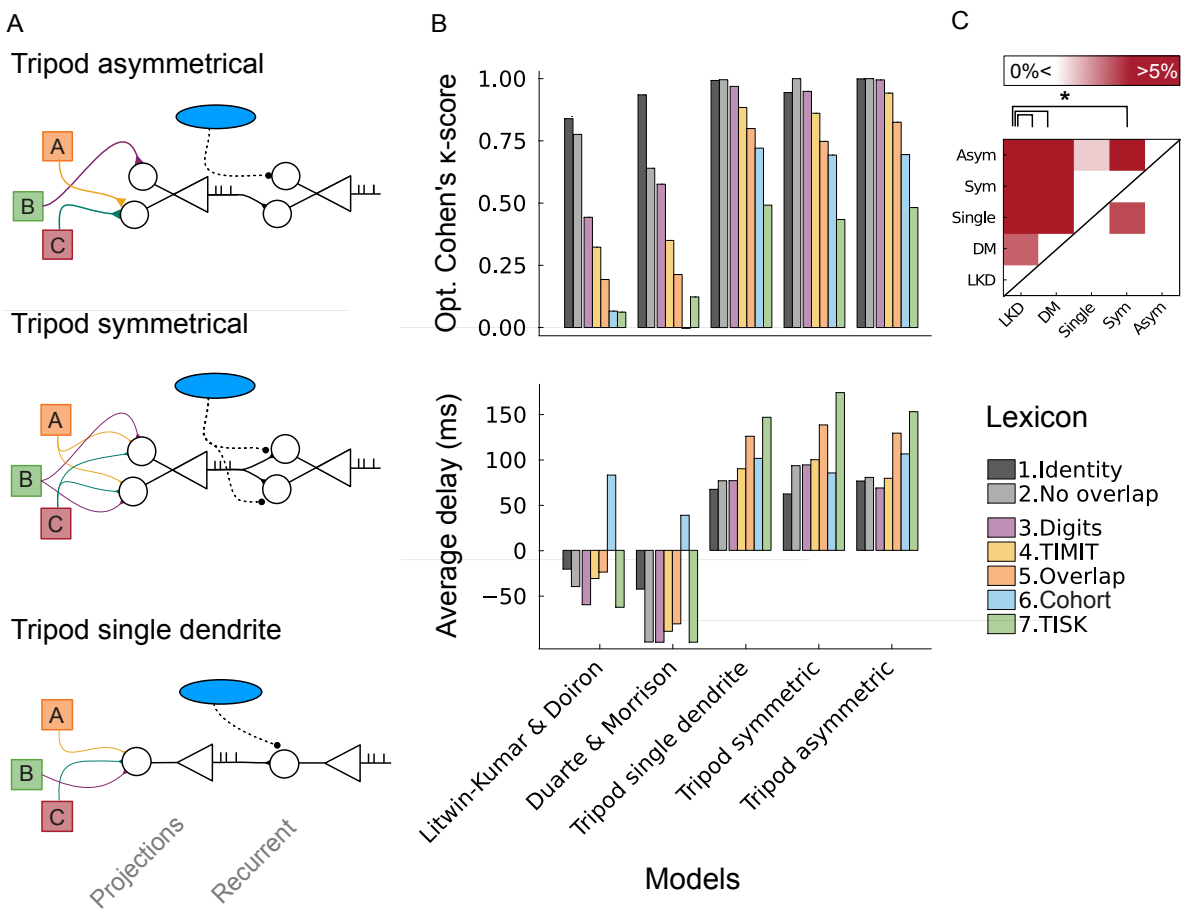

Figure 4: **Comparison of point-neuron and dendritic networks** (**A**) Dendritic configurations and connectivity patterns of the Tripod network models. The asymmetrical configuration has dendrites with length uniformly distributed in $150\,\mu m$ to $400\,\mu m$ and pre-synaptic connections independent from one to the other dendrite; it is the default for the Tripod network. The symmetrical configuration also has two heterogeneous compartments, but they receive the same external and recurrent connections. Eventually, the single dendrite model has one single compartment that hosts all the synapses. (**B**) The optimal recognition score and the optimal delay for the five network models and the seven lexica tested. The grayscale codes for the two memory-less lexica and the colors indicate the five lexica requiring temporal integration. The top panel illustrates the optimal recognition score achieved by the models, and the corresponding column in the lower panel indicates the average recognition delay (ARD). Point neuron models perform well on the first two lexica, with zero delays; conversely, the dendritic models integrate information over time, which is essential to reach higher scores on the lexica with overlapping phonemes. **C** The square-matrix reports the performance difference on the five lexica that require memory (3-7). The matrix expresses the difference between the model on the y-axis and the model on the x-axis, with a resolution of five point percentage.

From the Tripod network activity analysis, we hypothesize that heterogeneity in the average firing rates of the neurons is an indicator of the network performing the task correctly. Indeed, it is the case for the three dendritic networks, which have remarkably similar firing patterns with the Asymmetric Tripod model; the neurons that belong to an assembly are in the range $4\,\mathrm{Hz}$ to $12\,\mathrm{Hz}$, while the remainder of the network in the range $0\,\mathrm{Hz}$ to $5\,\mathrm{Hz}$. Conversely, the point-neuron models have fewer disparities among neurons within and without cell assembly, and their distributions are centered around the rate targeted by the iSTDP homeostasis. The firing rate distribution for the five models is portrayed in SI Fig.4 *B*.

From the comparison between the five models, we infer that neither the asymmetry of the dendrites nor their number of compartments provides substantial computational advantages. Neither does the inclusion of voltage-dependent iSTDP in point-neuron models. Excluding the alternative, the dendritic memory enabled by the NMDARs must be the crucial mechanism that allows the network to solve the word recognition task. However, the summation of pre-synaptic stimuli over a longer time interval is insufficient to solve the task, as the network must respond selectively to phonemes dispositions rather than their combinations, i.e., order matters. The sequence recognition capacity stems from the interaction of these structural features with the associative plasticity taking place in the synapses. In the following section, we show the network architecture emerging from the associative phase and prove that it is necessary for achieving the task.

Dendritic models have pronounced phonemes to word connections

Because the phoneme populations do not code for any temporal information, the network must rely on recurrent connections to process their order and then reactivate the correct assembly. The correct reactivation of the two words *dog* and *god* (lexicon *5.Overlap*) offers an example. The words share the same phonemes, but their temporal arrangement is different. Thus, upon the presentation of the phoneme sequence *D* and *O*, the word assembly *dog* and *god* must enter a different state; such that *D* before *O* rather than *O* before *D* prompts the reactivation of *dog*. The nature of the network architecture underlying this computation is unknown and is the focus of the following investigation.

First, we introduce a minimal notation that will allow us to be precise about analyzing the network structure. Tripod neurons have recurrent connections to $20\,\%$ of the dendritic compartments of the excitatory population. We identify the set of these connections as the matrix W^d_{ij}, where d is the compartment of post-synaptic neuron i, on which the pre-synaptic cell j fires. The three dendritic models have three different recurrent configurations. For the asymmetric model $W^1_{ij} \neq W^2_{ij}$, and the two matrices are drawn independently; the symmetrical model has $W^1_{ij} = W^2_{ij}$; and for the single dendritic model only W^1_{ij} is defined. Starting from the connectivity matrix, we can define

the *effective connectivity* (EC) matrix as the average of the connections between neurons from the pre-synaptic assembly A to the post-synaptic assembly B.

$$\mathcal{C}_{BA} = \frac{1}{N_A N_B} \sum_{i \in A;\ j \in B} ½(W^1_{ij} + W^2_{ij})$$

The EC is a square matrix and is not symmetrical, as the connections from pre- and post-synaptic neurons are not necessarily reciprocated. The EC is composed of four blocks, the connections between words ($\mathcal{C}_{WW}$), between phonemes ($\mathcal{C}_{PP}$), from phonemes to words ($\mathcal{C}_{WP}$), and from words to phonemes ($\mathcal{C}_{PW}$). An example of the EC matrix for the Tripod Asymmetric model on the *5. No overlap* lexicon is presented in Fig.5*A*; the EC is computed on the connectivity matrix at the end of the associative phase. The x-axis indicates the pre-synaptic population, and the y-axis the post-synaptic. The panel divides the group of phonemes and word populations and evidences the different modes of connectivity within recurrent (top-right, bottom-left) and feedforward/feedback (bottom-right and top-left, respectively) connections; the recurrent connections within the engram are on the diagonal. Based on this division, we calculated which phoneme and word population should be associated for the task to be correctly achieved. We define a group of *valid* connections: the connections from any phoneme to the words that contain them and reverse, and all the recurrent connections of phonemes and word assemblies. A scheme of these connections and the average weights for the five models considered is portrayed in Fig.5*B*. The averages are scaled to the maxima of each EC matrix to facilitate the model comparison. The minimum refers to the initial synaptic strength W_0, which sets the synaptic-weight budget for the STDP homeostasis; it is the same for the three dendritic models ($10\,\mathrm{pF}$) and different for the two point-neuron models (DM $0.45\,\mathrm{pF}$; LKD $2.76\,\mathrm{pF}$). For completeness, the ECs for the remaining models are portrayed in SI Fig.5 . Notice that here the scale changes for every model.

In the interest of brevity, we limit ourselves to describing the connectivity for the asymmetric Tripod model, depicted in Fig.5*A*. To start with, the strongest engrams are those connecting the word populations. Such connections determine the mutual contribution between word assemblies. In connectionist models of word recognition, (McClelland & Elman, 1986) they are usually inhibitory and are thought to mediate the competition between words. Such a mechanism is absent in our model, as the synapses follow Dale's principle and are only glutamatergic. The homeostatic plasticity in vSTDP is the only mechanism that can mediate competition, and its effect is expressed in the weak reciprocal connections among words. For the two pair of words with shared phonemes, *lop, poll* and *dog, god*, homeostasis seems ineffective as the assemblies have significant excitatory interactions. The second block of interest is the quadrant of phonemes to word connections (bottom right). These connections are strictly necessary for the network to solve the task because the externally-driven activity in the phoneme populations is forwarded to the word assemblies via these glutamatergic pathways. The

panel shows that the synaptic weights are significantly above the average only if the pre-synaptic phoneme is contained in the post-synaptic word (valid connections). Thus, the connections emerge because of the externally-driven co-activation of phonemes and word populations. Interestingly, the phonemes-to-words connections have a similar, although much weaker, structure in the DM network but are absent in the LKD network.

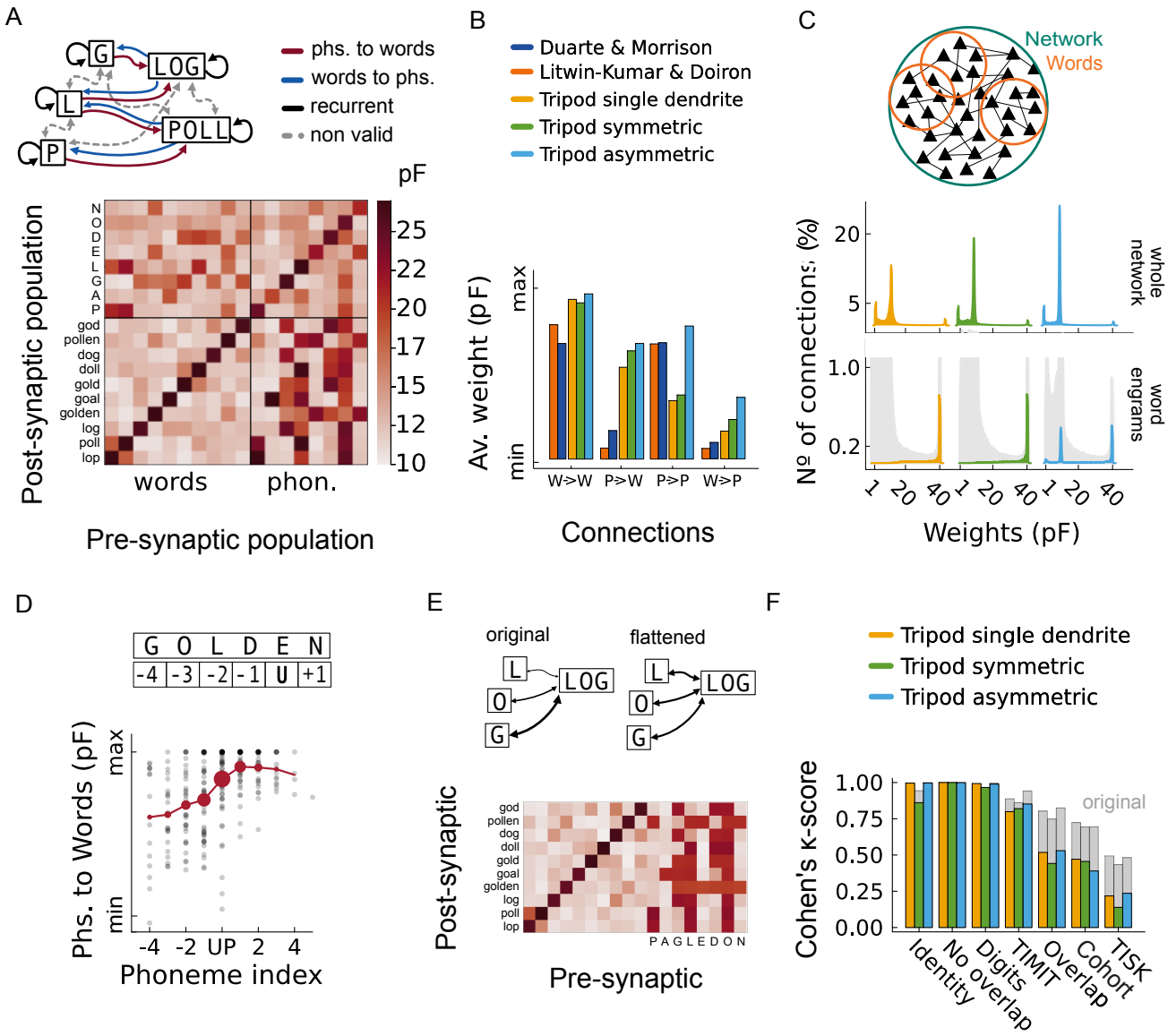

Figure 5: **Dendritic networks have feedforward structures that are necessary for correct word recognition**

The remaining two blocks involve those synapses that target post-synaptically the phoneme populations. These connections have a lesser impact on the network dynamics because they are in-

Figure 5: (**A**) Effective connectivity matrix of the network's assemblies. The heatmap expresses the average synaptic weight among two assemblies; weights are measured in pF. The matrix refers to the asymmetric model on the *5. Overlap* lexicon. The four quadrants correspond to the recurrent phonemes and word connections, the feedforward connections from phonemes to words, and the feedback connections from words to phonemes. The schematic illustrates these four types of connections and separates them from the *non-valid* connections, which are not expected to contribute to the word recognition task. The *non-valid* connections bind phonemes with words that do not contain them or those among different words. (**B**) Average weights of the four types of connections aforementioned for the point-neuron and dendritic models. All models have strong word connections, but only dendritic models have strong phonemes-to-word connections. (**C**) Distribution of network weights for the entire network (top panels) and the word recurrent connections (bottom panels). Most of the connections maintain the synaptic weight of the initialization $10\,\mathrm{pF}$; conversely, the recurrent word connections have strong synapses ($41\,\mathrm{pF}$). The asymmetric Tripod has half weak and half strong connections within the recurrent assemblies; the other models have only strong connections. (**D**) Strength of the phonemes-to-word connections organized by the order of apparition in the word. The phoneme index is computed starting from the uniqueness point. The panel shows together the words from the five lexica that require memory. The red dots and lines indicate the average among words, the size of the dots the size of the sample. (**E**) Connectivity matrix resulting from removing the order structure in the phonemes-to-word connections. The scheme illustrates the operation for the connections of the word *log*. (**F**) The panel shows the word recognition scores achieved in networks with flattened connections; in grey are those achieved with the original connectivity matrix. The κ-score diminishes for the three dendritic models, and the loss worsens for the lexica that require temporal integration.

herently weaker than the external projections that continuously stimulate the phonemes populations. However, the architecture of the phoneme engrams can favor the reactivation of the correct word assembly if it increases the stimulus gain for specific patterns of phoneme sequences. Similarly, the words-to-phonemes connections can indirectly contribute to routing the network dynamics toward the correct word selection. In this regard, it is crucial to notice that the phonemes-to-phonemes and words-to-phonemes present the only substantial difference between the dendritic models hitherto analyzed; for the asymmetric Tripod, the block of phonemes connections has diagonal and off-diagonal connections larger than the average. Preliminary analyses indicate that phonemes-to-phonemes connections play a role in the repairment of corrupt input sequences, thus suggesting a better fit of the Tripod asymmetric model as a computational model of the human language faculty. For the following, we instead focus on the connections that bind post-synaptically to word assemblies because they are central to the word recognition task.

A picture of the average differences and similarities between the five models can be grasped in Fig.5*B*. The dendritic models have highly structured EC matrices where word-to-words and phonemes-to-words connections dominate. Point-neuron models have weak connections binding the pre-lexical units to the word assemblies; conversely, their recurrent phonemic connections are more robust than the symmetrical and single dendrite models and equal to the asymmetrical Tripod. All models have weak word-to-phonemes connections. Overall, the analysis of the effective connectivity matrices indicates that the presence of dendrites favors the strengthening of non-recurrent connections among phonemes and word assemblies, and more so for the asymmetrical Tripod model.

One explanation for the differences among the dendritic models is suggested from the analysis of the weight landscape for the three networks. The histograms in Fig.5*C* portray the percentage of connections (y-axis) with their respective weights (x-axis), for the entire network (top) and within the recurrent word engrams (bottom, grey shades show the entire network for comparison). The top row indicates that the single dendrite, symmetrical, and asymmetrical Tripod models reach similar network configurations. The weights distribution have three peaks about 1,10, and $40\,\mathrm{pF}$. The latter is the maximum synaptic strength allowed among two compartments, and it is a stable point of recurrent glutamatergic connections. The homeostatic process requires that if a compartment has a strong incoming synapse, some other synapses have to be small; considering the hard limits set for the dendritic network STDP, 1 fully potentiated synapse imposes about 20 to 25 fully depressed ones. Thus, synaptic homeostasis explains the leftmost peak at $1\,\mathrm{pF}$ to $2\,\mathrm{pF}$. The synaptic weight distribution's third, most pronounced peak is at $10\,\mathrm{pF}$. It is the initial synaptic strength; such connections are idle because they do not actively contribute to any engram. Interestingly, the

asymmetrical Tripod has twice the idle connections of the single dendritic model and $40\,\%$ more than the symmetrical Tripod. The difference is markedly visible in the lower panel; for the two Tripod variations, most of the recurrent connections are fully potentiated, but only half are for the asymmetrical Tripod. The weaker words engrams in the asymmetrical configuration spare a large number of synaptic resources that the network can rearrange for other connections, which explains why the asymmetrical Tripod has stronger phonemes-to-phonemes and words-to-phonemes engrams than the other models.

Structures in the effective connectivity matrix mediate correct word recognition

Returning to the central topic of the study, we question if the variability in the phonemes-to-words connections, visible in the lower-bottom block of Fig.5*A*, is causal to the increased word-recognition capacities of the Tripo models. In other words, is it a matter of chance or reflects the underlying sub-lexical structure of the lexica? We hypothesized that in order to reactivate the correct word population, the network has to encode the reciprocal order of each pre-lexical unit. To test this idea, we assigned to each phoneme an index, matching its order in the word, and measured the average synaptic connection between the phoneme and the word assemblies. To counteract the inherent randomness in the network weights, we increased the statistical power by considering together the five lexica that require memory. The weights of the phoneme-to-word connections are not matched across the networks because of the varying number of phonemes and words. Therefore, before stacking together the words from different lexica, we normalized the synaptic weights by the strongest phoneme-to-word connection in each word. Similarly, because words have different lengths and the average is diverse across lexica, we centered all the samples to their uniqueness points. The scheme on the upper part of Fig.5*D* illustrates the indices of the word *golden* from the *5. Overlap* lexicon, the phoneme *E* corresponds to the uniqueness point. The lower panel shows the average synaptic weight of each phoneme-to-word connection (y-axis) matched with the index of the phoneme in the word (x-axis); the size of the dot indicates the number of words that contained that index (e.g., the word *golden* does not contain +3 but contains -2). The lower extreme corresponds to the initial synaptic strength of the connections (W_0).

The plot in Fig.5*D* shows a significant interaction between the phoneme index and the strength of the feedforward connections between phoneme and word assemblies. Early phonemes have $15\,\%$ less strength than phonemes at or after the uniqueness point. The curve shows a steep increase between the phonemes right before (-1) and right after (+1) the UP. Interestingly, it also steers down for phonemes three or four indices ahead of the UP, although we cannot conclude on the right side

of the UP because our sample is limited. We can, however, offer an interpretation of the results consistent with the preliminary observations on the network. During the associative phase, the early phonemes start firing because of the external stimulation; simultaneously, the word populations start firing because of the word projections, the feedforward, and the recurrent connections; with time, the dendritic compartments of the neurons in the word assemblies become more depolarized, which favor the strengthening of synaptic weights with later phonemes. Eventually, at the uniqueness point, the word assembly reaches its activity peak, and neuronal adaptation and feedback inhibition start hyperpolarizing the cells. The result is that later phonemes engrams fire on the word cells when the dendrites depolarize, and the STDP is less effective. This hypothesis remained untested in the context of the present work because it would require direct manipulations of the membrane potential and STDP parameters.

On the other hand, we could test if the weight structure is causal for the word recognition task. For this aim, we manipulated the network weights and flattened the phoneme-to-words connections, such that the curve in Fig.5*D* would appear flat. The panels in Fig.5*E* show a schematic of the transformation and the resulting EC matrix for the asymmetric Tripod network with the *5. Overlap* lexicon. For each post-synaptic word (y-axis), the strength of the incoming connections from the phoneme assemblies (x-axis) is homogeneous. To achieve this, we updated the presynaptic connections of all the word populations, such that given the words A and the phonemes contained in it P_1^A, ..., P_N^A all the non-zero synapses result:

$$W_{i,j}^d = \sum_{i,j} \frac{(W_{ij}^1 + W_{ij}^2)}{2N_p} \quad \text{for: } i \in A; \;\; j \in \left[P_1^A, ..., P_N^A\right] \tag{1}$$

We then ran the recognition task as in the previous sections and computed Cohen's score for all the delays. The κ of the three models on the seven lexica are indicated in Fig.5*F*, compared with the scores achieved in the original condition. Not surprisingly, the internal structure of the phonemes-to-words connections matters for the lexica that require memory more than the others. The network performance degrades more or less the same for the three dendritic models, losing $40\,\%$ to $60\,\%$ for the three lexica with the highest degree of phoneme overlap. In addition, the average delay before word recognition decrease in the flattened condition, suggesting that the temporal integration capacity is altogether damaged by the transformation in Eq.1.

The present results illuminate dendrites' pivotal role in forming a structured and functional connectivity matrix. The particular arrangement of the average synaptic efficacies implies that in the memory formation processes, multiple factors act in concert, such as the external stimuli, the density of the recurrent projections, and the available feedforward pathways. In the interest of a synthetic presentation of the results, we omit from the main text two insights that we deem valuable and set the ground for future work. The first concerns the dynamics behind the formation and the

maintenance of word memories; in short, formation is quick ($50\,\mathrm{s}$ to $100\,\mathrm{s}$), and word memories are robust and remain stable when the plasticity is switched on in the recall phase; Appendix C offers insights into the word memory formation and maintenance process.

We now move to the last section and eventually shift the focus from the network structure to the dendritic memory mechanism. We show that the finely tuned connectivity matrix is necessary for word recognition but is insufficient. The network and its dendrites must be governed by feedback inhibition at a specific operational point.

Dendritic non-linearity governs temporal integration

Our results have shown the neat superiority of dendritic networks in the word recognition task, which follows from delayed temporal integration and fine-tuned assembly connectivity. However, to leverage the dendritic capacity, the network's cells must be at a specific operational point that allows the expression of NMDA spikes while suppressing the non-linear integration of spurious signals. Such a computational state is allowed by the voltage dependency in the NMDARs and tight dendritic inhibition. We now show that the coupling of these mechanisms is necessary for the network to achieve correct word recognition.

In the Tripod network, the dendritic non-linearity and inhibitory control are governed by three parameters, the ratio of NMDA-to-AMPA receptors in the synapse (NARs), the decay timescale of the NMDAR receptors (τ_d), and the target membrane potential of the voltage-dependent iSTDP of I_2neurons (V_0). Fig.6*A* illustrates the impact of variations in the NMDAR parameters on the dendritic membrane potential of the post-synaptic cell. The panels show the excitatory post-synaptic potential (EPSP) of a $300\,\mu\mathrm{m}$ dendrite following an excitatory spike on a synapse with weight $100\,\mathrm{nF}$. The left panel shows the variations of the NAR (0 to 2.7) and the right panel of the decay timescale ($2\,\mathrm{ms}$ to $50\,\mathrm{ms}$); in both panels, the dashed line indicates the baseline synaptic model (NAR$=1.8$, $\tau_d = 34\,\mathrm{ms}$). The curves indicate a threshold in the NAR for the onset of the non-linear response, which is approximately NAR $= 1.2$. Above this value of the NAR, the membrane enters into a long-lasting depolarized state named *plateau potential.* Durable depolarizations are the fundamental mechanisms for dendritic memory, and the conditions for such states are amply discussed in Quaresima *et al.* (2022). Conversely, varying the timescale does not affect the onset of the plateau; instead, it changes its duration in time. Similarly, Fig.6*B* shows the consequences on the inhibitory network weights of varying the v-iSTDP V_0 in the range $-90\,\mathrm{mV}$ to $-40\,\mathrm{mV}$. The histograms depict the distribution of the post-synaptic weights between I_2neurons and the dendritic compartment; for the baseline value (V_0), the weights accumulate at the maximum of the inhibitory weight range allowed ($243\,\mathrm{pF}$). When the target V_0 increases, the distribution tends to flatten toward smaller values.

Variations in the strength and timescale of the non-linearity and the v-iSTDP mechanism are also traceable in the network activity. To highlight the changes in neuronal dynamics, we compared the average dendritic membrane of a word population from the remainder of the network while presenting the associated phoneme sequence. Such a measure can be considered an event-related potential (ERP) applied to dendritic depolarization, where the network act as the control condition. The difference between the control and the dendritic potential of the target population (*poll*) is shown in Fig.6*C* for all the parameter variations hitherto presented. The left-most and central panels show similar trends; for weak and short dendritic non-linearity, the ERP is close to zero, and the word assembly activity peaks right after the first phoneme (*P*) but then remains close to the control condition. Conversely, suppose the non-linearity is as in the baseline model or stronger. In that case, the dendritic membrane potential builds up throughout the presentation of the phoneme sequence till it reaches a peak $50\,\mathrm{ms}$ to $100\,\mathrm{ms}$ after the word offset. This pattern is highly consistent with the measures of the ARD in previous sections. Despite the different governing mechanisms, a similar ERP effect can be observed when the variations are applied to the v-iSTDO V_0. The rightmost panel indicates that the ERP is wider when the dendritic inhibition is tighter, although the other conditions also maintain a remarkable distance from the control dendritic membrane. The ERP is the same when V_0 is equal to or larger than $-70\,\mathrm{mV}$ because the $\mathrm{I_2}$weights are the same; hence the plasticity rule appears the same to the network dynamics. However, the ERP's direction seems counterintuitive regarding the inhibitory function; more potent dendritic inhibition should cause weaker dendritic depolarizations. Hence, we inspected the average membrane potential for the word assembly and control condition to shed light on the ongoing dynamics. The four panels in SI Fig.2 show the soma and dendritic average potentials and solve the paradox on the role of v-iSTDP control. When the inhibitory target potential is larger, e.g., $V_0 = -50\,\mathrm{mV}$, the network's dendrites are permanently depolarized, and the difference between the target population activity and the control condition are less significant. Thus, the dendritic inhibition acts as a signal-to-noise control mechanism that allows the assembly to reactivate only upon the correct stimulus, maintaining the cells hyperpolarized otherwise.

The changes in the word assemblies activity do not entail that the network cannot process the phonemic stimuli correctly. Indeed, the connectivity matrix plays a significant role in the recognition task, and it may drive correct assembly reactivation against weaker non-linearity or higher dendritic noise. In the present analysis, the structure of the excitatory connections is frozen to the one obtained from the associative phase. Thus, it could be that the parameters change do not affect the task performance but only change the network dynamics. We measured the word recognition performance on all the parameters tested to test this hypothesis. The three panels in Fig.6*D* illustrate

the performance loss corresponding to networks simulated with each parameter variation. The score obtained for baseline values is used as the comparison term, and the loss is the difference between the score in the test condition and the score achieved in the baseline (κ-score ≈ 0.85). From the comparison of the three parameters swaps, it results that the inhibitory control plays the most significative role, with performance dropping of $60\,\%$ when the V_0 is increased to $-55\,\mathrm{mV}$. To match the same loss, the NAR has to diminish to 0 to 0.6 and the timescale to $2\,\mathrm{ms}$ to $10\,\mathrm{ms}$, which corresponds to absent *plateau potentials*. Crucially, the changes in the accuracy of the word recognition task match the ERPs' differences. indicating that the dynamics portrayed in Fig.6*B* are causal for the task.

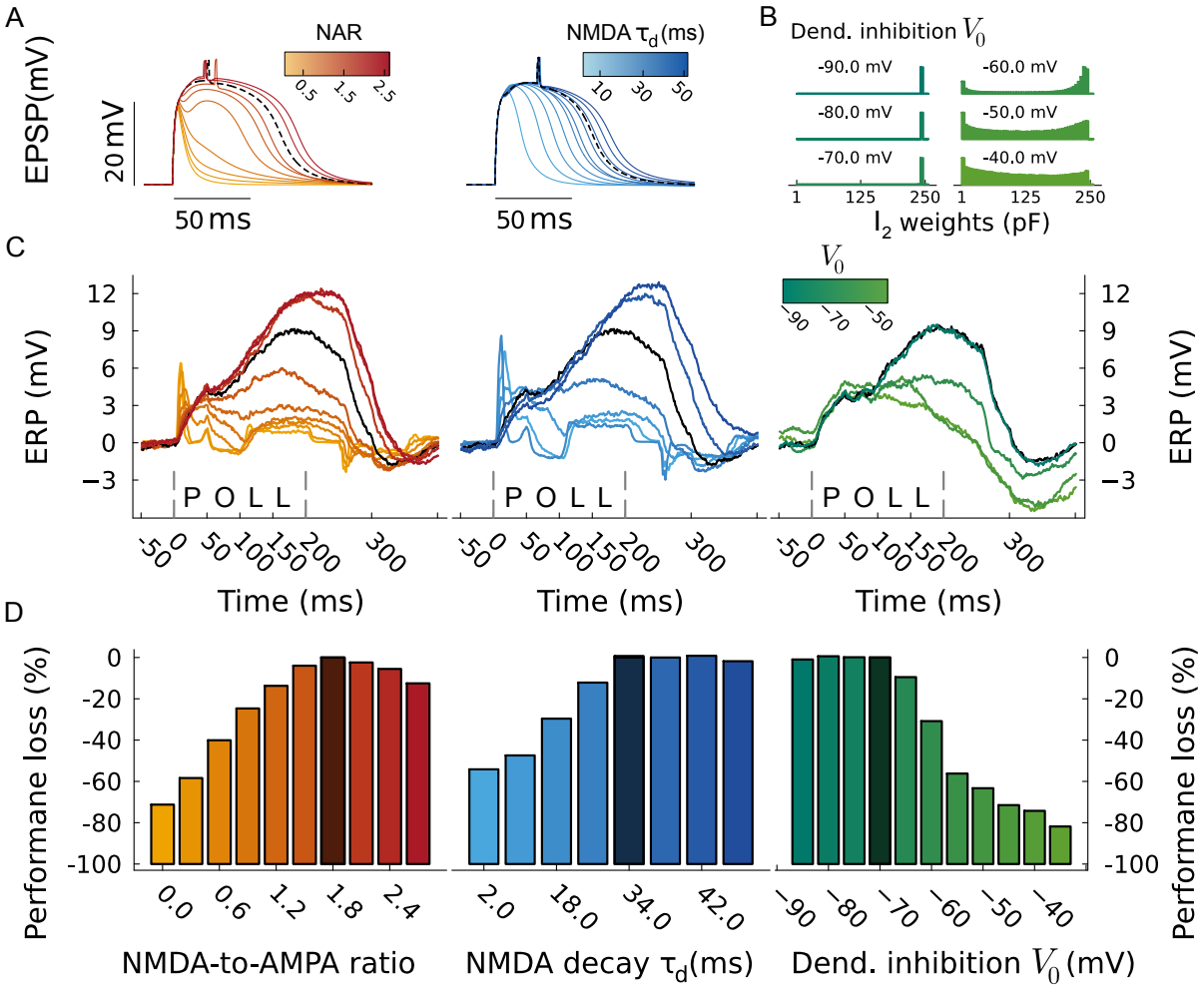

Figure 6: **Dendritic non-linearity and inhibitory control are necessary for correct word recognition** (**A**) EPSP of the dendritic membrane in Tripod neurons after the co-stimulation of a single synapse. The curves indicate the membrane potential for several NARs and NMDAR's decay timescale values. The dashed line indicates the model used, corresponding to the Tripod neuron with human synapses; NAR $= 1.8$; $\tau_d = 35\,\mathrm{ms}$. (**B**) Histograms of the inhibitory synaptic efficacies onto dendritic compartments, measured at the end of the recall phase simulation. The histograms refer to simulations with increasing V_0, the dendritic potential targeted by v-iSTDP. Weights accumulate on the maximum value for v-iSTDP for V_0 below $-70\,\mathrm{mV}$, the value used in all previous simulations. (**C**) Performance loss for parameter variations: NAR, τ_d, V_0. The bars with the darker shade correspond to the original model. Performances reduce up to $60\,\%$ when the dendritic non-linearity is short or absent. The changes are more dramatic for increases in V_0. (**D**) Event-related potentials (ERPs) upon presenting the phonemes *P, O, L, L*. The ERPs express the difference between the word *poll* assembly's average dendritic potential and the dendritic potential of the entire network. The ERP is measured for all the parameter variations, i.e., NAR, τ_d, V_0. The black solid line is the same for the three panels and refers to the baseline asymmetric Tripod model.

We reasoned that the loss in recognition could be caused by testing recall with different parameters than those used in the associative phase, hence the mismatch. Thus, to exclude this possibility, we modulated dendritic non-linearity and inhibitory control during the associative phase and again tested word recognition. The panels in Fig.7*A, B* show the κ score for a range of NARs, τ_s, and V_0 and indicate that early changes in the dendritic non-linearity do not solve the issue. In addition, the average strength of the phonemes-to-words connections has similar trends to the recognition score (Fig.7*C, D*). Altogether, the analysis suggests that both the NMDA spikes and tight inhibitory control are necessary during the associative phase for the formation of the phonemes-to-words connections which later support correct word recognition. As in the condition with unmodified network weights, the recognition drops for NAR below 1.0, for timescales shorter than $10\,\mathrm{ms}$ and inhibitory target membrane potential larger than $-60\,\mathrm{mV}$. An interesting difference in the previously discussed analysis is the interaction between the NMDARs' timescale and synaptic efficacy (NAR). The score can be maintained at the baseline level if the parameters counterbalance, i.e., increasing the decay timescale in the face of a smaller NAR or the reverse.

To conclude, we have shown that the non-linear excitability of cells' dendrites has direct consequences on the network's performance. This analysis elucidated the interaction between excitatory synaptic plasticity, dendritic non-linearity, and inhibitory control in supporting the formation of feedforward connections that allow the network to associate sequences of phonemes to words. These mechanisms can only be hosted in networks with dendrites; the segregated compartments endow the single cells with a long-lasting depolarization that results necessary for forming and accessing long-term memories through STDP.

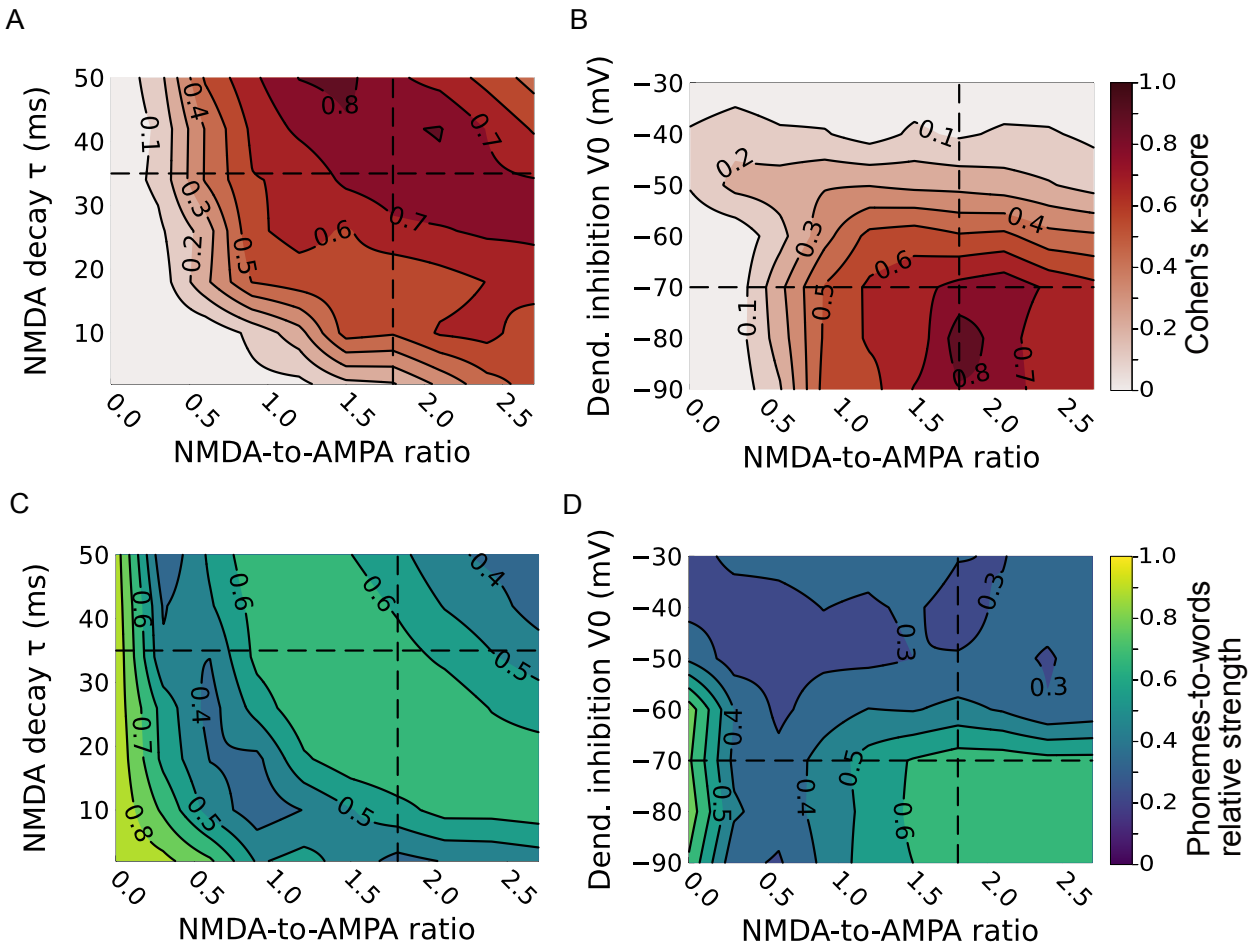

Figure 7: **Dendritic non-linearity and inhibitory control are necessary for correct word recognition** Cohen's κ-score of the Tripod asymmetric model with all NAR and τ_d pairs (**A**) and for pairs of NARs and V_0 (**B**). The model's parameters are varied in both the associative and recall phases. The dashed lines individuate the baseline model. The left panel shows that the score is stable if changes in the timescale are compensated with the NAR. Conversely, the interaction between the NAR and V_0 variables is weaker (right panel); the performances decay linearly on both axes. (**C**, **D**) Strength of the phonemes to word connections for the same parameter swap. Connections' strength is normalized over the effective connectivity matrix; larger values indicate a more significative average synaptic weight. The contour plot of the phonemes to words connection mirrors the score panels, indicating that the changes in the dendritic non-linearity and inhibitory control limit the formation of the connections, and these cause low recognition scores.

Discussion

Dendrites form most of the mammal's nervous tissue and constitute the spatial scaffold of brain networks; they primarily create the physical space for synapses. However, *in-silico*, these elements are unnecessary. Thus, do dendrites offer substantial computational contributions, or are they frills in the model that can be dispensed? The present work offered an answer to this question from a cognitive-computational perspective. It demonstrated that dendrites favor storing and retrieving memories with temporal structures in cell assemblies. We simulated two classes of networks, three dendritic and two point-neuron models. The networks received sparse and overlapping excitatory projections, representing words and phonemes. Through coordinated excitation and STDP plasticity, the targeted populations strengthened the synapses within their recurrent engrams. The recurrent phonemes connections are not sufficient to reactivate the word assemblies; our analysis showed that, for recognition to happen, structured and robust phonemes-to-word connections are necessary. Such feedforward connections are formed only in dendritic models suggesting that the dendritic structure may introduce a qualitatively different computational primitive in the network.

Recognizing sequences through dendritic non-linearity

Beyond its significance as a model of human language, the Tripod network presents a novel biologically-constrained model (BCM) for sequence processing. The perception of temporal patterns is a cognitive primitive; however, integration of information over time, and its maintenance, remain one of the great puzzles in neuroscience (Sreenivasan & D'Esposito, 2019; Fitz *et al.*, 2020; Gastaldi *et al.*, 2021; Spalla *et al.*, 2021; Mongillo *et al.*, 2008); let alone its interplay with synaptic long-term storage. In the present model, the segregated dendritic compartments of the Tripod neuron govern the interaction of short (dendritic) and long (synaptic) term memories. These two mechanisms have different computational trade-offs in terms of stability, robustness to noise, and duration of their transients (Chaudhuri & Fiete, 2016). The dendritic memory is encoded within a few tens of milliseconds and erased within hundreds. The possibility of accessing sequential information via the NMDARs' plateau-potential is a recent proposal (Branco *et al.*, 2010; Milojkovic *et al.*, 2005), which has been demonstrated successful in neural models with weak biological constraints (Hawkins & Ahmad, 2016; Leugering *et al.*, 2023; Bouhadjar *et al.*, 2022); the core mechanism is the feedforward pre-activation of a subset of the neuron's dendritic compartments, which facilitate the somatic firing when the remaining input sequences are presented. Our model implements the same dendritic process but leverages realistic synapses and membrane patch integration. The result is that the onset of depolarized dendritic states, or NMDA spikes, activates long-term potentiation in the dendritic

loci. Thus, the short-term dendritic memory engraves the temporal structure of the input in the recurrent and feedforward engrams. The synaptic changes happen on a scale 1000 time slower than the dendritic memory; stable connections form within $50\,\mathrm{s}$ to $100\,\mathrm{s}$ and do not return to the baseline values over the $5\,\mathrm{min}$ of the recall phase (Appendix B). The formation of stable memory via recurrent engrams forms in point-neuron models Litwin-Kumar & Doiron (2014), the novelty of our results is that the leveraging dendrites also feedforward connections are stable and can re-activate the short-term dendritic memories.

In cortical cells, dendrites host several synaptic clusters, organized spatially along the scaffold's spines; the co-activation of some of the clusters can elicit dendritic non-linearity (such as NMDA- or Calcium-spikes) and cause the neuron to burst (Larkum, 2022). Synaptic clusters allow the neuron to take part in multiple engrams because memories are maintained and reactivated independently (Legenstein & Maass, 2011; Kastellakis *et al.*, 2016; Kastellakis & Poirazi, 2019). Such engrams can be acquired via voltage-dependent STDP (Bono *et al.*, 2017; Bono & Clopath, 2017). The Tripod asymmetric model also has two autonomous synaptic pathways, and indeed it requires half of the specialized connections than the other dendritic models, retaining better performances on the task. In addition, preliminary analyses suggest that in virtue of this, its memory capacity scales better with the lexicon size. Orthogonally to the allocation of spatially distributed clusters, the segregated dendritic compartment offers better support for the vSTDP mechanism than the somatic membrane. The dendrites have a wider membrane potential range to interact with plasticity, which, in coordination with synaptic homeostasis, creates the conditions for synaptic stability and smooth weight dynamics in the associative phase (Bono *et al.*, 2017; Turrigiano, 2011; Triesch *et al.*, 2018).

Structured network connectivity achieved via Hebbian learning

Some of the early works on STDP (Clopath *et al.*, 2010; Gjorgjieva *et al.*, 2011; Markram *et al.*, 2011) showed that the statistics of the external projections can be encoded in the network structure of small networks via the unsupervised plasticity rule. Since then, extensive research has investigated how STDP can be leveraged to learn sequential stimuli on quick or behavioral timescales (Fiete *et al.*, 2010; Maes *et al.*, 2020; Rajan *et al.*, 2016; Duarte & Morrison, 2014; Reifenstein *et al.*, 2021; Riquelme *et al.*, 2023; Haga & Fukai, 2018; Gillett *et al.*, 2020). Much of this work aimed to provide a computational explanation for hippocampal and cortical replay, and the models acquire sequential patterns that can be promptly repeated upon the initial cues. However, it is crucial to distinguish between the word recognition problem we are discussing and cortical replay. The latter has more resemblance with the instantiation of motor programs, which is necessary to speech production rather than comprehension. Conversely, lexica access implies the recognition of a stimulus, whose

identity is jointly defined by the elements that compose it and their order. The goal is to activate a spatially-coded and time-compressed representation of the stimuli (Bagur *et al.*, 2022; Fox *et al.*, 2020). Two elements set our model apart from the previous studies. First, the word memory will be accessed as soon as its uniqueness point, which depends on the lexicon and not the word itself. Second, the sequences overlap for long intervals, which impose temporal integration on the scale of a hundred milliseconds. To our knowledge, few studies investigated whether STDP is sufficient for learning and recollecting word-like sequences. Crucially, they reach different conclusions. A BCM indicates no neat contribution of the glutamatergic plasticity Duarte & Morrison (2014). In contrast, a convolutional, feedforward network with weaker ties to biological mechanisms shows the formation of sharp spectro-temporal receptive fields (Dong *et al.*, 2018). We also attempted the same in the two control networks (LKD, DM), but we found that the point-neuron networks could not learn time-structured memories. In contrast, the Tripod network acquired time-structured memory with smooth and fast weights drift. The structure of our network connectivity was also somewhat surprising. The weights histogram has only three accumulation points for the synaptic weights, so each pre-post synaptic connection has one of these values. Nonetheless, the time structure of the word memories was encoded in the effective connectivity matrix rather than in the single connections, which also made the memory resilient to synaptic knockouts (data not shown).

Implementing cognitive-computational tasks with unsupervised plasticity algorithms comes with some difficulties, e.g., due to the intrinsic instability of the model and the lack of target values to optimize. Conversely, several algorithms have addressed and solved the sequence recognition problem with non-Hebbian learning rules. A classical case is the Tempotron (Gütig & Sompolinsky, 2006, 2009), a supervised plasticity rule for leaky-integrate-and-fire (LIF) that optimizes the pre-synaptic weights such that the neuron fires upon the correct class of spike patterns. Similarly, other approaches leverage reservoir computing (Buonomano & Maass, 2009; Nicola & Clopath, 2017; Duarte *et al.*, 2018; Fitz *et al.*, 2020) and backpropagation algorithms for SNN such as surrogate gradient (Bellec *et al.*, 2019; Zenke & Vogels, 2020). However, aside from the learning rule not being biologically grounded, these frameworks do not entail the formation of cell assemblies, and the sequence information is intertwined in the network latent space (Duarte & Morrison, 2014).

Inhibitory control governs memory access

The vSTDP rule supports the formation of cell assemblies in the Tripod network. Nonetheless, it creates a problem; strengthening excitatory connections makes the network prone to runaway excitation. The phenomenon is known as winner-take-all dynamics (Ermentrout, 1992) and destabilizes the network dynamics. Likewise Litwin-Kumar & Doiron (2014), we use inhibitory plasticity on the soma

as a rapid compensatory mechanism (RCM) that counteracts runaway activity (Zenke & Gerstner, 2017). In addition, the dendrites offer an additional RCM implemented via dendritic inhibition that does not act on the somatic firing directly. Rather, it selects the available dendritic pathways and can selectively silence dendritic branches. The on-off switch of dendritic compartments re-organizes the cell transfer function (Yang *et al.*, 2016) and prevents the overriding of synaptic memories, (Wilmes & Clopath, 2023). In our model, dendritic inhibition regulated both the signal-to-noise ratio of the dendritic response and helped maintain stable synapses during the recall phase. However, our implementation of the voltage-based iSTDP presents some limitations. The inhibitory plasticity drives all the weights to the maximum synaptic efficacy, so there is no pathway nor neuron specificity in the inhibitory connections. On the other hand, increasing the target dendritic potential releases the pressure from the synaptic rule and creates more heterogeneous landscapes in the histogram of synaptic weights; however, it also lowers the recognition score. Future work should explore the effect of the inhibitory plasticity mechanism that can modulate the dendritic response more selectively, this could be done by changing the plasticity rule with non-spike-time-dependent one (Pedrosa & Clopath, 2020; Miehl & Gjorgjieva, 2022), or changing the connection probability so to have more gabaergic synaptic contacts per dendrite.

Conclusions and future work

The present study proposed a highly biologically constrained model for spoken word recognition. We show that dendritic processes endow the network with the necessary computational primitives for storing and recalling word memories, short-term memory on the order of hundred milliseconds, and stable long-term memories via synaptic clustering. Beyond the limitations hitherto discussed, we think of two main improvements that can be made to the model to account for human word recognition. First, word populations should more actively compete among themselves. Blanket feedback inhibition usually works, but words with overlapping features tend to associate despite their mutual contribution, which is harmful to recognition. Anti-Hebbian plasticity implemented between inhibitory and excitatory neurons could introduce this active competition (Luz & Shamir, 2012) and carve part of the lexical memory in the inhibitory connections. Second, the next step is introducing semantic feedback in the network, such that lexical-decision is made based on phonetic and semantic information. Future work should investigate the pathways through which contextual information can be delivered to the network (feedforward pathways, apical dendrites, disinhibitory circuits, etc..., (Naumann *et al.*, 2022)) and how this will affect the network dynamics.

Methods

In this study, we compare two classes of network models, composed either of single-compartment neurons or few-compartment dendritic models, that is, Tripod neurons. In the following section, we first describe the neuronal dynamics of the neuron models, comprehending the neuron's equations and the synaptic dynamics. Then we describe the architecture, the recurrent connectivity, and the plasticity rules of each network class. Finally, we detail the experimental protocol used to associate and recall words.

Neurons and synapses

The networks are composed of two types of neurons, excitatory and inhibitory. The excitatory neurons have been modeled as three-compartment neurons in the dendritic networks and single-compartment units in the soma-only networks. The inhibitory neurons also have two sub-types, modeled as single-compartment neurons. The models and parameters used vary for each network, and the variations are constrained to ensure stability during the simulation. For the two previously published networks, namely the L2/3 cortical circuit by Duarte & Morrison (2019) and the balanced network by Litwin-Kumar & Doiron (2014), we used models and parameters identical with the original publications, if not differently stated. In the following particular attention will is reserved for the models of the dendritic network, as they are the most complex and constitute the novel contribution of the present work.

Excitatory neurons

There are two types of excitatory cells in the networks, soma-only and dendritic Tripod neurons. The Tripod neurons comprise three electronically segregated compartments: an axosomatic compartment and two membrane patches. A detailed description of the model is presented in Quaresima *et al.* (2022), which also constitutes the second chapter of the present manuscript. The somatic compartment is modeled with the adaptive exponential integrate-and-fire model (Brette & Gerstner, 2005), it is two-dimensional and comprises the dynamics of membrane potential and adaptive hyper-polarizing currents (Eqs.2,3).

$$
\begin{aligned}
C_m^s \frac{dV^s}{dt} &= -g_m^s \left[(V^s - V_r) + \Delta_T \exp \frac{V^s - \theta}{\Delta_T} \right] \\
&\quad - \sum_k g_k(t)(V^s - E_k) - w + \sum_{d=1,2} g_{ax}^d (V^d - V^s) \qquad (2) \\
\tau_w \frac{dw}{dt} &= -w + a(V^s - V_r) + b \sum \delta(t - t^f) \qquad (3) \\
\text{if} \quad V^s > \theta \quad &\rightarrow V^s = V_r \quad ; \qquad (4)
\end{aligned}
$$

$$
\begin{aligned}
C_m^d \frac{dV^d}{dt} &= -g_m^d (V^d - V_r) \\
&\quad - \sum_k g_k(t)(V^d - E_k) - g_{ax}^d (V^d - V^s); \quad \mathrm{d} = 1, 2 \qquad (5)
\end{aligned}
$$

The leak conductance g_m^s defines the electrical permeability of the somatic membrane, C_m^s its capacitance, and g_k the set of variable synaptic conductances. Spikes occur at times t^f when the potential V^s exceeds a threshold θ (Eq.4). The membrane potential V^s is reset to V_r after a spike, and the adaptation current w is increased by a constant value b (Eq.3). Parameters for the somatic compartment were fixed and set to the values used in Brette & Gerstner (2005) and are listed in Table1 The term g_{ax}^d in Eq.2 accounts for the axial conductance of the two dendritic compartments. The passive membrane-patch equation Eq.5 (Koch, 1998) governs the dendritic compartments. Its electrical properties depend on the dendritic geometry combined with physiological parameters. The physiology considered concerns the membrane permeability, the membrane capacitance, and the axial impedance to the soma ($\bar{g}_m$ =25.6 µS cm^2; $\bar{C}_m$ =0.5 µF/cm^2; $\bar{g}_{ax}$ =0.5 S cm^2) (Koch, 1998; Eyal *et al.*, 2016). In the present study, we use the parameters obtained from human pyramidal cells Eyal *et al.* (2016) because they show increased sequence processing capacities. The dendritic diameter is fixed at $4\,\mathrm{mm}$, and dendritic electrical properties are determined only by the dendritic length L_d, in agreement with the equations:

$$
\text{Dendrite}\,[l, d = 4\,\mu\mathrm{m}] \quad = \quad \begin{cases} C_m = & \pi l d \bar{c}_m \\ g_m^d = & \pi l d \bar{g}_m \\ g_{ax}^d = & \frac{\pi}{4} \frac{d^2}{l} \bar{g}_{ax} \end{cases}
$$

In all the dendritic models tested in the present study, the dendritic compartments have heterogeneous lengths in the range $150\,\mu\mathrm{m}$ to $400\,\mu\mathrm{m}$.

The current flow between soma and dendrites depends on g_{ax} $(V_d > V_s)$. It usually is positive (dromic); conversely, during somatic firing, the soma compartment's membrane potential is clamped to $20\,\mathrm{mV}$ for the duration of the spike (spike width, $1\,\mathrm{ms}$). During the spike interval, the current flow

is antidromic ($V_d < V_s$), and the current flowing to the dendrites acts as a backpropagating axon potential (the dendritic and soma potentials are illustrated in Fig.1A, the inset shows the potentials' dynamics during somatic firing).

The Tripod neurons express AMPA, NMDA, $GABA_A$, and $GABA_B$ receptors on the dendrites. The synaptic transmission is modeled with a double-exponential model (Roth & van Rossum, 2009). The synaptic conductance g_k is given by:

$$g_k(t) = \bar{g}_k^{\text{syn}} \mathcal{N}_k \left(\exp\left(-\frac{t-t_0}{\tau_k^r}\right) - \exp\left(-\frac{t-t_0}{\tau_k^d}\right) \right) \tag{6}$$

$$\mathcal{N}_k = \left(-e^{-t^{peak}/\tau^r} + e^{-t^{peak}/\tau_d^r}\right)^{-1} \tag{7}$$

$$t_k^{peak} = \frac{\tau^d \tau^r}{\tau^d - \tau^r} \ln \frac{\tau^d}{\tau^r} \tag{8}$$

with $k \in \{\text{AMPA}, \text{NMDA}, \text{GABA}_\text{A}, \text{GABA}_\text{B}\}$ indicating that each receptor has specific parameters. The equation describes the rise and decay of the receptor conductance g_k. The timescales of rise and decay are given by τ_r and τ_d while the amplitude of the curve is defined by the maximal conductance parameter g_{syn}. To ensure that the amplitude equals $\bar{g}_k^{\text{syn}}$, the conductance was scaled by the fixed normalization factor $\mathcal{N}_k$. In the Tripod neuron we fixed the NMDA-to-AMPA (NAR) ratio as in human-like cells in Eyal *et al.* (2018); Quaresima *et al.* (2022); Eyal *et al.* (2018). The GABAergic receptors on the dendrites express both $GABA_A$ and $GABA_B$, whereas the soma only expresses AMPA and $GABA_A$. The synaptic parameters of the Tripod neuron are listed in Table3.

A Tripod neuron is portrayed in Fig.1A. The neuron is composed of a soma and two dendrites of different lengths (basal, proximal 150 and basal distal $400\,\mu\text{m}$). Each dendrite expresses AMPA, NMDA, $GABA_A$, and $GABA_B$ receptors. The dendritic receptors' dynamics and relative maximal conductance are illustrated in the panel below. The bottom right panel portrays the dynamics during balanced excitatory-inhibitory stimulation of the Tripod model with distal-proximal dendrites. The inset shows the antidromic current following a somatic spike. Note, the superscript s is used throughout the manuscript to refer to variables and parameters of somatic compartments, whereas the superscript d refers to dendritic compartments.

The excitatory cells of the point-neuron networks are also implemented with the AdEx model (Eq.2). However, they introduce an additional adaptive parameter, expressed by a variable membrane potential threshold θ, which increases after each spike. The threshold is used to facilitate voltage-dependent synaptic plasticity during bursting intervals, it is governed by the equation:

$$\tau_\theta \frac{d\theta}{dt} = -(\theta - \theta_0) + \theta_+ \sum \delta(t - t^f) \tag{9}$$

$$\tag{10}$$

For the network borrowed from Duarte & Morrison (2019), the soma hosts four types of receptors, as in the Tripod. Conversely, the model from Litwin-Kumar & Doiron (2014) only has AMPA and $GABA_A$ receptors. The parameters of the point-neuron AdEx models and synapses are listed in Table1 and Table3 .

Inhibitory interneurons

Inhibitory neurons are modeled as single, isopotential compartments with leaky integrate-and-fire dynamics. The membrane potential V_i is governed by the following equations:

$$\begin{aligned} C_m \frac{dV_i}{dt} &= -g_m(V_i - V_r) - \sum_k g_k(t)(V_i - E_k) - w \\ \tau_w \frac{dw}{dt} &= a(V_i - V_r) - w \end{aligned} \tag{11}$$

$$\text{if } V > V_T \quad \rightarrow \quad V = V_r \tag{12}$$

Following previous modeling efforts (e.g. Park & Geffen (2020); Yang *et al.* (2016); Duarte & Morrison (2019)) and consistently with physiological description (Tremblay *et al.*, 2016), we implement two types of inhibitory interneurons: fast-spiking and non-fast-spiking. The first class corresponds to a set of well-defined parvalbumin-expressing (PV) GABAergic neurons, among which basket-cells; these cells primarily target the perisomatic regions and are indicated in this work as I_1population. Conversely, the second group includes parvalbumin-negative GABAergic neurons expressing somatostatin. For notational simplicity, we refer to the second group as I_2. Following Duarte & Morrison (2019), the fast-spiking neurons are modeled as not adaptive while the I_2 neurons are. The two classes differ in LIF model parameters (Table2), and synaptic properties (Table3). In our model, they also differ for the compartment of the Tripod neuron they connect to. I_1 neurons target the soma compartment, and their synapses express fast $GABA_A$ receptors. I_2 neurons, on the other hand, target the dendrites and, due to the additional slow $GABA_B$ component of the dendritic synapses, have a longer timescale. The inhibitory neurons' synapses express AMPA, $GABA_A$ receptors, and the synaptic transmission is modeled with a double-exponential model. The parameters for the inhibitory neurons are borrowed from Duarte & Morrison (2019) and are reported in Table1 and Table4.

Network architectures

We implement three main classes of networks that vary for the number and types of neurons considered and for the connectivity between them. The novel network model is composed of Tripod neurons as excitatory cells. The network counts 2000 Tripod neurons with dendritic lengths uniformly distributed in the range$150\,\mu\mathrm{m}$ to $400\,\mu\mathrm{m}$; along with 175 fast-spiking neurons, and 325 slow-spiking neurons. The proportion of inhibitory neurons follows (Duarte & Morrison, 2019). The network lacks geometrical structure and is conceived as a local cortical circuit. The Tripod neurons connect reciprocally through the dendrites and to the soma of inhibitory neurons. The fast-spiking neurons target the Tripod neurons' soma, while the slow-spiking neurons target the dendrites. Inhibitory neurons also connect reciprocally. All connections have a probability of $p = 0.2$. The connections targeting the two dendrites of a Tripod neuron are drawn independently from each other. Hence some neurons will connect through one dendrite only, and others will connect through both dendrites (pre-synaptic spikes arrive separately on both compartments). The connectivity pattern described is the default network configuration and is identified as the *asymmetrical* network. As a result, the Tripod network has 10 different connection types; the Tripod-Tripod and I_2-Tripod are doubled because they account for the two dendrites. An illustration of the Tripod network connectivity is offered in Fig.1*B*, and the full parameters are listed in Table7.

The remaining two networks are composed of point-neuron models, illustrated in the right panel of Fig.1*B*. We implemented an exact copy of the network described in Litwin-Kumar & Doiron (2014). The network is composed of 4000 excitatory neurons and 1000 inhibitory neurons. Then neurons are sparsely connected with a probability, $p = 0.2$, the synaptic weights change according to the four connection types (exc-exc, inh-inh, exc-inh, inh-exc), listed in Table6. The network is dubbed LKD network. The third network is composed of 2000 excitatory neurons, 200 fast-spiking neurons and 200 slow-spiking neurons. The network is modeled after the non-heterogeneous model of cortical L2/3 (Duarte & Morrison, 2019). Neurons are connected with probability and connectivity strength that depends on the type of connection, for a total of nine types of connections. The network is dubbed the DM network. The networks' neuron types and sizes are listed in Table6; their connection probability is presented in Table7.

Synaptic learning rules

The synapses connecting excitatory neurons reciprocally (EE) and from the inhibitory to the excitatory neurons (IE) are subject to spike-time-dependent plasticity (STDP). The two plasticity rules are illustrated in Fig.1*C*.

The glutamatergic synapses undergo a voltage-dependent STDP rule (Clopath *et al.*, 2010). The

voltage-dependent STDP (vSTDP) strengthens the connection when, following a pre-synaptic spike, the post-synaptic neuron's membrane potential is depolarized. The equations governing the vSTDP rule are:

$$\begin{aligned} J_{ij}^{EE} \leftarrow \quad J_{ij}^{EE} \quad & +dtA_{LTD} \quad s_j \left[u_i(t) - \theta^-\right]_+ \\ & +dtA_{LTP} \quad x_j \left[V_i - \theta^+\right]_+ \left[v_i(t) - \theta^-\right]_+ \end{aligned} \tag{13}$$

$$\tau_u du_j = \quad -(u_j - V_j)dt \tag{14}$$

$$\tau_v dv_j = \quad -(v_j - V_j)dt \tag{15}$$

$$\tau_x dx_i = \quad -x_i dt + \delta(t - t_f) \tag{16}$$

The J_{ij}^{EE} indicates the connection between the excitatory neurons, post-synaptic i, and pre-synaptic j. The variable $s_j = \sum_t \delta(t - t_j^f)$ is the spike train of the presynaptic neuron, and x_i is the filtered spike train of the post-synaptic cell. In addition, the variables V_j, u_j, v_j are the membrane potential, the slow, and the fast-filtered membrane potential of the targeted compartment. The two thresholds (θ^+, θ^-) mark the region where long-term potentiation (LTP) and long-term depression (LTD) occur, coupled with the learning rates A_{LTP} and A_{LTD}. In the point-neuron models, the vSTDP is applied to the soma ($V_i = V_i^s$), while in the Tripod neurons, the vSTDP is applied to the dendrites ($V_i = V_i^{d=1,2}$). For each couple of connected Tripod neurons, there are two values $J_{ij}^{E^1E}$ and $J_{ij}^{E^2E}$, one for each dendrite, and they are updated independently. In addition, the vSTDP rule includes a homeostatic mechanism that maintains the sum of pre-synaptic weights fixed; the homeostatic rule follows the prescription of synaptic scaling (Turrigiano, 2011; Tetzlaff *et al.*, 2011; Triesch *et al.*, 2018) and is multiplicative. Conversely, the homeostatic rule in Litwin-Kumar & Doiron (2014) is additive. The multiplicative homeostatic rule is implemented as follows:

$$J_{ij}^{EE} \leftarrow J_{ij}^{EE} \frac{\sum_k J_{ik}}{\sum_k J_{ik}^0} [1ex] \tag{17}$$

where J_{ij}^0 is the initial value of the synaptic weight. Homeostasis is applied every $20\,\mathrm{ms}$.

For the connections from inhibitory to excitatory neurons, the networks implement the iSTDP rule from Vogels *et al.* (2011). The iSTDP governs the network stability by regulating the inhibitory synaptic strength. It is a necessary ingredient for the formation of assemblies in the original LKD network because it prevents winner-takes-all dynamics. We implemented the same mechanism for the I_1or fast-spiking interneurons in the three networks, ($E^s\mathrm{I}_1$); the iSTDP maintains the firing rate of the excitatory cells by modulating the synaptic strength of the inhibitory pathways. In addition, we modified the iSTDP rule for the slow interneuron type such that they can control the

dendritic non-linearity of the Tripod's dendrites. The I_2iSTDP targets the membrane potential of the post-synaptic neuron, rather than its firing rate. We refer to it as voltage-dependent iSTDP or v-iSTDP. It applies to synapses on the soma ($E[s]I_2$) in the DM and to synapses on the dendrites ($E[d]I_2$) in the Tripod network, it is absent in the LKD network. The two iSTDP rules follow the equations:

$$\text{if} \quad j \in I_1, \quad i \in \text{Exc}:$$
$$J_{ij}^{E[s]I} \leftarrow J_{ij}^{E[s]I} + \eta dt \left[x_j s_i^E + (x_i - \alpha)\, s_j^I \right] \tag{18}$$
$$\text{if} \quad j \in I_2, \quad i \in \text{Exc}:$$
$$J_{ij}^{E[d]I} \leftarrow J_{ij}^{E[d]I} + \eta dt \left[x_j s_i^E + \left(v_i^d - V_0\right) s_j^I \right] \tag{19}$$

where $J_{ij}^{E[s]I}$ and $J_{ij}^{E[d]I}$ are the synaptic weights from the inhibitory to the excitatory neurons, respectively, for the soma and the dendrites. The s_j^I is the spike train of the inhibitory neuron j, while x_i and v_i^d are the filtered spike trains of the excitatory neuron i in the soma and the dendrites, respectively. The α is the target firing rate for the iSTDP of I_1neurons, while the V_0 is the target membrane potential for the iSTDP of I_2neurons. Finally, η is the learning rate. The parameters are listed in Table5.

Stimuli

Phonemes and word projections The networks were stimulated through external projections encoding phoneme or word inputs. Each set of projections targets $\rho = 5\,\%$ of the network, creating an assembly of co-activated cells. The assemblies were approximately 100 neurons in the Tripod and DM network and 200 in the LKD network. The connections are distributed randomly to the network's neurons; therefore, the assemblies lack any geometrical structure. The phoneme and word assemblies were stimulated with spike trains. All pre-synaptic projections have fixed firing rate ($8\,\mathrm{kHz}$) and synaptic efficacy ($2.78\,\mathrm{pF}$ for point-neuron models, $20\,\mathrm{pF}$ for the Tripod model). These numbers were optimized to have a strong neuronal response but not unstabilize the network. Stimuli spike trains are drawn from a Poisson distribution at every presentation, with the rate being the same for all projections; as a consequence, the stimuli cannot be interpreted as rate- or time-coded. The only information available to the networks to discriminate between the stimuli is the identity of the activated assemblies, which constitutes a spatial code. Depending on the lexicon used, each network has 20 to 35 distinct projections, each representing a phoneme or a word. The phoneme stimuli were presented for $50\,\mathrm{ms}$ each, and the word stimuli were presented for the entire duration of the stimulus. Each word was followed by a silent interval without projections, the pause lasts $50\,\mathrm{ms}$.

Sequence association and recall The stimulation protocol consists of two main phases during which the phonemes to words associations are first established and then tested, both phases last $5\,\mathrm{min}$. During the associative phase, we present the network with external stimuli on both phonemes and word projections. The stimuli are presented in sequence; one word is drawn randomly, and the corresponding phoneme projections are activated. The phonemes populations are activated for $50\,\mathrm{ms}$ each, while the word ensembles are activated for the entire duration of the word (that is $50\,\mathrm{ms}$ times the number of phonemes contained in the word). The number of words presented in the two phases depends on the average length and the interval duration; for the experiments discussed here, each word is presented 50 to 100 times. All words are presented approximately the same number of times; therefore, there are no frequency effects. The phase duration has been chosen to be long enough for the synaptic weights to converge, but short enough to allow for several simulations and maintain the study reproducible.

Word recognition measure

We individuate two salient time points for word recognition for every phoneme sequence presented during the recall phase. One is the offset of the word (offset point, OP), that is, the last phoneme of the sequence. The other is the word's uniqueness point (UP). The UP corresponds to the phoneme that permits distinguishing a word from any other in the lexicon (Marslen-Wilson, 1973); when the word is contained in another word (such as *gold* and *golden*), the UP is the phoneme following the last shared phoneme, i.e., the silence after the end of the word in case of *gold*. As in the previous section, we estimate word recognition for every trial by measuring the firing rate in the adjacent interval. We associate the OP with an interval lasting the entire length of the phonemes' sequence, from the word onset to the word offset ($50\,\mathrm{ms}$ times the length of the word); this measure is the same one portrayed in Fig.2*D*. Conversely, the UP interval is shorter and corresponds to the duration of one single phoneme ($50\,\mathrm{ms}$); in this case, the interval starts and ends with the UP phoneme. An example of the uniqueness and offset points of the word *golden*, and their intervals, is illustrated in Fig.3*A*; the offset and UP time-points are marked on the phoneme sequence, respectively, by the blue and green arrow marks. Because words are longer than one phoneme, the interval associated with the offset is, on average, longer than the one associated with the UP.

Because the reactivation of word assemblies appears delayed to the presented phoneme sequence, we introduced a variable time shift in the measure of the firing rate to individuate the optimal interval. For both the offset and UP measures, the measured intervals were probed with delays from $-150\,\mathrm{ms}$ to $200\,\mathrm{ms}$, sampling every $5\,\mathrm{ms}$. The delayed interval is also illustrated in Fig.3*A*; the applied time shift is indicated with the orange arrow. We computed the lexicon confusion matrix

(CM) for any given delay. The confusion matrix indicates which word is reactivated when a target word is presented; the sum of each column is one and corresponds to the fraction of trials (column) assigned to each possible target (row). Values close to one on the CM's diagonal indicate correct recognition, while off-diagonal elements indicate mistaken phoneme sequences. As explained in the Methods, the average recognition score is computed from the confusion matrix using Cohen's κ to correct against chance. The confusion matrices of all lexica, computed with $100\,\mathrm{ms}$ delay, for all dictionaries and averaged over the ten samples are reported in the supplementary materials (SI Fig.3).

Table 1: Parameters for the axosomatic compartment (Quaresima *et al.*, 2022; Litwin-Kumar & Doiron, 2014; Duarte & Morrison, 2019).

Symbol	**Description**	**Tripod**	**LKD**	**DM**	**Unit**
V_r	Resting membrane potential	-70.6	-70.	-76.43	mV
g_L	Membrane leak conductance	40	15.	4.64	nS
C_m	Membrane capacitance	281	300	116.5	pF
V_T	Threshold potential	-50.	-52.	-44.45	mV
u_{th}	Spike onset threshold	0	0	0	mV
u_r	Reset potential	-55	-60	-54.18	mV
Δ_T	Slope factor	2	2	2	mV
a	Adaptation conductance	4	4	4	nS
τ_w	STA[1] timescale	144	150	144	ms
b	STA[1] increment	80.5	80.5	80.5	pA
t_{up}	Spike width ($20\,\mathrm{mV}$)	1	1	1	ms
t_{ref}	Refractory period	2	2	2.1	ms

[1] Spike-triggered adaptation

Table 2: Parameters for the axosomatic compartment of the inhibitory neurons for the three network models (Litwin-Kumar & Doiron, 2014; Duarte & Morrison, 2019). The inhibitory neurons of the Tripod network have the same parameters as those in the Duarte and Morrison model.

Symbol	**Description**	**I_1 (LKD)**	**I_1 (DM)**	**I_2 (DM)**	**Unit**
g_L	Membrane leak conductance	15	9.75	4.61	nS
C_m	Membrane capacitance	300	104.52	102.86	pF
V_r	Resting membrane potential	-62	-64.33	-61	mV
V_T	Threshold potential	-52	-38.97	-34.4	mV
u_{th}	Spike onset threshold	V_T	V_T	V_T	mV
u_r	Reset potential	-57	-57.47	-47.11	mV
a	Adaptation conductance	-	-	144	nS
τ_w	STA[1] time scale	-	-	4	ms
b	STA[1] increment	-	-	80.5	pA
t_{up}	Spike width (20 mV)	1	1	1	ms
t_{ref}	Refractory period	1	0.5	1.3	ms

[1] Spike-triggered adaptation

Table 3: Synaptic parameters for the excitatory and inhibitory neurons in the Tripod network.

Receptor	E_r(mV)	τ_{rise}(ms)	τ_{decay}(ms)	g_{syn}(nS)	γ $(\frac{1}{\text{mV}})$
Tripod soma					
AMPA	0.0	0.26	2.	0.73	-
$GABA_A$	-70	0.1	15	0.38	
Tripod dendrites					
AMPA	0.0	0.26	2.	0.73	-
NMDA	0.0	8.	35.	$0.73 \cdot \text{NAR}$	0.075
$GABA_A$	-70	4.8	29.0	0.27	
$GABA_B$	- 90	30	400	0.006	
I_1soma					
AMPA	0.0	0.18	0.7	1.04	-
$GABA_A$	-75.0	0.19	2.5	0.84	-
I_2soma					
AMPA	0.0	0.18	1.8	0.56	-
$GABA_A$	-75.0	0.19	5.0	0.59	-

Table 4: Synaptic parameters for the excitatory and inhibitory neurons in the point neuron models (Duarte & Morrison, 2019; Litwin-Kumar & Doiron, 2014).

Receptor	E_r(mV)	τ_{rise}(ms)	τ_{decay}(ms)	g_{syn}(nS)	γ $\left(\frac{1}{\text{mV}}\right)$
LKD soma					
AMPA	0.0	1.0	6.0	1.0	-
$GABA_A$	-75.0	0.5	2.0	1.0	-
LKD I_1					
AMPA	0.0	1.0	6.0	1.0	-
$GABA_A$	-75.0	0.5	2.0	1.0	-
Duarte soma					
AMPA	0.0	0.25	2.0	0.73	-
NMDA	0.0	0.99	100.0	0.16	-
$GABA_A$	-75.0	0.5	6.0	0.26	-
$GABA_B$	-90.0	30.0	100.0	0.01	-
Duarte I_1					
AMPA	0.0	0.1	0.7	1.6	-
NMDA	0.0	1.0	100.0	0.0	-
$GABA_A$	-75.0	0.1	2.5	1.0	-
$GABA_B$	-90.0	25.0	400.0	0.02	-
LKD I_2					
AMPA	0.0	0.2	1.8	0.8	-
NMDA	0.0	1.0	100.0	0.01	-
$GABA_A$	-75.0	0.2	5.0	0.7	-
$GABA_B$	-90.0	25.0	500.0	0.02	-

Table 5: Parameters for the v-STDP, iSTDP, and v-iSTDP plasticity rules. The v-STDP parameters for the point-neuron models (LKD and DM) are obtained from Clopath *et al.* (2010), and identical to the original study by Litwin-Kumar & Doiron (2014); for the dendritic model is obtained from Bono & Clopath (2017). The iSTDP and v-iSTDP rules follow the parameters by Vogels *et al.* (2011) with variations on the learning rate η and the target of the homeostasis for the two interneurons.

Symbol	Description	LKD	DM	Tripod	Unit
v-STDP					
a^-	LTD learning rate	8.0E-05	8.0E-05	4.0E-05	Hz
a^+	LTP learning rate	1.4E-04	1.4E-04	1.4E-04	Hz
θ^-	LTD threshold	-70	-70	-40	mV
θ^+	LTD threshold	-49	-49	-20	mV
τ_u	slow membrane filter	10.0	10.0	15.0	ms
τ_v	fast membrane filter	7.0	7.0	45.0	ms
τ_x	spike train filter	15.0	15.0	20.0	ms
τ_s	homeostasis timescale	20.0	20.0	20.0	ms
j^-	minimum synaptic weight	2.78	2.78	2.78	pF
j^+	maximum synaptic weight	21.0	4.0	41.4	pF
iSTDP					
η	learning rate	1.0	0.2	0.2	Hz
j^+	maximum synaptic weight	243.0	243.0	243.0	pF
j^-	minimum synaptic weight	2.78	2.78	2.78	pF
I_1neurons					
r_0	target firing rate	5	10	10	Hz
τ_y	rate filter timescale	20.0	20.0	20.0	ms
I_2neurons					
vd	target dend. potential	-	-70	-70	mV
τ_d	dend. potential filter timescale	-	5.0	5.0	ms

Table 6: Number and types of neurons for each model, intensity of the background noise.

Parameter	**Description**	**Tripod**	**DM**	**LKD**	**Unit**
Neurons	Number of neurons	2500	2500	5000	-
Types	Neuron types	ET, I_1, I_2	E, I_1, I_2	E, I_1	-
Excitatory Noise	Excitatory noise	4	4	4	kHz
SST Noise	SST noise	0.5	0.5	0.5	kHz
PV Noise	PV noise	0.5	0.5	0.5	kHz

Table 7: Connectivity for the three network models. The connections subject to STDP and iSTDP are in bold text, the synapses are listed in the couple post-pre synaptic and the squared parentheses indicate the post-synaptic target compartment. The numbers indicate respectively the connection probability and initial synaptic strength. The synaptic strength is in parenthesis, expressed in pF.

Synapse	**LitwinKumar & Doiron**	**Duarte & Morrison**	**Tripod**
E[d]E	-	-	0.2 (10.78)
E[s]E	0.2 (2.76)	0.168 (0.45)	-
I_1E	0.2 (1.27)	0.575 (1.65)	0.2 (5.27)
I_2E	0.2	0.244 (0.638)	0.2 (5.27)
E[s]I_1	0.2 (48.7)	0.6 (5.148)	0.2 (15.8)
E[d]I_2	-	-	0.2 (15.8)
$\mathbf{E}^{s}I_2$	-	0.465 (4.85)	-
I_2I_2	-	0.381 (0.83)	0.2 (16.2)
I_1I_1	0.2 (16.2)	0.55 (2.22)	0.2 (16.2)
I_1I_2	-	0.379 (1.47)	0.2 (1.47)
I_2I_1	-	0.24 (1.4)	0.2 (0.83)

References

Almeida-Filho DG, Lopes-dos-Santos V, Vasconcelos NAP, Miranda JGV, Tort ABL & Ribeiro S (2014). An investigation of Hebbian phase sequences as assembly graphs. *Front Neural Circuits* **8**, 34.

Bagur S, Bourg J, Kempf A, Tarpin T, Bergaoui K, Guo Y, Ceballo S, Schwenkgrub J, Puel J, Bourien J & Bathellier B (2022), Emergence of a time-independent population code in auditory cortex enables sound categorization and discrimination learning. preprint, Neuroscience.

Bellec G, Scherr F, Subramoney A, Hajek E, Salaj D, Legenstein R & Maass W (2019). A solution to the learning dilemma for recurrent networks of spiking neurons. *Nat Commun* .

Bono J & Clopath C (2017). Modeling somatic and dendritic spike mediated plasticity at the single neuron and network level. *Nat Commun* **8**.

Bono J, Wilmes KA & Clopath C (2017). Modelling plasticity in dendrites: from single cells to networks. *Curr Opin Neurobiol* **46**, 136–141.

Bouhadjar Y, Wouters DJ, Diesmann M & Tetzlaff T (2022). Sequence learning, prediction, and replay in networks of spiking neurons. *PLOS Computational Biology* **18**, e1010233.

Branco T, Clark BA & Häusser M (2010). Dendritic discrimination of temporal input sequences in cortical neurons. *Science* **329**, 1671–1675.

Brette R & Gerstner W (2005). Adaptive exponential integrate-and-fire model as an effective description of neuronal activity. *J Neurophysiol* **94**, 3637–3642.

Buonomano DV & Maass W (2009). State-dependent computations: spatiotemporal processing in cortical networks. *Nat Rev Neurosci* **10**, 113–125.

Buzsáki G (2010). Neural Syntax: Cell Assemblies, Synapsembles, and Readers. *Neuron* **68**, 362–385.

Chan AM, Dykstra AR, Jayaram V, Leonard MK, Travis KE, Gygi B, Baker JM, Eskandar E, Hochberg LR, Halgren E & Cash SS (2014). Speech-Specific Tuning of Neurons in Human Superior Temporal Gyrus. *Cereb Cortex* **24**, 2679–2693.

Chaudhuri R & Fiete I (2016). Computational principles of memory. *Nat Neurosci* **19**, 394–403.

Clopath C, Büsing L, Vasilaki E & Gerstner W (2010). Connectivity reflects coding: a model of voltage-based STDP with homeostasis. *Nat Neurosci* **13**, 344–352.

Cohen Y, Shen J, Semu D, Leman DP, Liberti WA, Perkins LN, Liberti DC, Kotton DN & Gardner TJ (2020). Hidden neural states underlie canary song syntax. *Nature* , 1–6.

Cone I & Shouval HZ (2021). Learning precise spatiotemporal sequences via biophysically realistic learning rules in a modular, spiking network. *eLife* **10**, e63751.

Dong M, Huang X & Xu B (2018). Unsupervised speech recognition through spike-timing-dependent plasticity in a convolutional spiking neural network. *PLoS ONE* **13**, e0204596.

Dringenberg HC (2020). The history of long-term potentiation as a memory mechanism: Controversies, confirmation, and some lessons to remember. *Hippocampus* **30**, 987–1012.

Duarte RCF & Morrison A (2014). Dynamic stability of sequential stimulus representations in adapting neuronal networks. *Front Comput Neurosci* **8**.

Duarte R & Morrison A (2019). Leveraging heterogeneity for neural computation with fading memory in layer 2/3 cortical microcircuits. *PLoS Comput Biol* **15**, e1006781.

Duarte R, Uhlmann M, den van Broek D, Fitz H, Petersson KM & Morrison A (2018), Encoding symbolic sequences with spiking neural reservoirs. In *2018 Int. Jt. Conf. Neural Netw. IJCNN*, IEEE, Rio de Janeiro, pp. 1–8.

Ermentrout B (1992). Complex dynamics in winner-take-all neural nets with slow inhibition. *Neural Networks* **5**, 415–431.

Eyal G, Verhoog MB, Testa-Silva G, Deitcher Y, Benavides-Piccione R, DeFelipe J, de Kock CPJ, Mansvelder HD & Segev I (2018). Human cortical pyramidal neurons: from spines to spikes via models. *Front Cell Neurosci* **12**.

Eyal G, Verhoog MB, Testa-Silva G, Deitcher Y, Lodder JC, Benavides-Piccione R, Morales J, DeFelipe J, de Kock CP, Mansvelder HD & Segev I (2016). Unique membrane properties and enhanced signal processing in human neocortical neurons. *eLife* **5**, e16553.

Fiete IR, Senn W, Wang CZH & Hahnloser RHR (2010). Spike-time-dependent plasticity and heterosynaptic competition organize networks to produce long scale-free sequences of neural activity. *Neuron* **65**, 563–576.

Fitz H, Uhlmann M, van den Broek D, Duarte R, Hagoort P & Petersson KM (2020). Neuronal spike-rate adaptation supports working memory in language processing. *PNAS* **117**, 20881–20889.

Fox NP, Leonard M, Sjerps MJ & Chang EF (2020). Transformation of a temporal speech cue to a spatial neural code in human auditory cortex. *eLife* **9**, e53051.

Gallistel C (2021). The physical basis of memory. *Cognition* **213**, 104533.

Gastaldi C, Schwalger T, De Falco E, Quiroga RQ & Gerstner W (2021). When shared concept cells support associations: Theory of overlapping memory engrams. *PLoS Comput Biol* **17**, e1009691.

Gillett M, Pereira U & Brunel N (2020). Characteristics of sequential activity in networks with temporally asymmetric Hebbian learning. *Proc Natl Acad Sci U S A* **117**, 29948–29958.

Gjorgjieva J, Clopath C, Audet J & Pfister JP (2011). A triplet spike-timing-dependent plasticity model generalizes the Bienenstock-Cooper-Munro rule to higher-order spatiotemporal correlations. *Proceedings of the National Academy of Sciences* **108**, 19383–19388.

Gütig R & Sompolinsky H (2006). The tempotron: a neuron that learns spike timing–based decisions. *Nat Neurosci* **9**, 420–428.

Gütig R & Sompolinsky H (2009). Time-warp-invariant neuronal processing. *PLoS Biol* **7**, e1000141.

Haga T & Fukai T (2018). Dendritic processing of spontaneous neuronal sequences for single-trial learning. *Sci Rep* **8**, 15166.

Hawkins J & Ahmad S (2016). Why neurons have thousands of synapses, a theory of sequence memory in neocortex. *Front Neural Circuits* **10**.

Hebb DO (1949). *The organization of behavior; a neuropsychological theory*. The organization of behavior; a neuropsychological theory, Wiley, Oxford, England.

Hemberger M, Shein-Idelson M, Pammer L & Laurent G (2019). Reliable Sequential Activation of Neural Assemblies by Single Pyramidal Cells in a Three-Layered Cortex. *Neuron* .

Holt GR, Softky WR, Koch C & Douglas RJ (1996). Comparison of discharge variability in vitro and in vivo in cat visual cortex neurons. *Journal of Neurophysiology* **75**, 1806–1814.

Jones IS & Kording KP (2021). Might a Single Neuron Solve Interesting Machine Learning Problems Through Successive Computations on Its Dendritic Tree? *Neural Computation* **33**, 1554–1571.

Kastellakis G & Poirazi P (2019). Synaptic Clustering and Memory Formation. *Front Mol Neurosci* **12**.

Kastellakis G, Silva AJ & Poirazi P (2016). Linking memories across time via neuronal and dendritic overlaps in model neurons with active dendrites. *Cell Reports* **17**, 1491–1504.

Koch C (1998). *Biophysics of computation: information processing in single neurons, by*, volume 22. Elsevier, oxford university press. edition.

Langille JJ & Brown RE (2018). The Synaptic Theory of Memory: A Historical Survey and Reconciliation of Recent Opposition. *Front Syst Neurosci* **12**, 52.

Langille JJ & Gallistel CR (2020). Locating the engram: Should we look for plastic synapses or information-storing molecules? *Neurobiol Learn Mem* **169**, 107164.

Larkum ME (2022). Are dendrites conceptually useful? *Neuroscience* **489**, 4–14.

Legenstein R & Maass W (2011). Branch-Specific Plasticity Enables Self-Organization of Nonlinear Computation in Single Neurons. *J Neurosci* **31**, 10787–10802.

Leugering J, Nieters P & Pipa G (2023). Dendritic plateau potentials can process spike sequences across multiple time-scales. *Front Cogn* **2**.

Litwin-Kumar A & Doiron B (2014). Formation and maintenance of neuronal assemblies through synaptic plasticity. *Nat Commun* **5**.

Luz Y & Shamir M (2012). Balancing Feed-Forward Excitation and Inhibition via Hebbian Inhibitory Synaptic Plasticity. *PLOS Computational Biology* **8**, e1002334.

Maes A, Barahona M & Clopath C (2020). Learning spatiotemporal signals using a recurrent spiking network that discretizes time. *PLOS Computational Biology* **16**, e1007606.

Markram H, Gerstner W & Sjöström PJ (2011). A History of Spike-Timing-Dependent Plasticity. *Front Synaptic Neurosci* **3**.

Marslen-Wilson W (1973). Linguistic structure and speech shadowing at very short latencies. *Nature* **244**, 522–523.

McClelland JL & Elman JL (1986). The TRACE model of speech perception. *Cognitive Psychology* **18**, 1–86.

Miehl C & Gjorgjieva J (2022), Stability and learning in excitatory synapses by nonlinear inhibitory plasticity.

Milojkovic BA, Radojicic MS & Antic SD (2005). A strict correlation between dendritic and somatic plateau depolarizations in the rat prefrontal cortex pyramidal neurons. *J Neurosci* **25**, 3940–3951.

Mongillo G, Barak O & Tsodyks M (2008). Synaptic Theory of Working Memory. *Science* **319**, 1543–1546.

Naumann LB, Keijser J & Sprekeler H (2022). Invariant neural subspaces maintained by feedback modulation. *eLife* **11**, e76096.

Nicola W & Clopath C (2017). Supervised learning in spiking neural networks with FORCE training. *Nat Commun* **8**, 2208.

Papoutsi A, Sidiropoulou K & Poirazi P (2014). Dendritic nonlinearities reduce network size requirements and mediate ON and OFF states of persistent activity in a PFC microcircuit model. *PLoS Comput Biol* **10**, e1003764.

Park Y & Geffen MN (2020). A circuit model of auditory cortex. *PLOS Computational Biology* **16**, e1008016.

Pedrosa V & Clopath C (2020), Voltage-based inhibitory synaptic plasticity: network regulation, diversity, and flexibility.

Poeppel D (2012). The maps problem and the mapping problem: Two challenges for a cognitive neuroscience of speech and language. *Cogn Neuropsychol* **29**, 34–55.

Poeppel D & Idsardi W (2022). We don't know how the brain stores anything, let alone words. *Trends in Cognitive Sciences* , S1364661322002066.

Poirazi P, Brannon T & Mel BW (2003). Pyramidal neuron as two-layer neural network. *Neuron* **37**, 989–999.

Poo Mm, Pignatelli M, Ryan TJ, Tonegawa S, Bonhoeffer T, Martin KC, Rudenko A, Tsai LH, Tsien RW, Fishell G, Mullins C, Gonçalves JT, Shtrahman M, Johnston ST, Gage FH, Dan Y, Long J, Buzsáki G & Stevens C (2016). What is memory? The present state of the engram. *BMC Biology* **14**, 40.

Pulvermuller F (1999). Words in the brain's language. *Behav Brain Sci* **22**, 253–+.

Quaresima A, Fitz H, Duarte R, Hagoort P & Petersson KM (2022). The Tripod neuron: a minimal structural reduction of the dendritic tree. *J Neurophysiol* , 58.

Rajan K, Harvey CD & Tank DW (2016). Recurrent Network Models of Sequence Generation and Memory. *Neuron* **90**, 128–142.

Reifenstein ET, Bin Khalid I & Kempter R (2021). Synaptic learning rules for sequence learning. *eLife* **10**, e67171.

Riquelme JL, Hemberger M, Laurent G & Gjorgjieva J (2023). Single spikes drive sequential propagation and routing of activity in a cortical network. *eLife* **12**, e79928.

Roth A & van Rossum MCW (2009), Modeling synapses. In *Computational Modeling Methods for Neuroscientists*, ed. De Schutter E, The MIT Press, pp. 139–160.

Sakurai Y, Osako Y, Tanisumi Y, Ishihara E, Hirokawa J & Manabe H (2018). Multiple Approaches to the Investigation of Cell Assembly in Memory Research—Present and Future. *Front Syst Neurosci* **12**.

Salaj D, Subramoney A, Kraisnikovic C, Bellec G, Legenstein R & Maass W (????). Spike frequency adaptation supports network computations on temporally dispersed information. *eLife* **10**, e65459.

Scott DN & Frank MJ (2023). Adaptive control of synaptic plasticity integrates micro- and macroscopic network function. *Neuropsychopharmacology* **48**, 121–144.

Shouval HZ (2011). What is the appropriate description level for synaptic plasticity? *Proc Natl Acad Sci U S A* **108**, 19103–19104.

Spalla D, Cornacchia IM & Treves A (2021). Continuous attractors for dynamic memories. *eLife* **10**, e69499.

Sreenivasan KK & D'Esposito M (2019). The what, where and how of delay activity. *Nat Rev Neurosci* , 1.

Tetzlaff C, Kolodziejski C, Timme M & Wörgötter F (2011). Synaptic Scaling in Combination with Many Generic Plasticity Mechanisms Stabilizes Circuit Connectivity. *Front Comput Neurosci* **5**, 47.

Tomasello R, Garagnani M, Wennekers T & Pulvermüller F (2018). A Neurobiologically Constrained Cortex Model of Semantic Grounding With Spiking Neurons and Brain-Like Connectivity. *Front Comput Neurosci* **12**.

Tremblay R, Lee S & Rudy B (2016). GABAergic interneurons in the neocortex: from cellular properties to circuits. *Neuron* **91**, 260–292.

Triesch J, Vo AD & Hafner AS (2018). Competition for synaptic building blocks shapes synaptic plasticity. *eLife* **7**, e37836.

Turrigiano G (2011). Too many cooks? Intrinsic and synaptic homeostatic mechanisms in cortical circuit refinement. *Annu Rev Neurosci* **34**, 89–103.

Ujfalussy BB, Makara JK, Lengyel M & Branco T (2018). Global and multiplexed dendritic computations under in vivo-like conditions. *Neuron* **100**, 579–592.e5.

Vaz AP, Wittig JH, Inati SK & Zaghloul KA (2020). Replay of cortical spiking sequences during human memory retrieval. *Science* **367**, 1131–1134.

Vogels TP, Sprekeler H, Zenke F, Clopath C & Gerstner W (2011). Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573.

Wang XJ (1999). Synaptic basis of cortical persistent activity: the importance of NMDA receptors to working memory. *J Neurosci* **19**, 9587–9603.

Wang XJ (2021). 50 years of mnemonic persistent activity: quo vadis? *Trends in Neurosciences* **44**, 888–902.

Wilmes KA & Clopath C (2023). Dendrites help mitigate the plasticity-stability dilemma. *Sci Rep* **13**, 6543.

Yang GR, Murray JD & Wang XJ (2016). A dendritic disinhibitory circuit mechanism for pathway-specific gating. *Nat Commun* **7**, 12815–12815.

Zenke F, Agnes EJ & Gerstner W (2015). Diverse synaptic plasticity mechanisms orchestrated to form and retrieve memories in spiking neural networks. *Nat Commun* **6**, 6922.

Zenke F & Gerstner W (2017). Hebbian plasticity requires compensatory processes on multiple timescales. *Phil Trans R Soc B* **372**, 20160259.

Zenke F & Vogels TP (2020). The remarkable robustness of surrogate gradient learning for instilling complex function in spiking neural networks. *bioRxiv* , 2020.06.29.176925.