

ALMA MATER STUDIORUM - UNIVERSITÀ DI BOLOGNA
CAMPUS DI CESENA

DIPARTIMENTO DI INGEGNERIA DELL'ENERGIA ELETTRICA E
DELL'INFORMAZIONE "*GUGLIELMO MARCONI*"

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA BIOMEDICA

**SVILUPPO DI UNA RETE NEURALE
MULTISTRATO PER L'ANALISI
DELL'ACCOPPIAMENTO THETA-GAMMA
NELL'IPPOCAMPO E NELLE REGIONI
PREFRONTALI ASCRITTE ALLA MEMORIA DI
LAVORO**

Tesi in
Sistemi Neurali

Relatore
Ing. Mauro Ursino

Presentata da
Nicole Cesaretti

Anno Accademico 2019-2020
Sessione di Laurea III – Appello I

INDICE

<u>Introduzione</u>	1
<u>L'accoppiamento theta-gamma nella memoria di lavoro</u>	1
<u>Approccio alla modellazione</u>	4
<u>Descrizione del modello</u>	5
<u>Modello per la singola colonna corticale</u>	5
<u>Architettura della rete neurale</u>	12
<u>Set di pattern di lavoro</u>	18
<u>Addestramento della rete</u>	20
<u>Fase 1: sinapsi laterali in L₁</u>	20
<u>Fase 2: sinapsi laterali in L₂ e L₃</u>	22
<u>Fase 3: feedback da L₃ a L₂</u>	24
<u>Risultati</u>	27
<u>Test parziali</u>	28
<u>Performance in modalità <i>recall</i></u>	33
<u>Performance in modalità <i>desynchronize</i></u>	39
<u>Test in condizioni alterate</u>	44
<u>Sviluppi futuri</u>	52
<u>Conclusioni</u>	56
<u>Bibliografia</u>	59

Introduzione

Esistono diversi approcci per registrare l'attività elettrica cerebrale; uno di questi è l'elettroencefalografia di superficie (EEG). I tracciati EEG sono prodotti dall'attività complessiva di numerose cellule cerebrali: i principali contributi sono costituiti da potenziali postsinaptici corticali, i quali possono essere modulati da segnali provenienti da regioni sottocorticali.

Quando nel tracciato sono presenti onde il cui spettro di potenza presenta frequenze (f_r) all'interno di una determinata banda, si parla di ritmi cerebrali. In particolare, si distinguono: 1) ritmo delta, $f_r < 4$ Hz; 2) ritmo theta, $f_r \in [4, 7]$ Hz; 3) ritmo alpha, $f_r \in [8, 13]$ Hz; ritmo beta, $f_r \in [14, 25]$ Hz; ritmo gamma, $f_r > 25$ Hz. I ritmi sono largamente implicati in molte funzioni cerebrali, ed oscillazioni appartenenti a bande differenti possono ovviamente sovrapporsi.

La rete neurale realizzata e discussa nel presente elaborato si incentra sulla simulazione dei ritmi theta e gamma, in particolare nell'ippocampo e nelle regioni corticali prefrontali legate alla working memory. L'obiettivo del modello è consentire l'analisi del ruolo dei due ritmi e del loro accoppiamento, in varie condizioni di lavoro.

Nella sezione seguente saranno brevemente esposte nozioni di neurofisiologia e di neuroscienze pertinenti a questi temi; tali nozioni sono le fondamenta sulle base delle quali il modello è stato progettato, e ne consentiranno pertanto una migliore comprensione.

L'accoppiamento theta-gamma nella memoria di lavoro

La working memory, o memoria di lavoro, è una tipologia di memoria a breve termine che mantiene online informazioni strettamente legate al compito che il soggetto sta svolgendo in un dato istante. Il suo contenuto può dipendere da informazioni sensoriali, da processi attentivi, da rappresentazioni evocate da altre tipologie di memoria e da piani d'azione [1], e viene aggiornato quando il contesto

lo richiede. Il substrato neurale della memoria di lavoro è dato dall'attività in regioni prefrontali.

L'ippocampo è una struttura collocata sul lato mediale del lobo temporale di ciascun emisfero cerebrale. Svolge un ruolo fondamentale nel consolidamento della memoria a lungo termine, nel *retrieval* di ricordi e nella navigazione spaziale. Oscillazioni nelle bande theta e gamma caratterizzano le regioni ippocampali. Le oscillazioni in banda theta sono le più robuste; a lungo è stato supposto che esse fossero generate da circuiti al di fuori dell'ippocampo, ma dati più recenti hanno mostrato che circuiti interni alla struttura sono in grado di generare da sé oscillazioni theta [2]. Queste oscillazioni identificano delle “finestre di attività” entro le quali compare l'attività in banda gamma (i due ritmi sono dunque accoppiati).

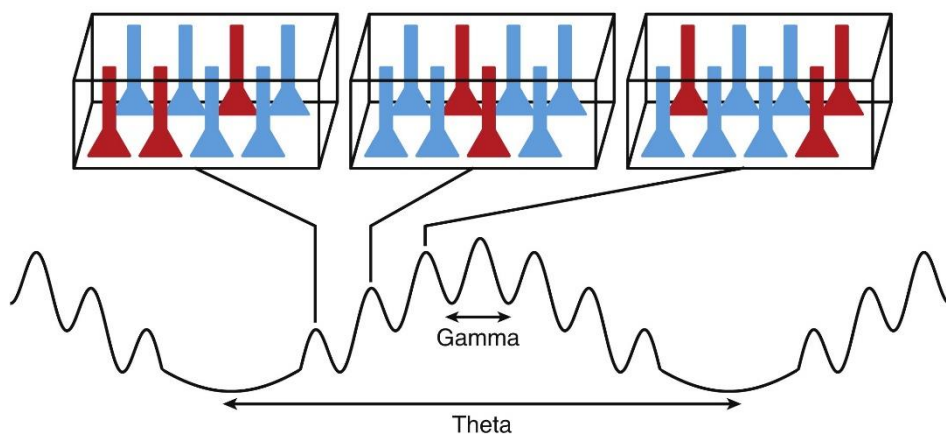


Figura 1 - Oscillazioni theta con ritmo gamma "annidato". Le frecce indicano l'estensione dei rispettivi periodi. Ogni periodo gamma corrisponde all'attività di un certo gruppo di neuroni (neuroni rossi: attivi, neuroni blu: non attivi), i quali codificano un certo oggetto (un "oggetto" può essere l'elemento di una scena visiva, un concetto astratto o una posizione spaziale).

Le oscillazioni in banda gamma “annidate” nei cicli theta sono legate alle operazioni di *binding* (unione delle *features* che caratterizzano un oggetto) e *segmentation* (separazione delle *features* di un oggetto dalle *features* di altri oggetti): il segnale in ciascun periodo gamma sarebbe da associare all'attività di cellule che codificano informazioni relative ad uno stesso oggetto; una sequenza di periodi gamma corrisponderebbe dunque ad una precessione di oggetti [3-5] (fig. 1; si noti che con “oggetti” ci si riferisce a pattern di attività – un “oggetto”

può dunque far riferimento all'elemento di una scena visiva, un concetto astratto o una posizione spaziale). A supporto di questa interpretazione, esistono ad esempio dati sperimentali che esaminano il caso di un ratto posto ad esplorare un labirinto. I dati si riferiscono, in particolare, alle cosiddette *place cells* ippocampali. In base alla loro attività, nel singolo ciclo gamma emerge il pattern che codifica una posizione nel labirinto, mentre una successione di cicli gamma codifica un movimento da una posizione iniziale a una finale [6].

Diversi set di dati sperimentali hanno evidenziato come l'attività prefrontale possa presentare le stesse caratteristiche dell'attività ippocampale (accoppiamento dei ritmi theta e gamma), soprattutto quando il soggetto è impegnato in compiti che coinvolgono la working memory [7-9]. In particolare, oscillazioni più lente (nelle bande theta e alpha) sarebbero generate negli strati corticali che si trovano più in profondità, mentre oscillazioni più rapide (in banda gamma) sarebbero generate negli strati corticali più in superficie. Le oscillazioni prodotte negli strati in profondità modulerebbero quelli prodotti più superficialmente [10]. Questi ritrovamenti sperimentali suggeriscono dunque che nella memoria di lavoro i vari "oggetti" siano rappresentati in maniera sequenziale, e che sussista la stessa precessione in ritmo gamma, contenuta in una finestra di lavoro delineata dal ritmo theta, descritta in precedenza per l'ippocampo e mostrata in fig. 1. È stato inoltre supposto che in fase di *recall* ("lettura" della working memory, cioè quando è necessario accedere ai suoi contenuti per completare il task in corso) la sequenza di "oggetti" in ritmo gamma venga opportunamente scorsa [11], esattamente come avviene nell'ippocampo per il *recall* di ricordi e di posizioni spaziali.

L'ippocampo ha connessioni bidirezionali con le aree frontali – può cioè sia ricevere da esse informazioni, sia inviarne a sua volta, e ciò consente all'attività delle regioni di sincronizzarsi. A tal proposito, esistono anche dati che supportano l'ipotesi di un accoppiamento diretto tra l'attività ippocampale e quella prefrontale in task incentrati sulla working memory [12-14]. Da notare che la sincronizzazione dell'attività in regioni diverse, e in particolare la sincronizzazione in banda theta, è stata in alcuni studi interpretata come una tecnica di "reclutamento", da parte della memoria di lavoro, delle regioni cerebrali di cui essa ha necessità in un dato momento, in dipendenza dal compito in corso

[15]. In tale ottica, si può supporre che la working memory sia in grado di reclutare, tramite *entrainment* in banda theta, le zone ippocampali nel momento in cui il task che il soggetto deve compiere richiede, ad esempio, l'evocazione di ricordi.

Approccio alla modellazione

Per descrivere le unità di base della rete progettata si è scelto di impiegare modelli di massa neurale (*neural mass models*, NMMs). I NMMs sono caratterizzati dal fatto che la loro unità di calcolo non rappresenta un solo neurone (come accade ad esempio nelle reti neurali spiking), ma piuttosto rappresenta una popolazione neuronale, composta da cellule aventi caratteristiche comuni. Nel nostro caso, le popolazioni modellate sono quelle dei neuroni individuabili in una colonna corticale, la quale può essere considerata un'unità di calcolo della corteccia cerebrale (o dell'archicorteccia, se si fa riferimento all'ippocampo). Maggiori dettagli in merito al NMM impiegato nel presente lavoro sono forniti nel capitolo seguente.

I NMMs sono organizzati in modo da formare i vari strati (o *layers*) della rete. Ogni strato modella una certa regione cerebrale, ma non in modo univoco – a seconda delle condizioni di funzionamento, la regione rappresentata dal *layer* può essere diversa (può trattarsi, alternativamente, di regioni prefrontali o ippocampali). L'interpretazione del ruolo degli strati in ciascuna delle simulazioni è uno dei punti più delicati nell'analisi del modello.

Descrizione del modello

Il modello realizzato, ispirato ad architetture proposte in letteratura [16-18], è una rete neurale costituita da quattro *layers* organizzati gerarchicamente; nell'ordine: WM (che modella la Working Memory), L_1 (che modella una regione ippocampale volta alla ricostruzione di ricordi), L_2 e L_3 (che modellano regioni ippocampali o prefrontali, a seconda delle condizioni di funzionamento). Ciascuno strato è composto da 400 unità elementari, ognuna rappresentante una colonna corticale, disposte in una griglia $M \times N$ ($M, N = 20$). L'input esterno è fornito a WM; l'informazione è poi propagata in avanti e, grazie a connessioni sinaptiche, elaborata, fino a raggiungere L_3 , la cui attività è considerata essere l'output.

Tutti i codici sono stati sviluppati in ambiente Matlab. Si è ovviamente effettuata una discretizzazione nel dominio temporale per la risoluzione delle equazioni, e le integrazioni numeriche sono state eseguite con il metodo di Eulero.

Nei paragrafi che seguono, saranno descritte la struttura delle unità elementari, le caratteristiche degli strati, la loro architettura sinaptica e la natura degli input impiegati. Saranno inoltre illustrati i comportamenti desiderati per la rete nelle sue due modalità di funzionamento principali: *recall* e *desynchronize*.

Modello per la singola colonna corticale

Ogni colonna corticale è rappresentata da un modello di massa neurale a quattro popolazioni, che consistono in 1) neuroni piramidali (pedice p), 2) interneuroni eccitatori (pedice e), 3) interneuroni inibitori *slow* (chiamati anche $GABA_{slow}$, pedice s) e 4) interneuroni inibitori *fast* (chiamati anche $GABA_{fast}$, pedice f). Tutte le popolazioni sono descritte da un simile formalismo matematico. Siano $i, j = p, e, s$ oppure f : il potenziale di membrana medio della popolazione i (v_i), influenzato dalla comunicazione sinaptica con le altre popolazioni, è elaborato tramite la funzione d'attivazione $S(v_i)$, che fornisce l'attività media, o

frequenza media di sparo, z_i . $S(v_i)$ ha andamento sigmoidale e consente perciò di simulare sia la presenza di una soglia (se v_i è al di sotto di tale soglia, la popolazione è silente, $z_i \approx 0$ Hz), sia il fenomeno di saturazione (anche per v_i molto elevati, z_i non può crescere oltre un valore massimo e_0). L'attività media z_i è impiegata come ingresso al blocco dinamico che modella le sinapsi e che genera come uscita y_i , un contributo “normalizzato” al potenziale di membrana postsinaptico; “normalizzato” in quanto esso dovrà essere moltiplicato per il corretto peso sinaptico C_{ji} per fornire l'effettivo contributo al potenziale v_j .

Le sinapsi tra popolazioni sono descritte con equazioni differenziali del secondo ordine aventi la seguente formulazione:

$$\frac{d^2 y_i(t)}{dt^2} = \frac{G_n}{\tau_n} z_i(t) - \frac{2}{\tau_n} \frac{dy_i(t)}{dt} - \frac{y_i(t)}{\tau_n^2} \quad (X1)$$

e sono caratterizzate dunque dai parametri G_n (guadagno) e τ_n (costante di tempo). Tale formulazione corrisponde alla risposta impulsiva $h_n(t)$:

$$h_n(t) = \frac{G_n}{\tau_n} t \cdot e^{-\frac{t}{\tau_n}} \quad (X2)$$

nonché alla coppia di equazioni di primo ordine:

$$\begin{cases} \frac{dy_i(t)}{dt} = x_i(t) \\ \frac{dx_i(t)}{dt} = \frac{G_n}{\tau_n} z_i(t) - \frac{2}{\tau_n} x_i(t) - \frac{y_i(t)}{\tau_n^2} \end{cases} \quad (X3)$$

Nel modello sono incluse sinapsi eccitatorie (pedice $n = e$, impiegate da neuroni piramidali e interneuroni eccitatori), inibitorie con dinamica lenta (pedice $n = s$, impiegate dai $GABA_{\text{slow}}$) e inibitorie con dinamica veloce (pedice $n = f$, impiegate dai $GABA_{\text{fast}}$). Per le unità elementari che compongono gli strati L_1 , L_2 e L_3 vale quanto segue: i neuroni piramidali inviano sinapsi a ciascuna delle altre tre popolazioni; i $GABA_{\text{slow}}$ inviano sinapsi sia ai piramidali, sia ai $GABA_{\text{fast}}$; gli interneuroni eccitatori inviano sinapsi unicamente ai piramidali; e per concludere, i $GABA_{\text{fast}}$ inviano sinapsi sia ai piramidali, sia a sé stessi. In altri lavori, è stato dimostrato come questo loop sia fondamentale per consentire, in certe condizioni, alla colonna corticale di generare oscillazioni in banda gamma [19], le quali

svolgono un ruolo importante in molti processi cognitivi, ad esempio nelle operazioni di *binding* e *segmentation* [20].

Unicamente per le unità elementari dello strato WM, oltre ad essere valido quanto appena detto per gli altri strati, è stato inserito un auto-anello anche per la popolazione di neuroni piramidali. Tale loop consente alle unità dello strato di mantenere l'informazione relativa all'input esterno quando esso, dopo essere stato fornito per un certo intervallo di tempo, viene rimosso. Si tratta di un tipo di connessione già proposto in precedenza per modellare regioni cerebrali [21]. Nel presente modello tale anello è posto a zero nel caso in cui l'ingresso all'unità sia non nullo, supponendo quindi una sua inattivazione dall'esterno. Gli auto-anelli vanno inoltre incontro a un reset (cioè, smettono di agire) nel caso in cui venga fornito un nuovo un input. Nel momento cui il nuovo input sarà anch'esso rimosso, indicando la necessità di una nuova memorizzazione a breve termine, si attiveranno gli auto-anelli solo delle unità appropriate, senza memoria dei loop che erano in azione prima del reset. L'esistenza di fenomeni di reset nella memoria di lavoro è documentata in letteratura [22-23].

Complessivamente, l'architettura del modello consente dunque di implementare le capacità della working memory di 1) modificare il proprio contenuto (stato interno) durante la ricezione d'informazioni rilevanti (auto-eccitazione soppressa) e 2) quando opportuno, arrestare l'aggiornamento dello stato per mantenere online le più recenti informazioni ricevute (auto-eccitazione selettiva attiva).

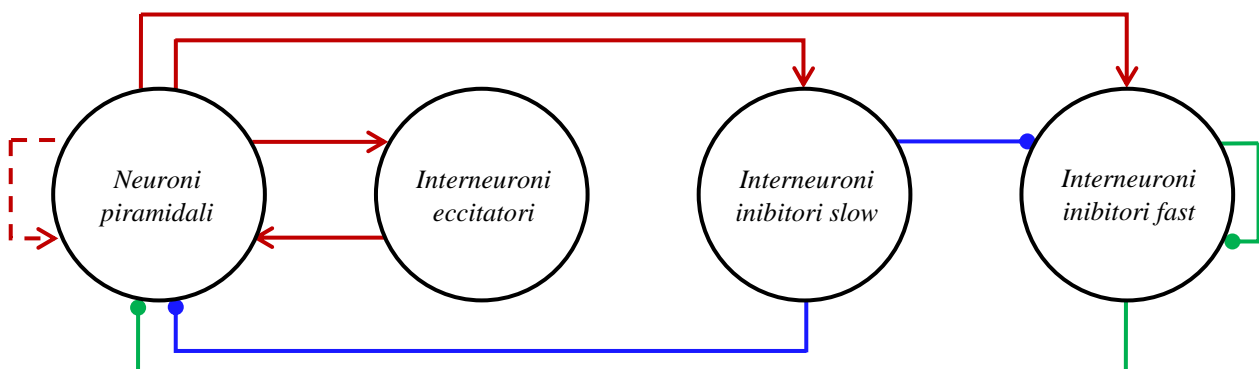


Figura 2 - Architettura del modello di massa neuronale (NMM) per la singola colonna corticale. In rosso, sinapsi eccitatorie; in blu, sinapsi inibitorie lente; in verde, sinapsi inibitorie veloci. La sinapsi eccitatoria tratteggiata rappresenta il loop presente solo per le unità di WM.

Il modello è predisposto a ricevere due tipi di input dall'esterno: u_p e u_f , rispettivamente indirizzati ai neuroni piramidali e ai GABA_{fast}. Tali input sono ottenuti aggiungendo all'eccitazione dall'esterno del rumore bianco. Sono stati dunque semplicemente modellati come rumore bianco con una certa varianza (σ_p^2 o σ_f^2 a seconda della popolazione) e un valore medio (m_p o m_f , di nuovo, a seconda della popolazione) pari all'eccitazione dall'esterno.

L'attività dei neuroni piramidali e dei GABA_{fast} può inoltre essere influenzata da contributi dovuti a comunicazioni sinaptiche a lungo raggio – cioè, da sinapsi che provengono da altre colonne corticali. Tali contributi sono denominati $E(t)$ e $I(t)$, rispettivamente, e verranno discussi nel dettaglio nella sezione successiva.

Si riportano di seguito le equazioni che riassumono quanto finora descritto relativamente alle quattro popolazioni di ogni unità, nonché una figura che schematizza nel dettaglio il modello.

Neuroni piramidali (L_1, L_2, L_3):

$$\left\{ \begin{array}{l} v_p(t) = C_{pe}y_e(t) - C_{ps}y_s(t) - C_{pf}y_f(t) + E(t) \\ S(v_p(t)) = z_p(t) = \frac{2e_0}{1 + e^{r(s_0 - v_p)}} \\ \frac{dy_p(t)}{dt} = x_p(t) \\ \frac{dx_p(t)}{dt} = \frac{G_e}{\tau_e}z_p(t) - \frac{2}{\tau_e}x_p(t) - \frac{y_p(t)}{\tau_e^2} \end{array} \right. \quad (X6)$$

Neuroni piramidali (WM):

$$\left\{ \begin{array}{l} v_p(t) = C_{pe}y_e(t) + \widehat{C_{pp}}(t)y_p(t) - C_{ps}y_s(t) - C_{pf}y_f(t) + E(t) \\ \widehat{C_{pp}}(t) = \begin{cases} C_{pp}, & m_p(t) = 0 \\ 0, & m_p(t) \neq 0 \end{cases} \\ S(v_p(t)) = z_p(t) = \frac{2e_0}{1 + e^{r(s_0 - v_p)}} \\ \frac{dy_p(t)}{dt} = x_p(t) \\ \frac{dx_p(t)}{dt} = \frac{G_e}{\tau_e}z_p(t) - \frac{2}{\tau_e}x_p(t) - \frac{y_p(t)}{\tau_e^2} \end{array} \right. \quad (X6')$$

Interneuroni eccitatori:

$$\left\{ \begin{array}{l} v_e(t) = C_{ep}y_p(t) \\ S(v_e(t)) = z_e(t) = \frac{2e_0}{1 + e^{r(s_0 - v_e)}} \\ \frac{dy_e(t)}{dt} = x_e(t) \\ \frac{dx_e(t)}{dt} = \frac{G_e}{\tau_e} \left(z_e(t) + \frac{u_p}{C_{pe}} \right) - \frac{2}{\tau_e} x_e(t) - \frac{y_e(t)}{\tau_e^2} \end{array} \right. \quad (X7)$$

Interneuroni inibitori *slow*:

$$\left\{ \begin{array}{l} v_s(t) = C_{sp}y_p(t) \\ S(v_s(t)) = z_s(t) = \frac{2e_0}{1 + e^{r(s_0 - v_s)}} \\ \frac{dy_s(t)}{dt} = x_s(t) \\ \frac{dx_s(t)}{dt} = \frac{G_s}{\tau_s} z_s(t) - \frac{2}{\tau_s} x_s(t) - \frac{y_s(t)}{\tau_s^2} \end{array} \right. \quad (X8)$$

Interneuroni inibitori *fast*:

$$\left\{ \begin{array}{l} v_f(t) = C_{fp}y_p(t) - C_{fs}y_s(t) - C_{ff}y_f(t) + y_l(t) + I(t) \\ S(v_f(t)) = z_f(t) = \frac{2e_0}{1 + e^{r(s_0 - v_f(t))}} \\ \frac{dy_f(t)}{dt} = x_f(t) \\ \frac{dx_f(t)}{dt} = \frac{G_f}{\tau_f} z_f(t) - \frac{2}{\tau_f} x_f(t) - \frac{y_f(t)}{\tau_f^2} \\ \frac{dy_l(t)}{dt} = x_l(t) \\ \frac{dx_l(t)}{dt} = \frac{G_e}{\tau_e} u_f(t) - \frac{2}{\tau_e} x_l(t) - \frac{y_l(t)}{\tau_e^2} \end{array} \right. \quad (X9)$$

Si noti che entrambi gli ingressi (u_p e u_f , che compaiono nei blocchi di equazioni X7 e X9) sono elaborati da una dinamica sinaptica di tipo eccitatorio per generare un contributo a v_p o v_f . Questo poiché assumiamo, in accordo con le attuali conoscenze neurofisiologiche, che ogni ingresso esterno alla colonna corticale provenga da assoni che fanno uso del neurotrasmettitore glutammato. In particolare, per l'elaborazione di u_f è introdotto nel modello un blocco dinamico

	Descrizione	Pedici
G_x	Guadagno della sinapsi di tipo x	e (eccitatoria), s (inibitoria lenta), f (inibitoria veloce)
τ_x	Costante di tempo della sinapsi di tipo x	
v_i	Potenziale di membrana medio della popolazione i	p: neuroni piramidali, e: interneuroni eccitatori, s: interneuroni inibitori lenti, f: interneuroni inibitori veloci. Solo per x e y : 1, relativo all'input dall'esterno ai GABA _{fast} .
z_i	Frequenza di sparo media della popolazione i	
y_i	Contributo "normalizzato" al potenziale postsinaptico generato dalla popolazione i (variabile di stato)	
x_i	Introdotta per trasformare l'equazione di ordine due in due equazioni di ordine uno (variabile di stato)	
u_i	Input dall'esterno (solo piramidali e GABA _{fast})	
C_{ij}	Peso per la connessione dalla popolazione presinaptica j alla popolazione postsinaptica i	

Tabella 1 - Descrittori del modello per la singola colonna corticale.

e_0	Frequenza di sparo massima
r	Regola la pendenza della funzione d'attivazione sigmoideale
s_0	Valore di v_i per cui $S(v_i) = e_0/2$.

Tabella 2 – Descrizione dei parametri della funzione d'attivazione sigmoideale.

Dinamica sinaptica				Funzione $S(v_i)$	
G_e (mV)	5.17	τ_e (ms)	7.7	e_0 (Hz)	5
G_s (mV)	4.45	τ_s (ms)	34	r (mV ⁻¹)	0.7
G_f (mV)	57.1	τ_f (ms)	6.8	s_0 (mV)	10
Ingressi		Connessioni intra-colonna		C_{pp}	300
m_p (Hz)	Vedasi testo	C_{ep}	31.7	C_{fp}	66.9
m_f (Hz)	Vedasi testo	C_{pe}	17.3	C_{fs}	100
σ_p^2 (s ⁻²)	5	C_{sp}	51.9	C_{pf}	16
σ_f^2 (s ⁻²)	5	C_{ps}	100	C_{ff}	18

Tabella 3 - Valori dei parametri impiegati.

Degno di nota è il fatto che, impiegando per i parametri i valori appena mostrati, se si esamina la z_p del modello quando $m_p \neq 0$ e $m_f = 0$ (condizioni standard per un'unità stimolata), la singola colonna corticale, non connessa ad altre, oscilla in banda alpha.

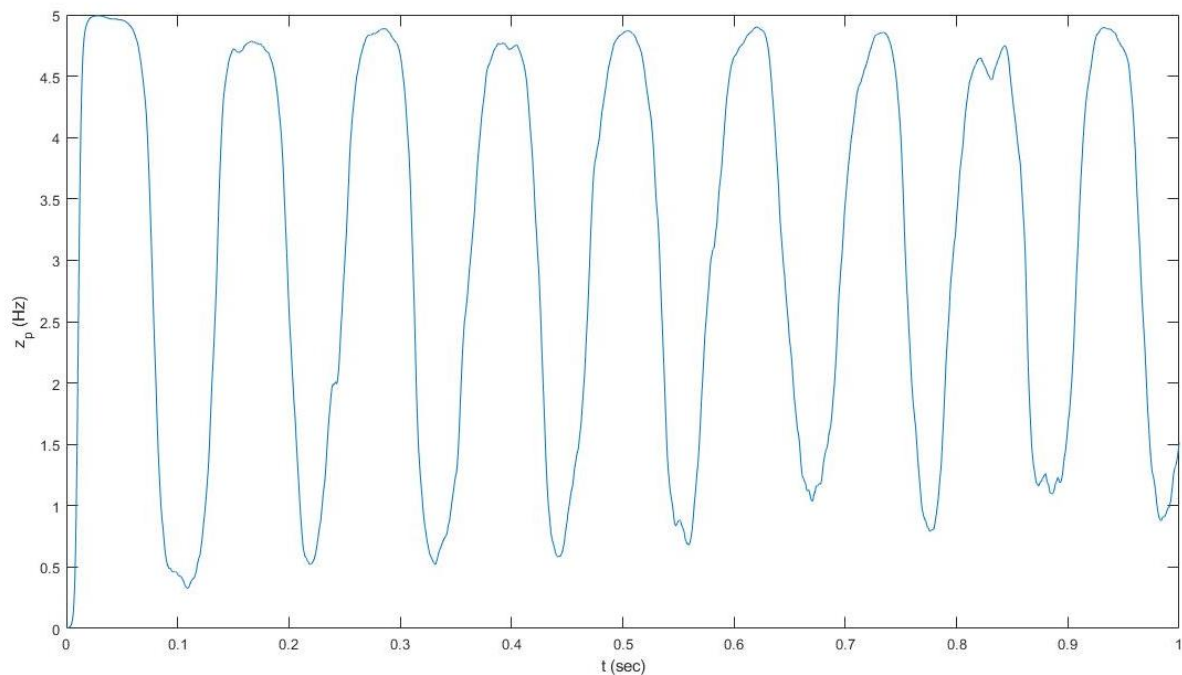


Figura 4 - Attività (frequenza media di sparo piramidale) prodotta dal modello di singola colonna, con ingresso dall'esterno verso la sola popolazione piramidale. Si evidenzia ritmo alpha (oscillazioni a circa 9 Hz).

Architettura della rete neurale

Come spiegato in precedenza, la rete è composta da quattro strati organizzati come griglie $M \times N$ ($M, N = 20$). Per semplicità di trattazione, tuttavia, tali griglie sono state vettorizzate – di conseguenza, ogni unità elementare di ciascuno strato può essere identificata da un solo pedice i ($i = 1, \dots, 400$).

A definire l'architettura della rete sono le già menzionate connessioni a lungo raggio – in avanti, laterali e in feedback. Per nessuna di queste sinapsi è stato introdotto un ritardo puro; le uniche dinamiche a entrare in gioco sono quelle dovute alle costanti di tempo delle connessioni, esattamente come accade anche per le sinapsi a corto raggio (intra-colonna). Le sinapsi a lungo raggio originano

da una popolazione di neuroni piramidali (sono dunque sempre eccitatorie ed impiegano il glutammato come neurotrasmettitore) e hanno come bersagli, alternativamente, un'altra popolazione di neuroni piramidali oppure una popolazione di interneuroni $GABA_{fast}$. È stato difatti riportato che sinapsi aventi tali bersagli hanno un ruolo cruciale nel far emergere i ritmi cerebrali d'interesse (theta e gamma), mentre connessioni verso popolazioni di interneuroni eccitatori o $GABA_{slow}$ avrebbero impatto marginale [18]. In particolare, nel modello sono presenti: 1) connessioni di tipo W, da piramidali a piramidali, mediate da meccanismi sinaptici lenti; 2) connessioni di tipo K, da piramidali a $GABA_{fast}$, mediate da meccanismi sinaptici lenti; 3) connessioni di tipo A, da piramidali a $GABA_{fast}$, mediate da meccanismi sinaptici rapidi. Le diverse velocità dei meccanismi sinaptici fanno riferimento alla presenza di differenti recettori per il neurotrasmettitore sulla membrana postsinaptica; di tipo NMDA per le sinapsi lente, AMPA per quelle rapide.

Per tutte le tipologie di sinapsi sarà di seguito impiegato lo stesso formalismo: si indicherà con S_{XY} una connessione dallo strato Y allo strato X, con $S = W, K$ o A . Dal punto di vista matematico, S_{XY} è sempre (per quanto detto relativamente alla vettorizzazione) una matrice $(M \cdot N) \times (M \cdot N)$, il cui elemento in posizione (i, j) è dunque il peso della connessione tra l'unità j -esima dello strato Y (presinaptica) e l'unità i -esima dello strato X (postsinaptica). Alcune di queste sinapsi sono fissate a priori, mentre altre sono ottenute tramite procedure di addestramento che saranno illustrate nel capitolo seguente.

Fatte queste premesse, è possibile descrivere i ruoli e le funzioni di ciascuno degli strati, nonché la connettività che li caratterizza.

WM modella la memoria di lavoro. Riceve l'ingresso dall'esterno e, se esso viene rimosso, mantiene le più recenti informazioni ricevute grazie all'architettura già descritta delle sue unità elementari. Nel caso in cui l'ingresso fosse corrotto, cioè fosse presentato con una certa percentuale di informazione mancante (ad es., un'immagine incompleta), WM non è in grado, da solo, di ricostruire il contenuto assente. Invia semplicemente l'informazione disponibile a L_1 , tramite le sinapsi fisse W_{L_1WM} , tali per cui l'unità i -esima di WM comunica con l'unità i -esima di L_1 (cioè, W_{L_1WM} è una matrice diagonale; inoltre, tutti i suoi elementi non nulli sono uguali tra loro).

A L_1 è affidato il compito di ricostruzione, in caso di input corrotti. È dunque strutturato come una rete auto-associativa: le connessioni laterali $W_{L_1L_1}$, definite a valle di una procedura di addestramento, sono tali da garantire che le *features* di un certo input si richiamino l'un l'altra. Eseguito il training, le colonne corticali di questo strato, se sufficientemente eccitate, oscillano con ritmo theta, in un'alternanza di tratti ON e tratti OFF (fig. 5). Questo ritmo emerge per via delle connessioni (intra-colonna) reciproche tra le popolazioni di neuroni piramidali e di interneuroni $GABA_{slow}$. I tratti ON definiscono le istanze in cui agli strati successivi è concesso lavorare. Le informazioni (eventualmente ricostruite) in L_1 sono sia riportate in feedback a WM tramite le sinapsi W_{WML_1} , sia propagate in avanti verso L_2 attraverso $W_{L_2L_1}$ – entrambe matrici diagonali fornite a priori (i cui elementi non nulli sono tutti uguali tra loro).

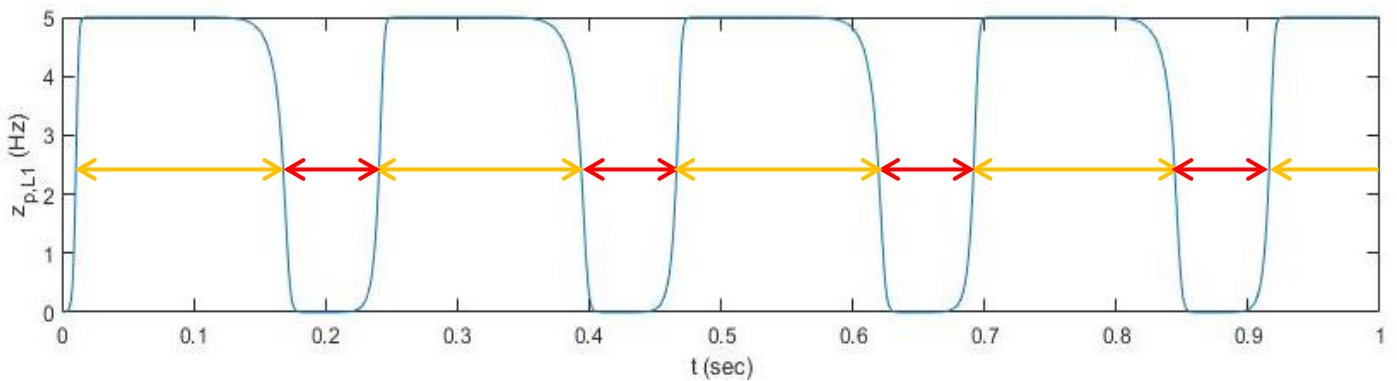


Figura 5 - Attività media delle colonne corticali in L_1 eccitate da un oggetto in input. Si evidenzia ritmo theta (oscillazioni a circa 4.5 Hz), interpretabile come un'alternanza di tratti ON (in giallo) e tratti OFF (in rosso).

Infine, gli strati L_2 e L_3 , il cui ritmo prevalente è in range gamma, eseguono le funzioni principali della rete. Tra di essi sussiste sia un set di connessioni in avanti fisse ($W_{L_3L_2}$, matrice diagonale i cui elementi non nulli sono tutti uguali), sia un set di connessioni in feedback ottenute con un addestramento ($W_{L_2L_3}$). Il feedback è tale per cui L_2 e L_3 formano una rete etero-associativa: ogni input che raggiunge L_2 , ed è dunque propagato a L_3 , richiama in L_2 (tramite $W_{L_2L_3}$) l'input successivo in una sequenza (appresa durante il training); ciò si ripete fin quando il ritmo theta in L_1 lo consente.

Le connessioni laterali nei due strati ($K_{L_2L_2}$, $A_{L_2L_2}$ e $K_{L_3L_3}$, $A_{L_3L_3}$), anch'esse ottenute tramite procedure di addestramento, implementano fenomeni di sincronizzazione (tipo K) e desincronizzazione (tipo A) in range gamma. Garantiscono, cioè, che le *features* di un certo pattern – dunque, le colonne corticali eccitate da un certo oggetto – oscillino in fase tra loro (*binding*) e siano invece sfasate rispetto alle *features* di pattern diversi (*segmentation*). Come anticipato nella sezione precedente, l'implementazione di questi fenomeni coinvolge i neuroni GABA_{fast} (attraverso $I_i(t)$).

La descrizione qualitativa appena presentata è riassunta nei blocchi di equazioni che seguono. Essi completano la descrizione formale del modello esprimendo, per tutti i *layers*, le formulazioni dei termini $E(t)$ e $I(t)$ presenti nei gruppi di equazioni X6, X6' e X9. Si ricorda che tali termini contribuiscono, rispettivamente, al potenziale di membrana medio delle popolazioni di neuroni piramidali (v_p) e di interneuroni GABA_{fast} (v_f) della *i*-esima unità dello strato.

WM:

$$\begin{aligned} E_i^{WM}(t) &= \sum_j W_{WML_1}(i,j) \cdot y_{p,j}^{L_1}(t) \\ I_i^{WM}(t) &= 0 \end{aligned} \quad (X10)$$

L₁:

$$\begin{aligned} E_i^{L_1}(t) &= \sum_j W_{L_1WM}(i,j) \cdot y_{p,j}^{WM}(t) + \sum_j W_{L_1L_1}(i,j) \cdot y_{p,j}^{L_1}(t) \\ I_i^{L_1}(t) &= 0 \end{aligned} \quad (X11)$$

L₂:

$$\begin{aligned} E_i^{L_2}(t) &= \sum_j W_{L_2L_1}(i,j) \cdot y_{p,j}^{L_1}(t) + \sum_j W_{L_2L_3}(i,j) \cdot y_{p,j}^{L_3}(t) \\ I_i^{L_2}(t) &= \sum_j K_{L_2L_2}(i,j) \cdot y_{p,j}^{L_2}(t) + \sum_j A_{L_2L_2}(i,j) \cdot z_{p,j}^{L_2}(t) + \text{Inibitore} \\ \text{Inibitore} &= R \cdot \frac{1}{2} \left(\left| T - \sum_j z_{p,j}^{L_1} \right| + \left(T - \sum_j z_{p,j}^{L_1} \right) \right) \end{aligned} \quad (X12)$$

L₃:

$$\begin{aligned} E_i^{L_3}(t) &= \sum_j W_{L_3L_2}(i,j) \cdot y_{p,j}^{L_2}(t) \\ I_i^{L_2}(t) &= \sum_j K_{L_3L_3}(i,j) \cdot y_{p,j}^{L_3}(t) + \sum_j A_{L_3L_3}(i,j) \cdot z_{p,j}^{L_3}(t) \end{aligned} \quad (X13)$$

Dove $i, j = 1, \dots, 400$.

L'effetto di modulazione sugli strati a valle da parte di L₁ è implementato tramite il termine *Inibitore* che compare nel blocco di equazioni X12. Non si tratta propriamente di un termine sinaptico; lo si è incluso tuttavia nella definizione di I_i(t) in L₂ poiché ha comunque una forte dipendenza dall'attività di colonne corticali diverse dalla *i*-esima in esame. Il valore di tale fattore è difatti individuato moltiplicando per uno scalare positivo R il risultato del confronto tra l'attività complessiva in L₁ e una soglia T. Per come è definito, *Inibitore* è non nullo solo se l'attività in L₁ è sottosoglia (accade nei sopracitati tratti OFF del ritmo theta): in tali circostanze le popolazioni GABA_{fast} di L₂ ricevono una forte eccitazione, e poiché esse sono popolazioni di neuroni inibitori, il risultato complessivo è un silenziamento dello strato L₂ (a cui segue inevitabilmente il silenziamento di L₃). Di conseguenza, lo strato L₂ è in grado di elaborare informazione solo in presenza di una sufficiente attività proveniente dal L₁ (complessivamente sopra la soglia T); lo strato L₂ è pertanto “spento” durante le fasi OFF dell'attività theta.

Si noti inoltre come, nei gruppi di equazioni X12 e X13, sinapsi di tipo A considerino, come attività presinaptica nel rispettivo strato, z_p piuttosto che y_p . Questa scelta, fatta anche in modelli precedenti [16-17], permette di simulare meccanismi sinaptici veloci; sono cioè implementate in questo modo le sinapsi che sfruttano i già menzionati recettori per il glutammato di tipo AMPA.

Si sottolinea che la rete può lavorare in due modalità distinte. Quella fin qui descritta sarà chiamata modalità *recall*, poiché è relativa al richiamo di una certa sequenza di oggetti. Essa è caratterizzata dall'oscillazione in ritmo theta di L₁, dalla presenza del feedback $W_{L_2L_3}$ e dal fatto che l'input dall'esterno consiste di un solo oggetto alla volta. Il funzionamento alternativo sarà invece chiamato *desynchronize* – in tale modalità, alla rete sono forniti in ingresso diversi oggetti

contemporaneamente, e le si richiede di risolvere un problema di *binding* e *segmentation*. In quest'ultima condizione, nei gruppi di equazioni X12 e X13, il feedback $W_{L_2L_3}$ è eliminato e sono adeguatamente rinforzate le sinapsi di tipo A. Inoltre, è bene che L_1 non presenti alcuna oscillazione; a tal scopo, rispetto alla modalità *recall* sono rafforzate le connessioni reciproche tra WM e L_1 ; differiscono cioè i pesi sinaptici contenuti in W_{L_1WM} e W_{WML_1} (ma non varia la forma di tali matrici). Si noti che il fatto che l'attività di L_1 non presenti oscillazione, e sia invece sempre a livelli alti una volta che alla rete è stato fornito l'input, implica che il termine *Inibitore* nel blocco di equazioni X12 sia sempre nullo. In questa modalità L_1 non esercita, dunque, alcuna modulazione sugli strati a valle, ma si limita a fornire ad essi un input.

Dal punto di vista neurofisiologico, ipotizziamo che a differenziare ulteriormente le due modalità di lavoro sia la rete di provenienza (cioè, la regione cerebrale di provenienza) dell'input. Difatti, in modalità *recall*, l'input dovrà essere un oggetto o un concetto che ne richiama altri da recuperare in memoria (compito mnemonico strutturato), mentre in modalità *desynchronize* l'input dovrà essere composto da varie informazioni sensoriali (compito di riconoscimento).

Si riportano di seguito uno schema che illustra l'architettura sinaptica della rete e una tabella che riassume i valori impiegati per costanti moltiplicative, soglie e sinapsi fisse finora citate, per entrambe le modalità di funzionamento.

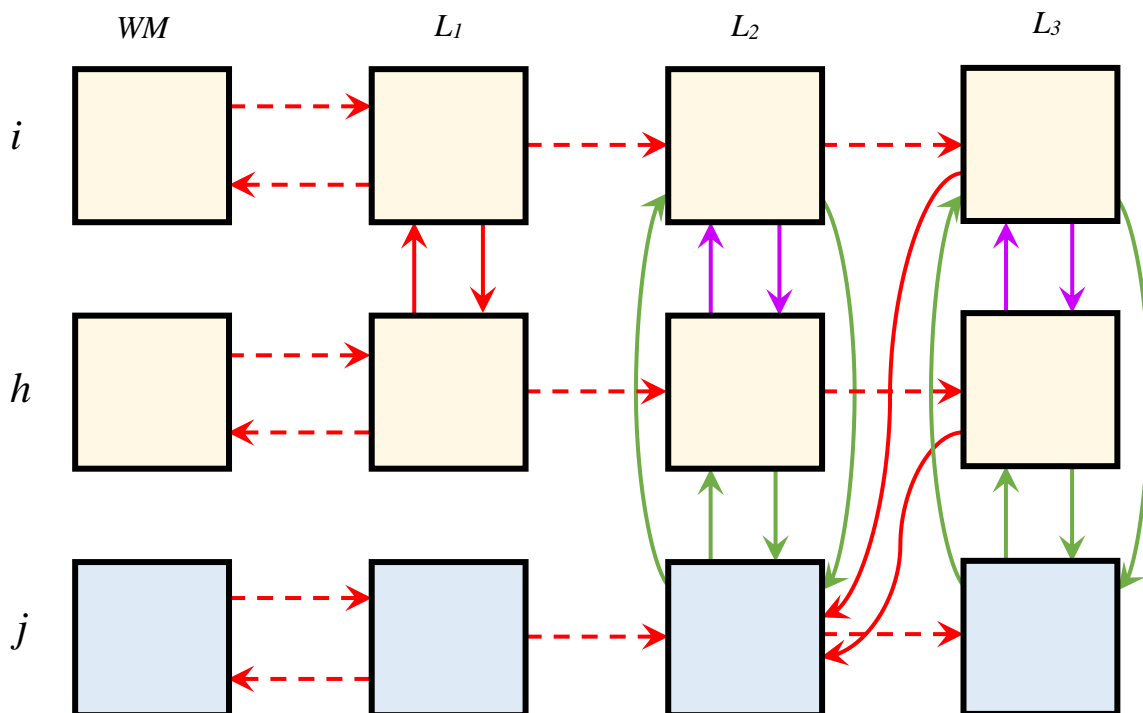


Figura 6 – Connettività della rete. Sono rappresentate le unità (riquadri) di indici i, h, j nei quattro strati, dove $i, h, j \in [1, 400]$ e non sono necessariamente posizioni consecutive l'una all'altra. Ogni riquadro contiene un NMM come quello mostrato in fig. 2. Le unità i ed h sono stimolate da uno stesso oggetto (OGG1, sfondo giallo), mentre l'unità j è stimolata da un oggetto differente (OGG2, sfondo blu), assumendo che OGG2 segua OGG1 nella sequenza appresa durante il training. Sono rappresentate con delle frecce le connessioni a lungo raggio con peso sinaptico non nullo; in particolare, omettendo i pedici per semplicità, si individuano: sinapsi W in rosso, sinapsi K in viola e sinapsi A in verde. Inoltre, sono indicate con linea continua le sinapsi ottenute da procedure di addestramento, e con linea tratteggiata quelle fisse fornite a priori. Con l'architettura mostrata, la rete opera in modalità recall. In modalità desynchronize, sarebbero assenti le sinapsi W (rosse) di feedback dagli elementi i ed h in L_3 verso l'elemento j in L_2 .

W_{L_1WM} (recall)	100	$W_{L_2L_1}$	120
W_{L_1WM} (desynchronize)	300	$W_{L_3L_2}$	186
W_{WML_1} (recall)	100	T	20
W_{WML_1} (desynchronize)	300	R	1000

Tabella 4 - Valori delle sinapsi a lungo raggio non addestrabili e di parametri inerenti.

Set di pattern di lavoro

I pattern impiegati per l'addestramento e il test delle performance della rete sono immagini binarie $20px \times 20px$ l'una. Ogni pixel "acceso" è considerato essere una *feature* del pattern. Come per gli strati che compongono la rete, anche per tali input si è eseguita per comodità matematica una vettorizzazione, che consente di indirizzare ciascun pixel con un solo indice (ad es., $i = 1, \dots, 400$). Sulla base del valore dell' i -esimo pixel viene determinato l'ingresso m_p (verso i neuroni piramidali) all' i -esima unità elementare dello strato target.

Sono stati utilizzati due diversi set di pattern, SET1 e SET2, formati da nove elementi l'uno. Definendo come "oggetti" l'insieme dei pixel "accesi" di un pattern, si può affermare che il primo set è composto da oggetti di dimensione costante (in particolare, 36px); il secondo set, invece, include oggetti di dimensione variabile (nel range 21÷42px). All'interno di ciascun set, i vari pattern

risultano essere ortogonali tra loro – non sussistono dunque “sovrapposizioni” tra un oggetto e l’altro.

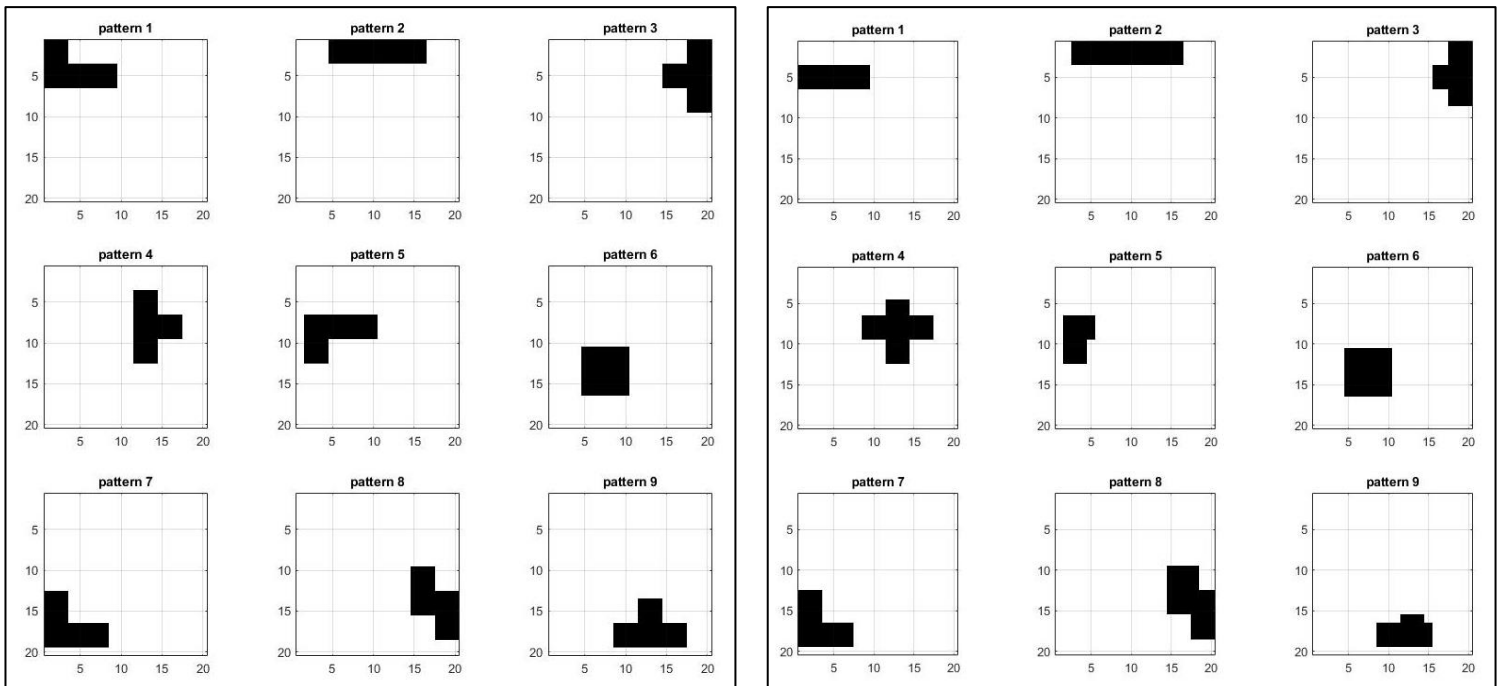


Figura 7 – A sinistra, il SET1, composto da oggetti a dimensione fissa; a destra, il SET2, composto da oggetti a dimensione variabile.

I singoli oggetti all’interno dei due set, come mostrato dalle figure, sono connessi. Tuttavia, questa non è una proprietà rilevante, in quanto il modello non si basa in alcun modo su coordinate spaziali – fatto testimoniato dall’applicabilità della vettorizzazione, attraverso la quale si viene largamente a perdere la vicinanza tra pixel “accesi”.

Per ogni set è stata considerata una sola possibile “sequenza”, quella che va, semplicemente, dal pattern1 al pattern9. In modalità *recall*, dunque, se è ad es. fornito in input il pattern2, si vorrà che gli strati L_2 e L_3 evochino (dopo aver ricevuto l’informazione sull’ingresso) i pattern dal terzo in poi, compatibilmente con la modulazione fornita da L_1 . La sequenza non è circolare: il pattern9 non richiama il pattern1.

Si consideri, infine, che per “pattern corrotti” si intendono pattern a cui è stato sottratto il 30% delle *features*; sono stati cioè portati a 0 il 30% dei pixel originariamente “accesi”. La selezione dei pixel da “spegnere” è fatta di volta in volta in maniera casuale.

Addestramento della rete

Il punto di partenza del training è una rete avente l'architettura descritta nel capitolo precedente, con connessioni sinaptiche fisse stabilite, e tutti i pesi delle sinapsi addestrabili inizializzati a zero. Posto che il procedimento di addestramento deve essere condotto nella sua interezza per ciascun set di pattern, quanto segue è valido per il singolo set: per individuare i valori corretti dei pesi, sono state eseguite tre fasi di apprendimento con maestro, distinte e consecutive l'una all'altra, basate su regole di tipo hebbiano e anti-hebbiano. Esse saranno esposte una per volta nelle sezioni seguenti.

Fase 1: sinapsi laterali in L_1

La prima sessione consiste nell'addestrare la rete auto-associativa contenuta in L_1 . L'obiettivo è dunque determinare gli elementi della matrice $W_{L_1L_1}$.

Durante la procedura, il maestro presenta allo strato L_1 i pattern del set correntemente in uso, non corrotti e uno per volta. Ciò è implementato agendo su m_p ; in particolare, è stato posto $m_p = 2000$ Hz (eccitazione molto forte) per tutte le colonne che corrispondono a pixel "accesi". Mentre L_1 sta ricevendo l' h -esimo pattern ($h = 1, \dots, 9$), la sinapsi tra le generiche unità i e j è rafforzata se entrambe risultano avere un'attività (proporzionale a z_p e normalizzata fra 0 e 1) superiore ad una certa soglia (T_{low1}), secondo una regola hebbiana (equazione X14). I pesi si incrementano progressivamente di un fattore che dipende sia dal learning rate (γ_w), sia dal valore massimo ammesso ($W_{L_1L_1}^{max}$) per la tipologia di pesi sinaptici in questione. In particolare, quanto più il peso si avvicina al valore massimo, tanto più le sue variazioni rallentano. Non sono ammessi auto-anelli (condizione $i \neq j$) – di conseguenza, la matrice $W_{L_1L_1}$ possiede una diagonale principale nulla.

L'effettivo aggiornamento dei pesi, eseguito con frequenza f_s (la quale dipende dal passo di campionamento nel dominio temporale, $T_s = f_s^{-1}$), inizia a rigore a seguito di un transitorio iniziale. Per diminuire però il costo

computazionale del processo, l'aggiornamento delle sinapsi avviene solo in una ristretta finestra temporale al termine dell'intervallo di presentazione di ciascun h -esimo pattern.

$$\Delta W_{L_1 L_1}^{i,j}(t) = \gamma_W \left(\frac{z_p^i(t)}{2e_0} - T_{low1} \right)^+ \left(\frac{z_p^j(t)}{2e_0} - T_{low1} \right)^+ (W_{L_1 L_1}^{max} - W_{L_1 L_1}^{i,j}(t)), i \neq j \quad (X14)$$

$$W_{L_1 L_1}^{i,j}(t + T_s) = W_{L_1 L_1}^{i,j}(t) + \Delta W_{L_1 L_1}^{i,j}(t)$$

I simboli “+” che compaiono come apice dei fattori relativi all'attività delle unità i e j indicano che del confronto con la soglia T_{low} viene considerata la sola parte positiva. Di conseguenza, se il termine tra parentesi è negativo perché l'attività della rispettiva colonna è sottosoglia, il fattore risulterà nullo e con esso anche la variazione complessiva del peso sinaptico. Risulta allora che la formula presentata non può che determinare $\Delta W_{L_1 L_1}^{i,j}(t) \geq 0$. L_1 , dunque, non possiede meccanismi per indebolire le connessioni apprese: non può “dimenticare”. Questa limitazione potrebbe essere superata in successivi sviluppi del presente modello.

Terminato l'aggiornamento delle sinapsi per ciascuno dei pattern del set in uso, viene eseguita, come ultimo step, un'operazione di normalizzazione: dopo aver fissato il valore massimo a cui può ammontare la somma dei pesi sinaptici in ciascuna riga di $W_{L_1 L_1}$ (TS_W) i pesi sono scalati di conseguenza. Questo consente di rappresentare il fatto che la colonna corticale postsinaptica è in grado di processare solo una certa quantità di neurotrasmettitore per volta; o in maniera equivalente, che le colonne presinaptiche posseggono quantità finite di neurotrasmettitore disponibile per la comunicazione sinaptica.

$$\forall i = 1, \dots, 400 \text{ sia } S_i^W = \sum_{j=1}^{400} W_{L_1 L_1}(i, j). \text{ Se } S_i^W > TS_W \quad (X15)$$

$$\Rightarrow W_{L_1 L_1}(i, j) \leftarrow W_{L_1 L_1}(i, j) \cdot \frac{TS_W}{S_i^W}$$

Terminato l'addestramento, le colonne eccitate da uno specifico oggetto in input si richiamano dunque l'un l'altra (eccitazione reciproca). Questa proprietà

della rete consente di ricostruire le informazioni mancanti in caso siano forniti in ingresso pattern corrotti. Inoltre, si sottolinea che quanto più è piccolo l'oggetto originale (non corrotto), tanto più forti saranno le sinapsi tra le relative colonne corticali; questo per via della scala dei fattori sinaptici descritta nel blocco X15.

Fase 2: sinapsi laterali in L₂ e L₃

Il secondo step dell'addestramento prevede di individuare i valori corretti per le sinapsi di tipo K e di tipo A, in L₂ e L₃. Viene imposto che le matrici di sinapsi siano addestrate in modo identico, e pertanto risultino uguali a coppie, cioè:

$$K_{L_2L_2} = K_{L_3L_3} \quad (X16)$$

$$A_{L_2L_2} = A_{L_3L_3} \quad (X17)$$

Per un'esposizione più concisa, saranno dunque di seguito riportate le sole equazioni di addestramento relative alle matrici di sinapsi laterali in L₂, tenendo presente che tali espressioni sono valide anche per L₃ previo un adeguato cambio di pedici.

Similmente a quanto accade per il training in L₁, il maestro fornisce allo strato da addestrare i pattern del set correntemente in uso, non corrotti e uno per volta; ciò è implementato ponendo $m_p = 2000$ Hz e $m_f = 2000$ Hz per tutte le colonne corticali che corrispondono ad un pixel "acceso". Da sottolineare dunque che per l'addestramento delle sinapsi K e A viene fornito input diretto anche alle popolazioni GABA_{fast} – cosa che non accade per l'addestramento di L₁ e delle sinapsi in feedback. Anche in questo caso, l'effettivo aggiornamento delle sinapsi avviene solo in una ristretta finestra temporale al termine della presentazione del singolo pattern.

Relativamente alle sinapsi di tipo K – la connessione tra le generiche unità i e j dello strato viene rafforzata, secondo una regola hebbiana, se l'attività piramidale presinaptica normalizzata (proporzionale a $z_{p,j}$) è maggiore di una certa soglia (T_{low2}) e lo è, al contempo, anche l'attività normalizzata degli interneuroni GABA_{fast} (proporzionale a $z_{f,i}$) nella colonna postsinaptica. La variazione del peso

è ovviamente impattata dal learning rate (γ_K), nonché da un fattore che dipende dal valore massimo stabilito per la tipologia di sinapsi in questione ($K_{L_2L_2}^{max}$).

$$\Delta K_{L_2L_2}^{i,j}(t) = \gamma_K \left(\frac{z_f^i(t)}{2e_0} - T_{low2} \right)^+ \left(\frac{z_p^j(t)}{2e_0} - T_{low2} \right)^+ \left(K_{L_2L_2}^{max} - K_{L_2L_2}^{i,j}(t) \right), i \neq j \quad (X18)$$

$$K_{L_2L_2}^{i,j}(t + T_s) = K_{L_2L_2}^{i,j}(t) + \Delta K_{L_2L_2}^{i,j}(t)$$

Parallelamente, le sinapsi di tipo A sono addestrate con una regola anti-hebbiana. In particolare, il peso tra le generiche unità i e j dello strato viene rafforzato se l'attività piramidale presinaptica normalizzata (proporzionale a $z_{p,j}$) è maggiore di una certa soglia (T_{low2}) e al contempo l'attività normalizzata degli interneuroni GABA_{fast} (proporzionale a $z_{f,i}$) nella colonna postsinaptica è al di sotto di una certa soglia (T_{up}). Brevemente, se la popolazione presinaptica è attiva e quella postsinaptica è silente. Influenzano l'aggiornamento anche il learning rate (γ_A) e il valore massimo stabilito per le sinapsi A ($A_{L_2L_2}^{max}$).

$$\Delta A_{L_2L_2}^{i,j}(t) = \gamma_A \left(T_{up} - \frac{z_f^i(t)}{2e_0} \right)^+ \left(\frac{z_p^j(t)}{2e_0} - T_{low2} \right)^+ \left(A_{L_2L_2}^{max} - A_{L_2L_2}^{i,j}(t) \right), i \neq j \quad (X19)$$

$$A_{L_2L_2}^{i,j}(t + T_s) = A_{L_2L_2}^{i,j}(t) + \Delta A_{L_2L_2}^{i,j}(t)$$

Per entrambe le tipologie di connessioni, non sono ammessi auto-anelli (condizione $i \neq j$). Inoltre, per i termini relativi al confronto con le soglie T_{low2} e T_{up} , si considera la sola parte positiva (simbolo “+” in apice); di conseguenza, come per le connessioni laterali in L_1 , le sinapsi K e A possono essere rafforzate ma non indebolite nel corso del tempo. Infine, per le stesse ragioni illustrate nella sezione precedente, una volta terminate le procedure di training per tutti i pattern del set, è eseguita un'operazione di normalizzazione sulle matrici ottenute.

$$\forall i = 1, \dots, 400 \text{ sia } S_i^K = \sum_{j=1}^{400} K_{L_2L_2}(i, j). \text{ Se } S_i^K > TS_K \quad (X20)$$

$$\Rightarrow K_{L_2L_2}(i, j) \leftarrow K_{L_2L_2}(i, j) \cdot \frac{TS_K}{S_i^K}$$

$$\forall i = 1, \dots, 400 \text{ sia } S_i^A = \sum_{j=1}^{400} A_{L_2L_2}(i, j). \text{ Se } S_i^A > TS_A \quad (X21)$$

$$\Rightarrow A_{L_2L_2}(i, j) \leftarrow A_{L_2L_2}(i, j) \cdot \frac{TS_A}{S_i^A}$$

Complessivamente, le matrici $K_{L_2L_2}$ e $K_{L_3L_3}$ consentono, all'interno del rispettivo *layer*, alle *features* di uno stesso oggetto non tanto di richiamarsi l'un l'altra (l'eccitazione piramidale, che durante l'addestramento è a carico del maestro, in simulazione sarà determinata dall'input sinaptico degli strati a monte), quanto piuttosto di sincronizzarsi in banda gamma. Per contro, attraverso le matrici $A_{L_2L_2}$ e $A_{L_3L_3}$, le *features* di un certo oggetto possono desincronizzarsi dalle *features* di qualunque altro oggetto (nonché dall'attività di background di colonne corticali non stimolate da alcuna *feature*). In questo modo, anche se le *features* di più pattern sono presentate alla rete contemporaneamente, solo uno di tali pattern tenderà a prevalere come “vincitore” della competizione ed emergerà all'interno del periodo gamma in corso. La normalizzazione delle sinapsi descritta nei blocchi X20 e X21 contribuisce a rendere questa competizione bilanciata, compensando per eventuali differenze di dimensione tra gli oggetti.

Fase 3: feedback da L₃ a L₂

L'ultimo step nell'addestramento della rete ha come obiettivo la determinazione degli elementi della matrice $W_{L_2L_3}$: deve essere stabilita l'etero-associazione tra i due strati che consente di richiamare sequenze di oggetti. A tal scopo, a differenza delle fasi di training precedenti, in questa procedura si agisce contemporaneamente sui due *layers* L₂ e L₃.

In sotto-sessioni separate, il maestro fornisce ingressi $m_p = 2700$ Hz agli strati in modo che, nell'*h*-esima sotto-sessione, mentre in L₂ sono stimolate le colonne dell'*h+1*-esimo oggetto, in L₃ siano stimolate quelle dell'*h*-esimo. Ciò è valido per $h = 1, \dots, 8$.

Anche il nono pattern potrebbe eventualmente essere presentato a L₃, ma poiché esso è sempre l'ultimo della rispettiva sequenza, e dunque non deve

richiamare alcun oggetto successivo, il maestro non dovrebbe parallelamente fornire alcun input in L_2 . Altro caso particolare è quello per cui in L_2 sia presentato il pattern1. Esso, in maniera speculare al pattern9, non deve essere richiamato da alcuna rappresentazione precedente; di conseguenza, il maestro in parallelo non dovrebbe fornire alcuna eccitazione in L_3 . Eseguire sotto-sessioni di addestramento in queste due condizioni peculiari è possibile, ma ammonta a un nulla di fatto, poiché non vengono formate sinapsi.

La regola di aggiornamento dei pesi è di tipo hebbiano, ed anche in questo caso viene effettivamente applicata in una breve finestra temporale al termine di ciascuna sotto-sessione. La sinapsi tra le generiche unità i (di L_2) e j (di L_3) è rafforzata se entrambe risultano avere un'attività (proporzionale a z_p e normalizzata) superiore ad una certa soglia (T_{low3}). Influiscono sulla variazione anche il learning rate (γ_{wb}) e il valore massimo fissato per questo tipo di connessioni ($W_{L_2L_3}^{max}$).

$$\Delta W_{L_2L_3}^{i,j}(t) = \gamma_{wb} \left(\frac{z_p^{L_2,i}(t)}{2e_0} - T_{low3} \right)^+ \left(\frac{z_p^{L_3,j}(t)}{2e_0} - T_{low3} \right)^+ (W_{L_2L_3}^{max} - W_{L_2L_3}^{i,j}(t)), i \neq j \quad (X22)$$

$$W_{L_2L_3}^{i,j}(t + T_s) = W_{L_2L_3}^{i,j}(t) + \Delta W_{L_2L_3}^{i,j}(t)$$

Anche in questo caso, per via del fatto che si considera la sola parte positiva dei confronti con la soglia T_{low3} , i pesi possono aumentare ma non diminuire. È inoltre per completezza inclusa la condizione $i \neq j$, anche se al momento essa non ha alcun particolare effetto, posto che si sta lavorando con due set i i cui rispettivi pattern sono tutti ortogonali tra loro.

A valle della procedura ora descritta non viene eseguita alcuna normalizzazione. Di conseguenza, tutti i pesi contenuti in $W_{L_2L_3}$ tendono a portarsi al valore $W_{L_2L_3}^{max}$. Il motivo per cui non si sono scalate queste sinapsi è legato al fatto che, contrariamente alle altre connessioni addestrabili, quelle in questione sono extra-layer. Supponendo che siano a più lungo raggio di quelle intra-layer, si è fatta l'ipotesi che ci siano meno restrizioni fisiologiche sulla disponibilità e sulla ricezione del neurotrasmettitore.

Per completare la descrizione sulle procedure di training, si riporta di seguito una tabella che contiene i valori impiegati per i parametri finora menzionati.

Parametro	Valore
T_s (s)	10^{-4}
T_{low1}	0.12
T_{low2}	0.8
T_{up}	0.6
T_{low3}	0.7
γ_w	0.1
γ_K	1
γ_A	1
γ_{wb}	10
$W_{L_1L_1}^{max}$	10
$K_{L_2L_2}^{max}, K_{L_3L_3}^{max}$	8
$A_{L_2L_2}^{max}, A_{L_3L_3}^{max}$	0.3
$W_{L_2L_3}^{max}$	11
TS_w	130
TS_K	160
TS_A	$\min(S_i^A)$

Tabella 5 - Valori parametri di training.

Risultati

In questo capitolo verranno discusse le performance della rete addestrata. In particolare, si illustreranno i risultati di 1) test parziali, che mettono alla prova un sottoinsieme degli strati della rete o impiegano una connettività limitata; 2) test completi, che sfruttano tutti i *layers* e tutte le sinapsi, ed esplorano le due modalità di funzionamento *recall* e *desynchronize*; 3) test in condizioni alterate, che mirano a simulare circostanze anomale (patologiche) o comunque peculiari.

In tutte le prove che saranno presentate, a meno che non sia specificato altrimenti, l'input dall'esterno porrà $m_p = 600$ Hz per le colonne corticali eccitate. Saranno di volta in volta riportati dati e figure relativi ad uno dei due set di pattern; si consideri implicito che i risultati sui due set sono paragonabili.

Come risultato preliminare, è anteposta alla discussione una figura (fig. 8) che mostra le matrici sinaptiche a valle delle procedure di training descritte in precedenza (per il solo SET2, per brevità). Osservando queste matrici si possono fare alcune considerazioni: $W_{L_1L_1}$ e $K_{L_2L_2}$, $K_{L_3L_3}$ sono simmetriche e soprattutto hanno la stessa forma. Ciò è in linea con le aspettative, in quanto entrambe le tipologie di sinapsi sono non nulle per le generiche unità i, j ($i \neq j$) che siano contemporaneamente attive (stimolate da uno stesso pattern). Inoltre, i pesi in queste matrici non hanno tutti il valore massimo rispettivo per via del fatto che il SET2 è composto da oggetti a dimensione variabile – le procedure di normalizzazione rendono più deboli le sinapsi relative a oggetti più grandi.

Si può notare poi che la struttura di $A_{L_2L_2}$ e $A_{L_3L_3}$ indica che tutte le unità del rispettivo strato ricevono tali sinapsi: non vi sono righe completamente a zero. Per contro, sono presenti colonne nulle (unità che non inviano alcuna sinapsi di tipo A), e in particolare esse fanno riferimento a quelle unità che non sono eccitate da nessun pattern. Questa struttura esemplifica il ruolo delle sinapsi A – ciascuna unità eccitata da un ingresso invia connessioni “desincronizzanti” a tutte le altre, meno quelle eccitate dallo stesso ingresso (ragion per cui i “vuoti” nelle matrici A ricalcano l'aspetto delle matrici K).

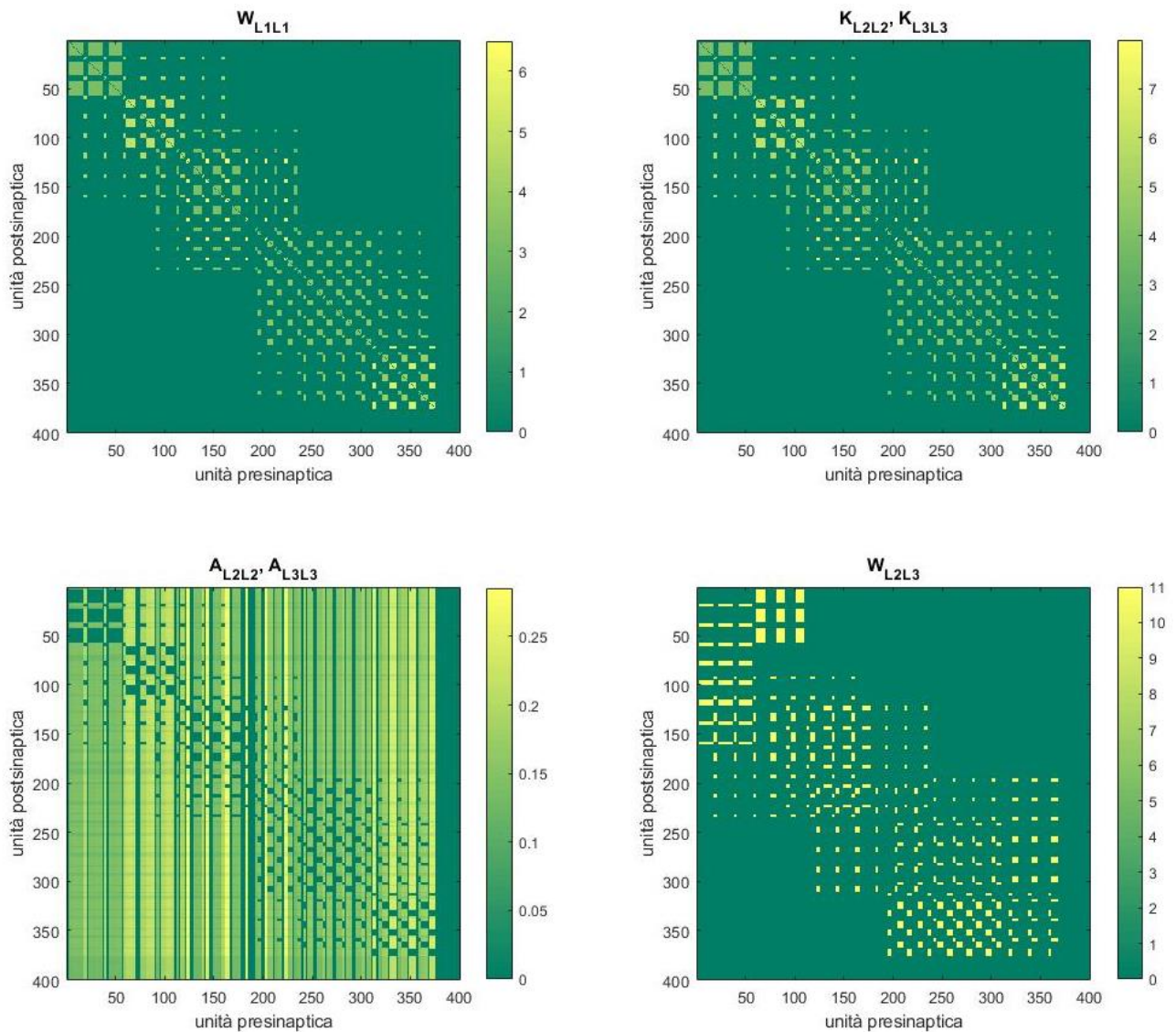


Figura 8 - Matrici delle sinapsi addestrabili a valle del training (SET2).

Test parziali

Il primo test parziale concerne WM e L_1 , e viene presentato per riassumere le caratteristiche salienti del comportamento di tali *layers*. Consiste nel fornire in ingresso a WM due pattern corrotti, singolarmente e ad un intervallo di 0.4 s l'uno dall'altro. Il fatto che i pattern siano corrotti implica che la media dell'ingresso ricevuto dalle colonne corticali associate a ogni pattern sarà ≈ 420 Hz, pari al 70% di 600 Hz. I valori dei parametri impiegati sono quelli relativi alla modalità *recall*.

I risultati sono mostrati in fig. 9. L'oscillazione in banda theta in L_1 è un risultato positivo subito individuabile. Si nota poi che, alla rimozione del primo ingresso (traccia blu), WM riesce a “tenere in memoria” il pattern – in particolare,

lo “ricorda” corrotto, cioè esattamente come è stato presentato. Possiamo affermare ciò poiché la frequenza media istantanea (piramidale) in WM quando L_1 è silente è pari a ≈ 3.5 Hz (il 70% dell’attività massima; dunque proprio il 70% delle colonne relative al pattern blu sono attive in quel momento). A conferire caratteristiche oscillatorie (in banda theta) all’andamento dell’attività in WM sono le connessioni in feedback provenienti da L_1 . Si può difatti osservare che la frequenza media istantanea in L_1 raggiunge i 3.5 Hz, e poi rapidamente sale a 5 Hz (segnalato dalla freccia nera continua) – dunque, lo strato ricostruisce con successo l’informazione mancante dell’oggetto in ingresso. Le *features* inizialmente corrotte e richiamate da L_1 attivano, attraverso il feedback W_{WML_1} , le corrispondenti unità in WM, la cui attività media riesce dunque anch’essa a raggiungere valori prossimi a 5 Hz (segnalato dalla freccia nera tratteggiata). Quando poi il corrente tratto ON del ritmo theta termina, le *features* che erano corrotte tornano a “spegnersi” in WM.

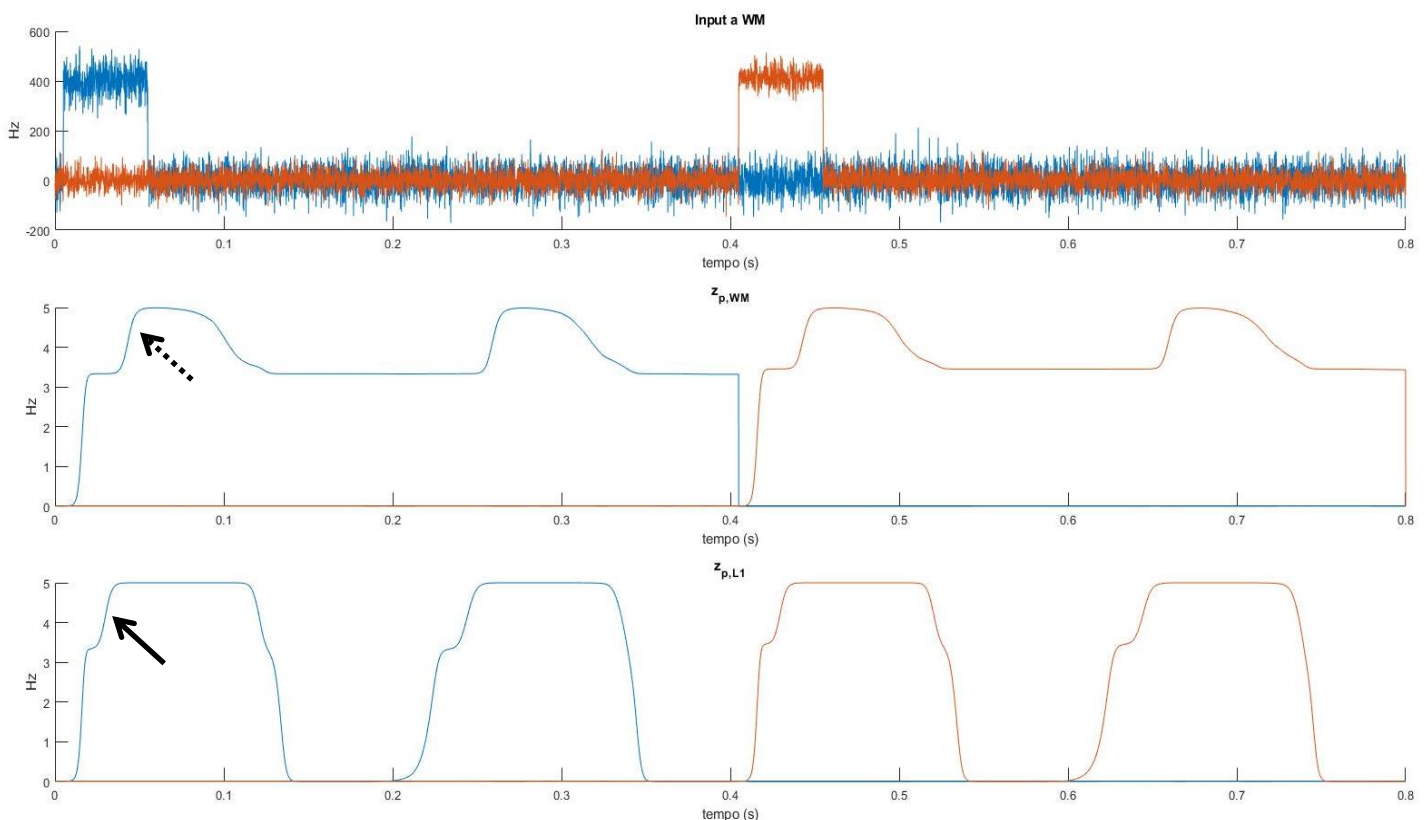


Figura 9 - Risultati test parziale su WM e L_1 . Dall'alto in basso: input a WM, frequenza media istantanea in WM, frequenza media istantanea in L_1 . Sono distinte con colori diversi le tracce relative a pattern differenti; ogni traccia fa riferimento alle sole unità eccitate dal rispettivo pattern. (SET2)

Si può fare un'ultima importante osservazione: al comparire del secondo input (traccia arancione), in WM è correttamente eseguito un “reset”, sono cioè portati a zero gli auto-anelli piramidali intra-colonna che mantenevano in memoria il primo pattern. Quando il secondo ingresso viene rimosso, è attivato un differente insieme di auto-anelli piramidali per il mantenimento in memoria di tale secondo input.

Il secondo test parziale è stato pensato per verificare che le sinapsi di tipo K e di tipo A riescano, rispettivamente, a sincronizzare e desincronizzare vari oggetti in banda gamma. L'ingresso, fornito a WM, è composto da due gruppi di pattern, presentati ad un intervallo di 0.4 s l'uno dall'altro. Ciascun gruppo è formato da tre pattern corrotti. La struttura degli strati WM e L₁ è equivalente a quella esposta per il test precedente.

Il test prende in esame tutti gli strati della rete, ma non include il feedback tra L₂ e L₃, né una modulazione diretta di L₂ da parte di L₁; quindi, complessivamente, nel blocco di equazioni X12 sono rimossi il termine dipendente da $W_{L_2L_3}$ e *Inibitore*. Si sottolinea che la combinazione di queste modifiche non altera il fatto che l'attività in L₂ sia possibile solo durante i tratti ON del ritmo theta (L₂ “spento” durante i tratti OFF). In assenza del feedback da L₃, difatti, l'unico segnale pilota di L₂ è l'attività in L₁; se L₁ è silente, lo sarà necessariamente anche L₂ (e di conseguenza L₃).

Gli andamenti degli ingressi e delle attività in WM e L₁ sono del tutto simili a quelli visti anche nel test precedente (fig. 10). L'unica differenza è data dal fatto che in questa prova le colonne corticali stimulate dai tre pattern che formano un gruppo sono attive contemporaneamente – e le tre tracce relative sono essenzialmente sovrapposte l'una all'altra. Si può quindi concludere che WM e L₁ non posseggono la capacità di discriminare pattern presentati insieme (giustamente).

È esaminando le attività degli strati L₂ e L₃ che si evidenzia, coerentemente con le aspettative, la discriminazione dei singoli pattern che formano un gruppo e quindi la simulazione dei fenomeni di *binding* e *segmentation* ad opera delle sinapsi laterali di tipo K e A (fig. 10). È importante sottolineare che le oscillazioni in questi *layers* rientrano in banda gamma. In entrambi gli strati si evidenzia la

presenza di una competizione tra i vari pattern, più caotica in L_2 di quanto non lo sia in L_3 , che individua di volta in volta un pattern “vincitore” che emerge per un periodo gamma. All’interno della finestra temporale individuata da ciascun tratto ON del ritmo theta, sono contenuti 4-5 periodi gamma in L_2 e L_3 , e ciascuno dei tre pattern che formano un gruppo tipicamente emerge almeno una volta.

Considerate le sue performance, si può interpretare il modello semplificato impiegato nel presente test parziale come un modello completo che, ricevuto un certo input, implementa i fenomeni di modulazione theta-gamma, *binding* e *segmentation*. Modelli simili sono stati difatti discussi in letteratura [15-16].

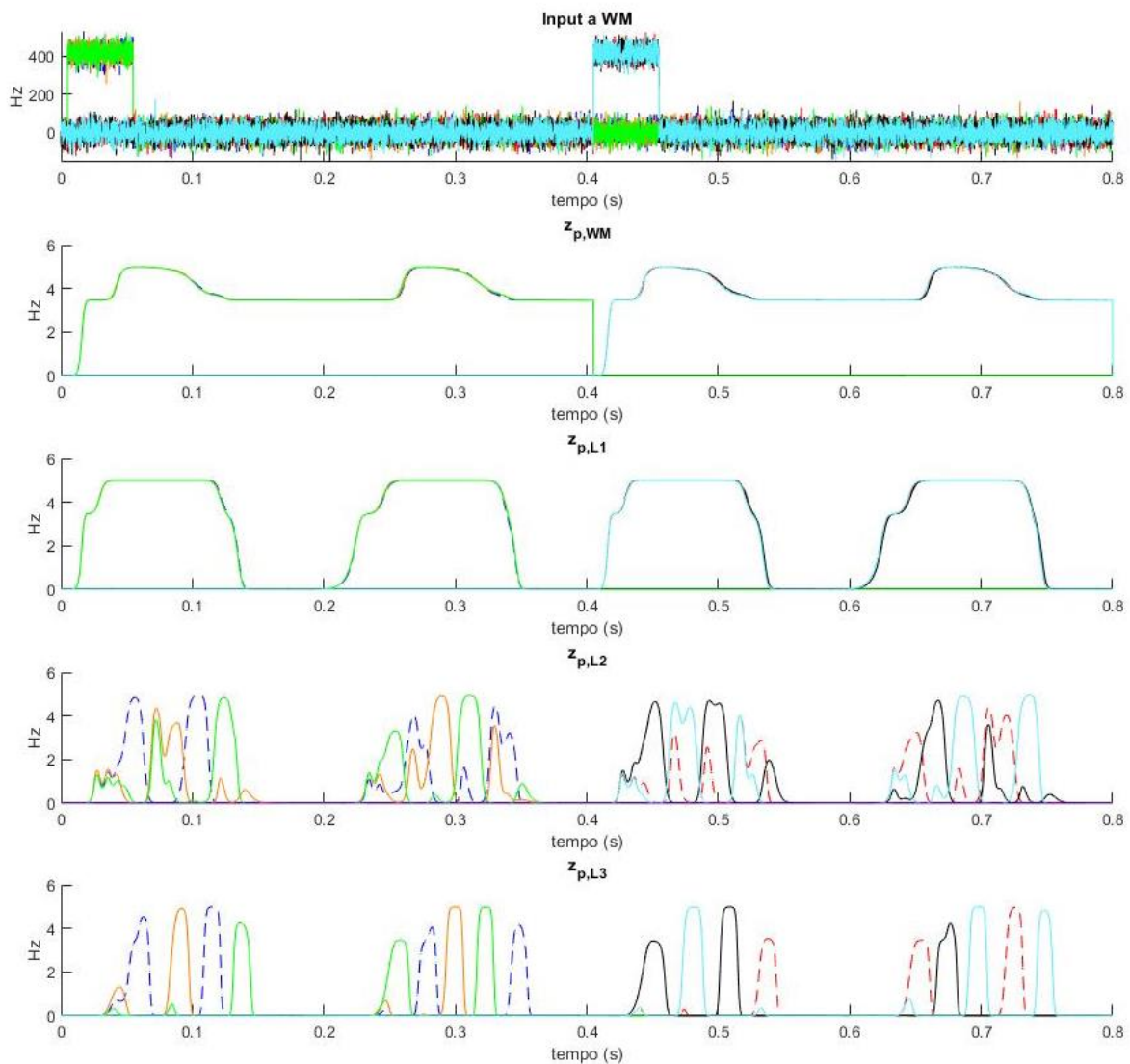


Figura 10 – Risultati test parziale di sincronizzazione e desincronizzazione in banda gamma. Dall’alto in basso: input a WM, frequenza media istantanea (piramidale) in WM, in L_1 , in L_2 , in L_3 (considerando solo le colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. (SET1)

Il terzo ed ultimo test parziale coinvolge solo L_2 e L_3 , ed ha lo scopo di sondare le prestazioni della rete etero-associativa che i due strati formano grazie alle sinapsi di feedback $W_{L_2L_3}$. L'ingresso è fornito direttamente a L_2 e consiste di una serie di pattern interi, presentati singolarmente ad un certo intervallo l'uno dall'altro. In questo caso i pattern non sono corrotti poiché, in assenza di L_1 , la rete è priva di un meccanismo che possa ripristinare le informazioni mancanti.

Le performance risultano del tutto soddisfacenti (fig. 11): la sequenza appresa durante il training viene rievocata con successo a seguito della breve presentazione di ciascuno dei pattern in input (si tratta, in particolare, dei pattern 2-4-6 del set, nell'ordine). Si noti però come, in queste condizioni, le oscillazioni nei due strati siano notevolmente più lente rispetto al test precedente. Questo è dovuto al fatto che nel test ora in esame, alla rimozione dell'input, i due strati sono pilotati unicamente dalla retroazione da L_3 verso L_2 (a cui segue una propagazione "passiva" da L_2 e L_3). Si può dunque concludere che l'eccitazione proveniente da L_1 impatta significativamente le caratteristiche dell'oscillazione in L_2 e L_3 .

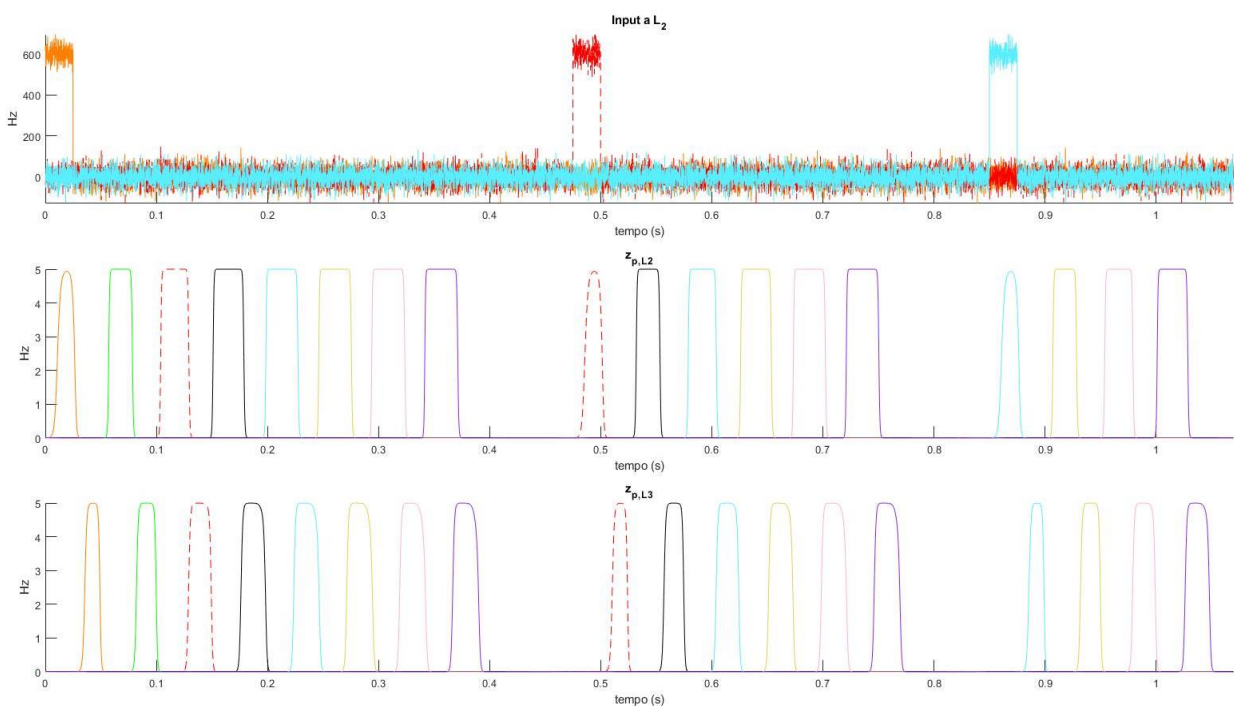


Figura 11 – Risultati test parziale di etero-associazione. Dall'alto verso il basso: ingresso a L_2 , frequenza media istantanea (piramidale) in L_2 , frequenza media istantanea (piramidale) in L_3 (considerando le sole colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. La sequenza è rievocata nell'ordine corretto per ciascuno dei tre input. (SET1)

Performance in modalità *recall*

Si ricorda che in questa modalità di funzionamento: 1) alla rete è fornito come ingresso un solo pattern alla volta; 2) è fondamentale che L_1 oscilli in banda theta; 3) L_1 esercita una modulazione diretta sull'attività degli strati a valle tramite il termine *Inibitore*; 4) L_2 e L_3 hanno il compito di rievocare la sequenza di oggetti appresa durante il training, oscillando in banda gamma.

Per rimarcare l'importanza del punto 3) appena esposto, si propone la figura seguente (fig. 12). Essa è stata ottenuta impiegando per semplicità i soli strati L_1 , L_2 e L_3 della rete, fornendo l'input dall'esterno a L_1 (in particolare l'input coincide con il pattern3 del set, corrotto e mantenuto poi costante per tutta la durata della simulazione) ed eliminando dalle equazioni che descrivono L_2 (X12) il termine *Inibitore*.

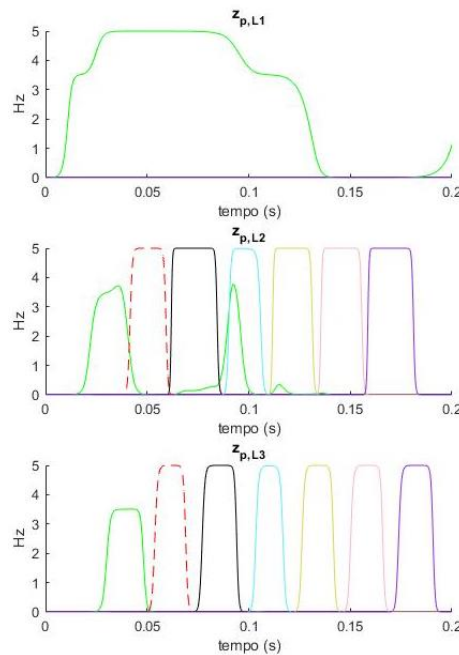
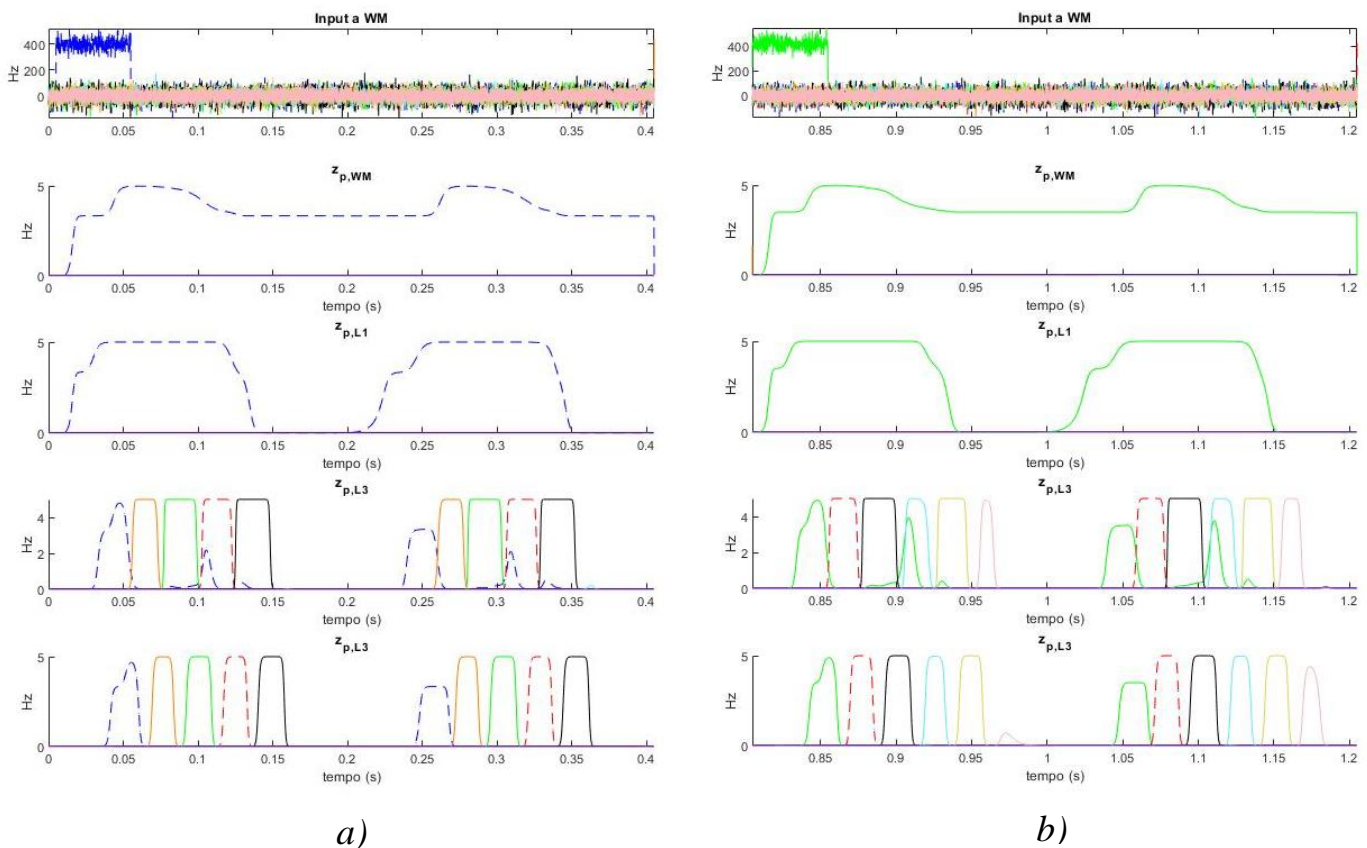


Figura 12 – Risultati di una prova in modalità *recall* condotta sugli strati L_1 , L_2 , L_3 . Dall'alto in basso: frequenza media istantanea (piramidale) in L_1 , in L_2 , in L_3 (considerando solo le colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. (SET2)

L'operazione fondamentale che la rete deve eseguire in questa modalità di funzionamento, cioè il *recall* di una sequenza eseguito mentre gli strati L_2 e L_3 oscillano in banda gamma, va chiaramente a buon fine: una volta che l'informazione sull'ingresso raggiunge L_3 , i pattern successivi nella sequenza sono evocati per un periodo gamma ciascuno (grazie alle connessioni di feedback

$W_{L_2L_3}$, il cui effetto prevale sull'eccitazione proveniente da L_1). L'output complessivo (frequenza media istantanea piramidale in L_3) è particolarmente ben strutturato. Tuttavia, il richiamo dei pattern non si arresta nel momento in cui il ritmo theta in L_1 entra nel tratto OFF. Questo è un comportamento non ottimale. Contraddice difatti i dati sperimentali, implicando che non sussista un vero e proprio accoppiamento tra i ritmi theta e gamma: il ritmo theta non guida il reset della sequenza evocata nello strato L_2 e la funzione di L_1 , al di là del recupero delle *features* corrotte dell'ingresso esterno, si riduce al fornire un input a L_2 .

Il termine *Inibitore* è stato incluso nel modello dopo che si sono fatte queste osservazioni, con l'obiettivo di stabilire un vero accoppiamento tra i ritmi nei vari *layers*. A tal proposito, si osservino le figure seguenti, che illustrano il comportamento della rete in modalità *recall* (tutti gli strati coinvolti, connettività completa). In input alla rete sono stati forniti, uno per volta e ad intervalli di 0.4 s, otto pattern del set (ciascuno presentato per 50 ms). Sono riportate le finestre temporali della simulazione per cui in input si hanno, nell'ordine, i pattern 1-3-5-7 del set.



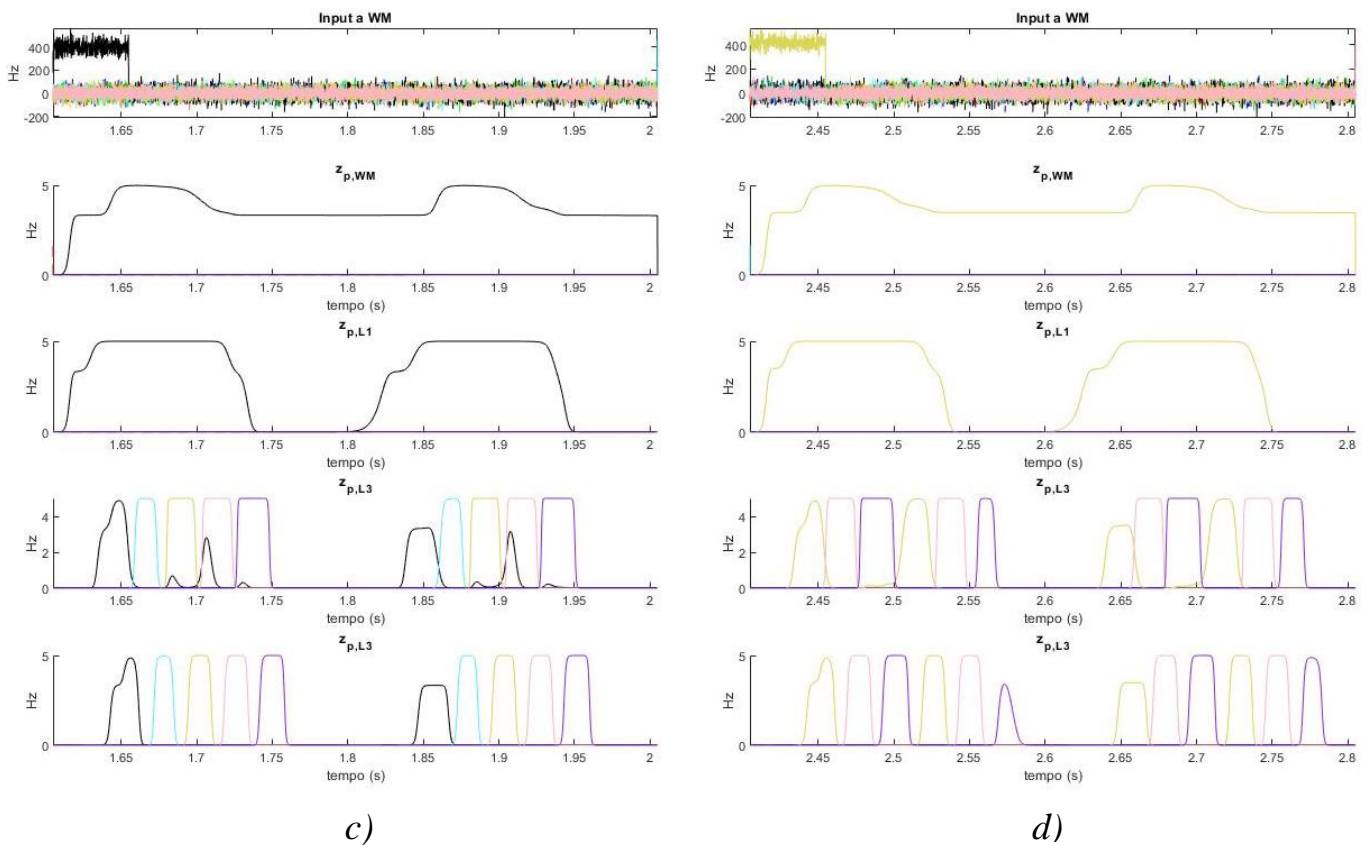


Figura 13 – Risultati di una prova in modalità recall condotta sulla rete intera, quando in ingresso a WM è posto a) il pattern1; b) il pattern3; c) il pattern5; d) il pattern7. Valido per tutti i pannelli: dall’alto verso il basso, input a WM, frequenza media istantanea (piramidale) in WM, in L₁, in L₂, in L₃ (considerando solo le colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. (SET2)

Coerentemente con le aspettative, la modulazione di L₂ e L₃ da parte di L₁ è tale da garantire l’accoppiamento dei ritmi theta e gamma: negli strati L₂ e L₃ è evocata la sequenza di oggetti opportuna fin quando L₁ è attivo, mentre durante i tratti OFF del ritmo theta il *recall* è bloccato. Inoltre, per ciascun tratto ON, sono evocati in L₂ e L₃ 5-6 pattern. Tal numero di elementi “ricordati” dalla rete in ogni finestra di lavoro è in accordo con le capacità della working memory documentate in letteratura [24-25].

Si noti anche che, se il pattern in ingresso è tale per cui la sequenza da evocare è composta da meno di 5-6 elementi (pannello *d* – sequenza di 3 elementi), la rete all’interno di uno stesso tratto ON cerca di evocarla più volte (sempre nel pannello *d*, negli strati L₂ e L₃ si ripete la serie di pattern 7-8-9 due volte per ogni tratto ON). Ciò è dovuto al fatto che, raggiunto l’ultimo elemento della sequenza appresa (pattern9), il feedback da L₃ non guida più l’attività in L₂,

che dunque torna ad essere pilotata da L_1 : riemerge così il pattern in ingresso in L_2 e il *recall* ricomincia da capo.

I valori delle sinapsi $W_{L_2L_3}$ e della soglia T (che compare nella definizione di *Inibitore*) hanno un impatto significativo sulle performance in questa modalità. Il modello risulta robusto a variazioni di T (testato $\pm 25\%$), ma non lo è particolarmente per variazioni dei pesi sinaptici $W_{L_2L_3}$.

Come ultima nota in merito alla modalità *recall*, si presenta la seguente prova: essa mira a testare la capacità di richiamo di sequenza della rete quando il set di pattern di lavoro è composto da elementi non tutti ortogonali tra loro. Si tratta di un caso rilevante per un modello che miri a simulare fenomeni legati alla memoria – è difatti più ragionevole, fisiologicamente parlando, lavorare con oggetti (in uno stesso set) parzialmente sovrapposti, piuttosto che con oggetti tutti ortogonali fra loro (il quale risulta essere un caso particolare, più semplice da trattare).

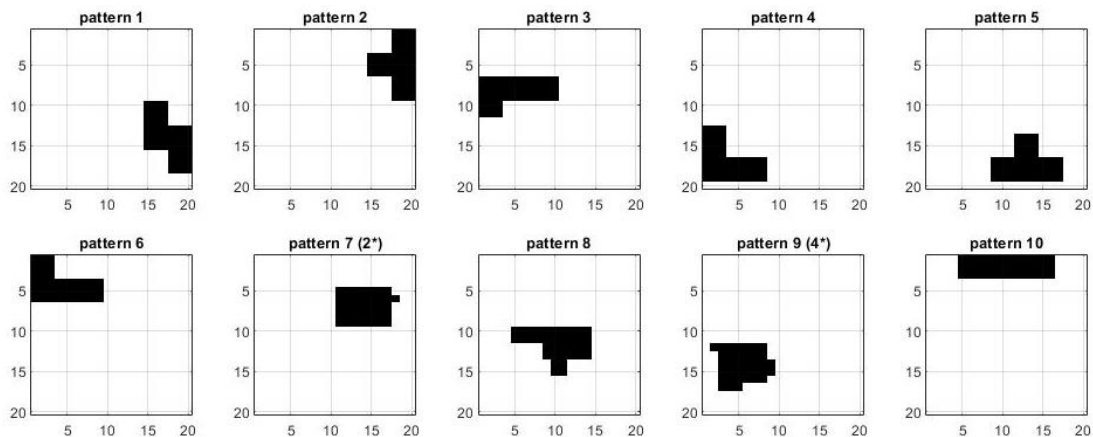


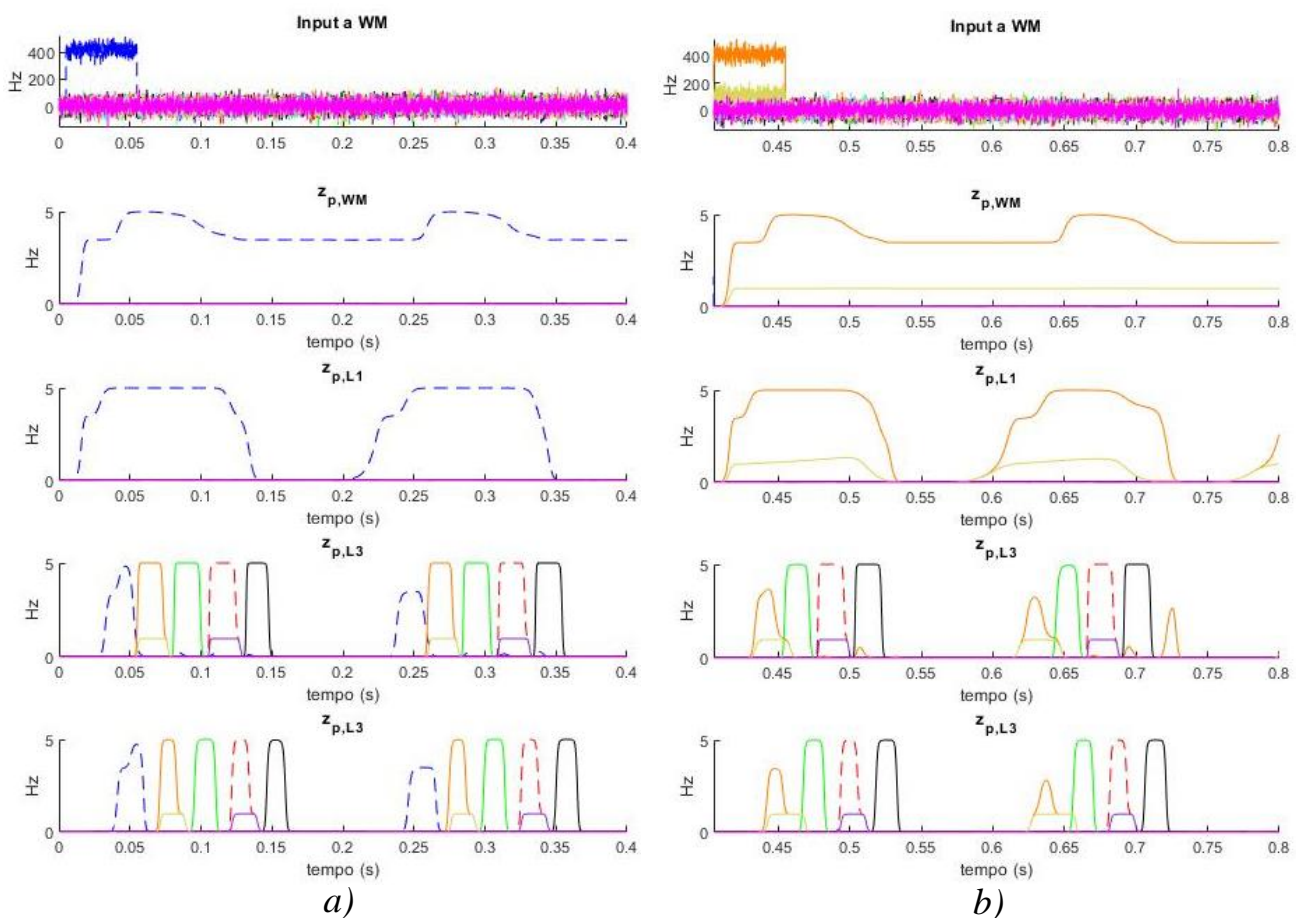
Figura 14 – SET3, composto da oggetti a dimensione fissa (36px), non tutti ortogonali tra loro. In particolare, a coppie, i pattern 2 e 7 ed i pattern 4 e 9 condividono circa il 20% delle features (pari a 7px).

È stato messo a punto, specificamente per questo test, un terzo set di pattern (SET3, fig. 14). Esso contiene 10 pattern, ciascuno dei quali ha 36px “accesi”. Due coppie, i pattern 2 e 7 ed i pattern 4 e 9, condividono circa il 20% delle features (7px). Tale percentuale di sovrapposizione è stata selezionata studiando le capacità di ricostruzione di L_1 : fornendo difatti a questo *layer* un pattern severamente corrotto, che mantiene solo circa il 20% delle proprie features, la ricostruzione fallisce – L_1 ha cioè troppe poche informazioni a disposizione per

richiamare le *features* non attive del pattern memorizzato. Impiegando allora una sovrapposizione del 20%, si assicura un corretto comportamento di L_1 anche nel caso peggiore: il caso cioè in cui si fornisce in ingresso alla rete un pattern corrotto (che mantiene il 70% delle proprie *features*) in cui nessuno dei pixel “sovrapposti” viene “spento” durante la corruzione, come sempre gestita da processi casuali.

Per questo set, la rete è stata addestrata tramite procedure simili a quelle illustrate nel precedente capitolo, con una sola accortezza aggiuntiva: le sinapsi relative a ogni pattern sono state addestrate individualmente, così che non si formassero connessioni spurie tra le colonne corticali eccitate dai pattern parzialmente sovrapposti. sono state considerate due sequenze: la prima, dal pattern1 al pattern5; la seconda, dal pattern6 al pattern10 – entrambe non circolari.

L’output generato dalla rete in modalità *recall* impiegando il SET3 è mostrato in fig. 15. In particolare, sono stati posti in ingresso a WM, per 50 ms ciascuno, i pattern del set, ad un intervallo di 0.4 s l’uno dall’altro. Nei pannelli della figura sono riportate per brevità solo alcune finestre temporali della simulazione completa.



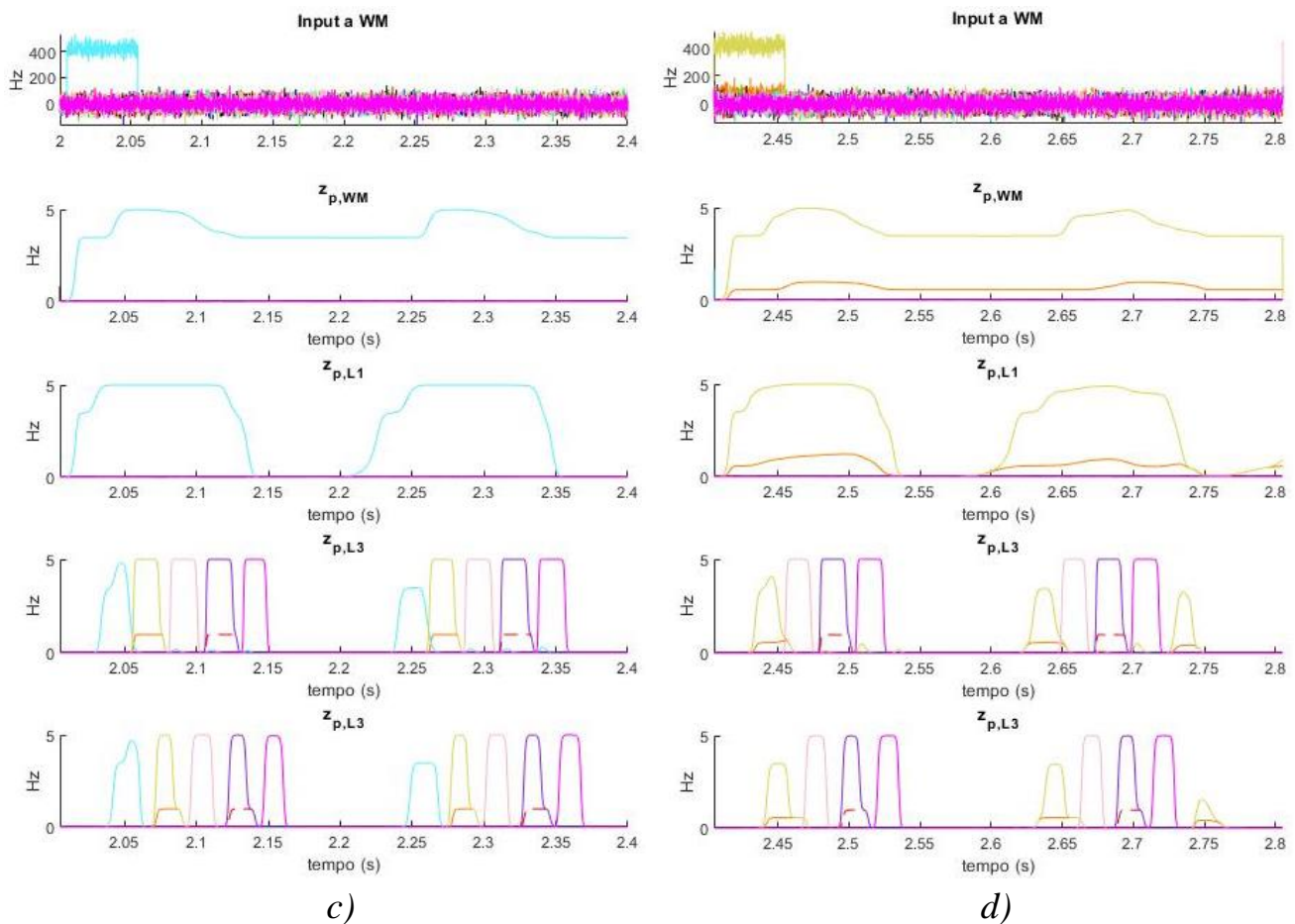


Figura 15 - Risultati di una prova in modalità recall, quando in ingresso a WM è posto a) il pattern1; b) il pattern2; c) il pattern6; d) il pattern7. Valido per tutti i pannelli: dall'alto verso il basso, input a WM, frequenza media istantanea (piramidale) in WM, in L₁, in L₂, in L₃ (considerando solo le colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. Le sequenze sono rievocate correttamente; la presenza di pattern parzialmente sovrapposti non ostacola le funzionalità della rete. (SET3)

Il comportamento dei *layers* è del tutto in linea con le aspettative: le attività in tutti gli strati hanno andamenti comparabili a quelli ottenibili lavorando con gli altri set di lavoro, il richiamo delle sequenze va a buon fine, i ritmi theta e gamma sono accoppiati e si evidenzia il fenomeno della precessione. Per ciascun tratto ON del ritmo theta sono evocati 4-5 pattern in L₂ e L₃, risultato sempre in linea con le nozioni disponibili in letteratura [24-25]. Si evidenzia un'unica differenza significativa rispetto ai risultati ottenuti con gli altri set, anch'essa attesa: in tutti gli strati, quando uno dei pattern non ortogonali viene evocato, è presente nei grafici una traccia che mostra, correttamente, come una certa percentuale delle colonne corticali attive faccia riferimento anche ad un altro pattern.

Performance in modalità *desynchronize*

In questa modalità di funzionamento: 1) sono forniti in ingresso a WM più pattern corrotti insieme; 2) l'oscillazione in L_1 e di conseguenza anche quella in WM sono soppresse: ciò è implementato rafforzando le sinapsi a lungo raggio reciproche tra i due *layers*; 3) è eliminato il feedback da L_3 a L_2 , dunque tali strati non recuperano più sequenze di oggetti. Piuttosto, modellando regioni prefrontali ascritte alla memoria di lavoro, risolvono un problema di *binding* e *segmentation* (ad es., segmentazione di una scena visiva).

Quanti più sono i pattern forniti contemporaneamente in input, tanto più il compito assegnato alla rete è complesso. Si è allora supposto che possano entrare in azione meccanismi di tipo attentivo quando il carico cognitivo è sufficientemente elevato, con l'obiettivo di facilitare la segmentazione dell'ingresso. I meccanismi attentivi sono stati modellati in maniera tale che influenzino la forza delle sinapsi laterali desincronizzanti in L_2 e L_3 ($A_{L_2L_2}$ e $A_{L_3L_3}$) e sono stati implementati impiegando la seguente formula:

$$\text{Se } nP > 3 \Rightarrow \begin{aligned} A_{L_2L_2} &\leftarrow A_{L_2L_2} \cdot (1 + 0.025 \cdot [nP - 3]) \\ A_{L_3L_3} &\leftarrow A_{L_3L_3} \cdot (1 + 0.025 \cdot [nP - 3]) \end{aligned} \quad (X23)$$

Dove nP rappresenta il numero dei pattern posti insieme in ingresso.

Dall'espressione si evince che le sinapsi di tipo A sono rafforzate quando in input sono presentati quattro o più pattern insieme. In particolare, per ogni pattern ulteriore dopo il terzo, la forza sinaptica è aumentata del 2.5% – il potenziamento varia dunque tra un minimo di +2.5% (quattro pattern) e un massimo di +15% (nove pattern).

Per iniziare a discutere il comportamento della rete in modalità *desynchronize*, si riporta il risultato di una simulazione per cui $nP = 3$ (fig. 16). Si tratta di un caso relativamente semplice, in cui non intervengono i meccanismi attentivi e che, per quel che riguarda la bontà delle operazioni di *binding* e *segmentation* eseguite dalla rete, si presta al confronto con la condizione descritta nel secondo test parziale illustrato in precedenza.

Come anticipato, i *layers* WM e L_1 non mostrano alcuna oscillazione. Continuano tuttavia a adempiere al proprio compito principale: WM mantiene in memoria le informazioni sugli ingressi quando essi sono rimossi, L_1 ricostruisce

le *features* corrotte. Da notare che le colonne corticali eccitate di entrambi gli strati si portano, a seguito di un breve transitorio, a una frequenza media istantanea di ≈ 5 Hz. Dunque, la ricostruzione in L_1 ha “permanentemente” aggiornato la rappresentazione dell’ingresso in WM (attraverso le sinapsi di feedback W_{WML_1}).

In quanto agli strati L_2 e L_3 – entrambi mostrano oscillazioni in range gamma, e le operazioni di *binding* e *segmentation* sono andate a buon fine. Similmente a quanto accade nel secondo test parziale, anche in questa prova si evidenzia una competizione tra i pattern in L_2 e L_3 , più caotica in L_2 di quanto non lo sia in L_3 . Laddove però nel secondo test parziale l’oscillazione di L_1 determina la segmentazione in L_3 di pattern “incompleti” (alcuni periodi gamma nella fig. 10 hanno picchi inferiori a 5 Hz), in questa prova quando un pattern emerge in L_3 esso viene segmentato nella sua interezza – solo raramente ciò non è verificato dopo il transitorio iniziale.

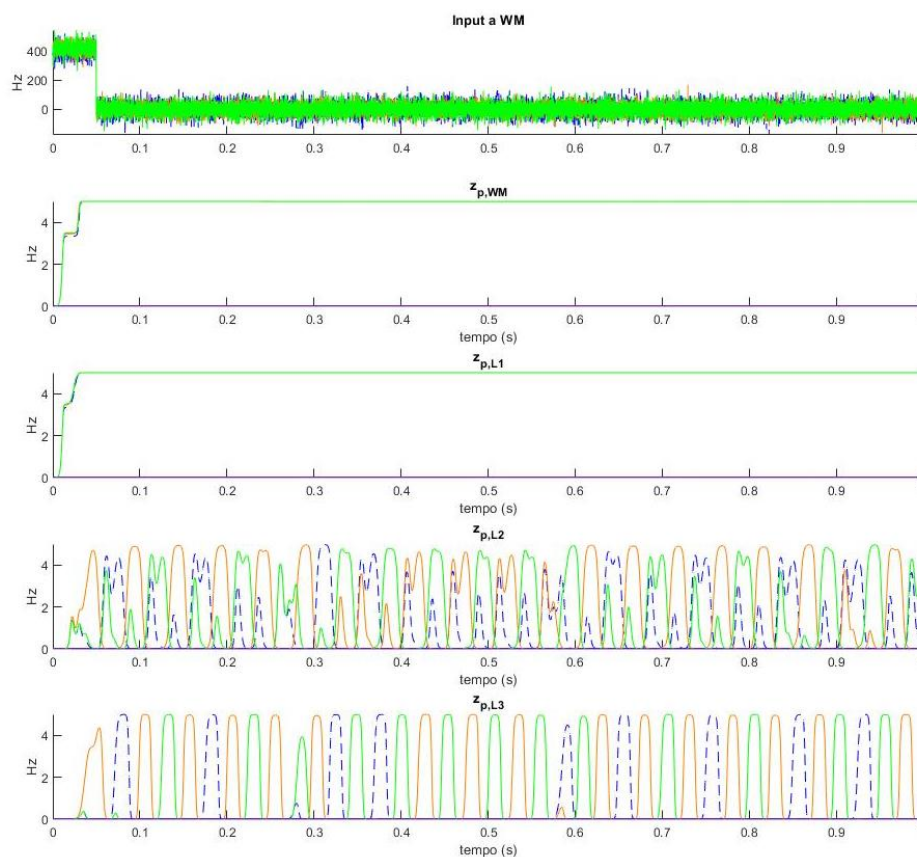


Figura 16 – Risultati di una prova in modalità *desynchronize*. Dall’alto in basso: input a WM, frequenza media istantanea (piramidale) in WM, in L_1 , in L_2 , in L_3 (considerando solo le colonne corticali eccitate dai rispettivi pattern). Sono distinte con stili diversi le tracce relative a pattern differenti. Come ingresso sono stati impiegati i pattern 1-2-3; utilizzando altre triplete di pattern si ottengono risultati del tutto comparabili. (SET2)

Si considerino ora i grafici seguenti (fig. 17), in cui sono state isolate le frequenze medie di spiking piramidale in L_3 al variare del numero di pattern in ingresso, nP , da 4 a 9 (dunque con meccanismi attentivi in azione). Si sono omessi gli andamenti delle attività negli altri strati e dell'ingresso poiché tali tracce non sono di particolare interesse e le loro caratteristiche qualitative sono già illustrate in fig. 16.

Come prima cosa si sottolinea che, per tutti i valori di nP , la rete riesce a segmentare con successo tutti i pattern presentati in ingresso (sono state eseguite numerose prove e questo risultato è robusto). Rispetto al caso con $nP = 3$, la maggiore complessità del compito per $nP = 4, 5, 6$ (pannelli *a, b, c*) si riflette in una attività in L_3 sensibilmente meno “pulita”: a volte la rete “perde un colpo” e non riesce a segmentare un pattern nel periodo gamma in corso. La frequenza di oscillazione rientra ancora in range gamma (≥ 25 Hz), sebbene essa diminuisca progressivamente all'aumentare di nP . La diminuzione della frequenza continua per $nP = 7, 8, 9$ (pannelli *d, e, f*), rientrando per tali casi in range beta (frequenza minima ≈ 20 Hz). La presenza del ritmo beta non è un risultato negativo: dati sperimentali ne supportano la pertinenza [14]. Il rinforzo progressivo delle sinapsi di tipo A ad opera dei meccanismi attentivi è il responsabile del rallentamento del ritmo, ed anche del ritorno ad un andamento più “ordinato” dell'attività in L_3 rispetto ai casi nei pannelli *a, b, c*.

Un'ultima osservazione che si può fare, valida per tutti i casi presentati (nP che varia da 3 a 9, fig. 16 e 17), è che i pattern in L_2 e L_3 non vengono segmentati in alcun ordine particolare. Le colonne corticali eccitate da ciascun pattern inibiscono le colonne corticali eccitate dagli altri (attraverso le sinapsi laterali di tipo A) ed in ogni periodo gamma (o beta) è segmentato il pattern vincitore di tale competizione. Nel codice che genera queste simulazioni è inclusa una componente aleatoria che consente di ottenere, a ogni ripetizione, un susseguirsi di pattern segmentati in ordine di volta in volta differente. L'assenza di un ordine specifico di evocazione si accompagna all'assenza dell'accoppiamento tra i ritmi theta e gamma (il ritmo theta in L_1 non è proprio presente); ciò è in accordo con i dati sperimentali, che mostrano come, in task che coinvolgono la memoria di lavoro, l'accoppiamento tra questi ritmi diviene man mano meno prominente quanto meno è importante l'ordinamento degli oggetti [26].

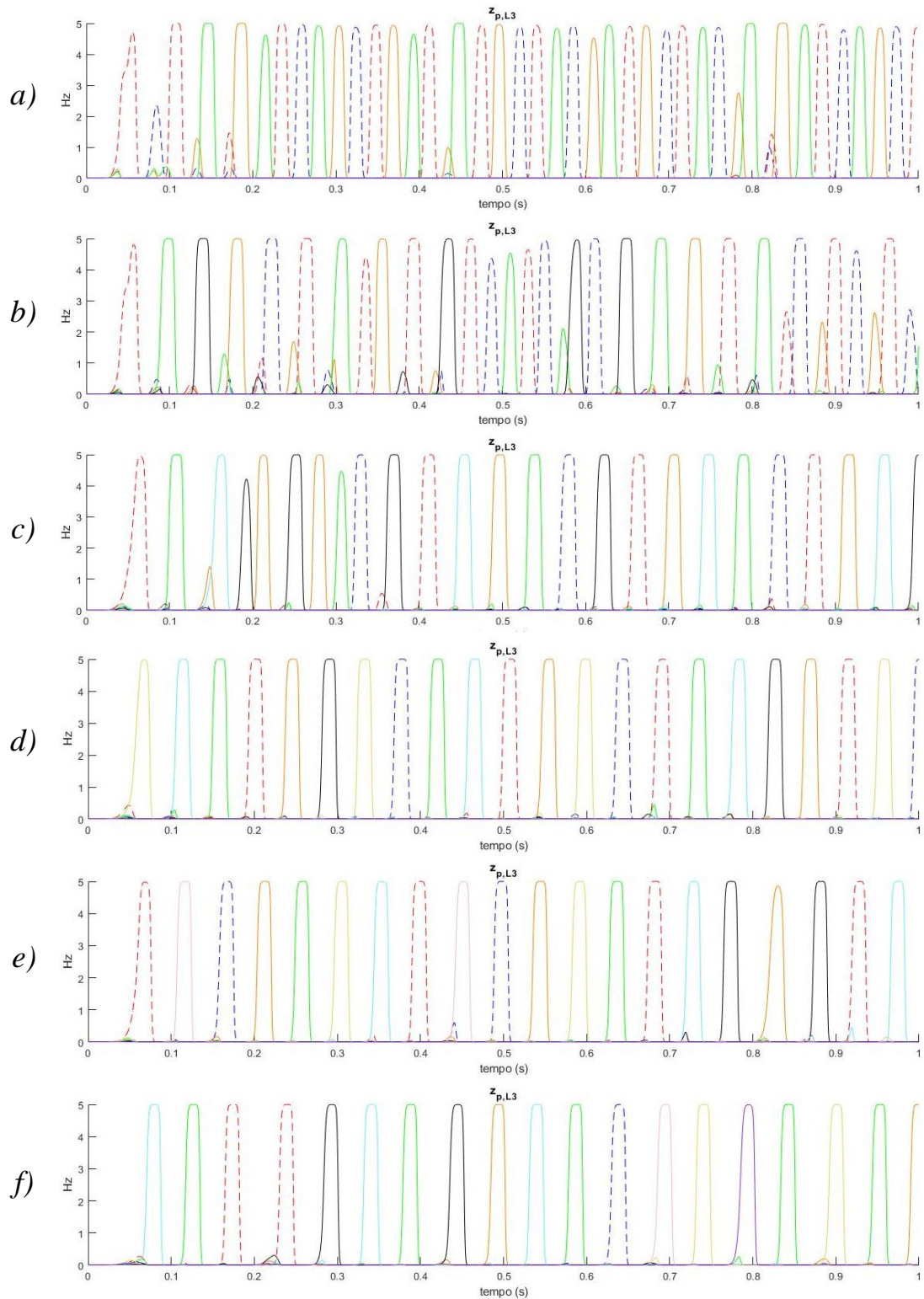


Figura 17 – Risultati di alcune prove in modalità desynchronize. È riportata la frequenza media istantanea in L_3 (considerando solo le colonne corticali eccitate dai rispettivi pattern). Rispettivamente, nei vari pannelli sono stati impiegati come ingresso: a) 4 pattern; b) 5 pattern; c) 6 pattern; d) 7 pattern; e) 8 pattern; f) 9 pattern. In tutti i pannelli, ciascuno dei pattern posti in ingresso viene evocato almeno una volta. Sono distinte con stili diversi le tracce relative a pattern differenti. (SET2)

Sempre in merito all'ordinamento dei pattern che vengono segmentati in L_2 e L_3 , si ritiene interessante menzionare un peculiare comportamento che la rete può avere in modalità *desynchronize*. Si consideri in caso in cui, in tale modalità, alla rete sia richiesto non di segmentare un insieme di oggetti presentati insieme (caso finora esaminato), ma piuttosto di tenere in memoria un certo numero di oggetti per un breve intervallo di tempo (che è esattamente ciò che accade, ad es., in task con delay). Mantenendo validi i meccanismi attentivi descritti nel blocco X23, se si riduce sensibilmente la forza delle sinapsi laterali di tipo A – in particolare, esse sono state ridotte a due terzi del loro valore originale – e si forniscono in ingresso a WM un certo numero di pattern insieme, emerge in L_3 l'andamento mostrato in fig. 18, in cui sono riportati i casi per cui $nP = 3, 4, 5$.

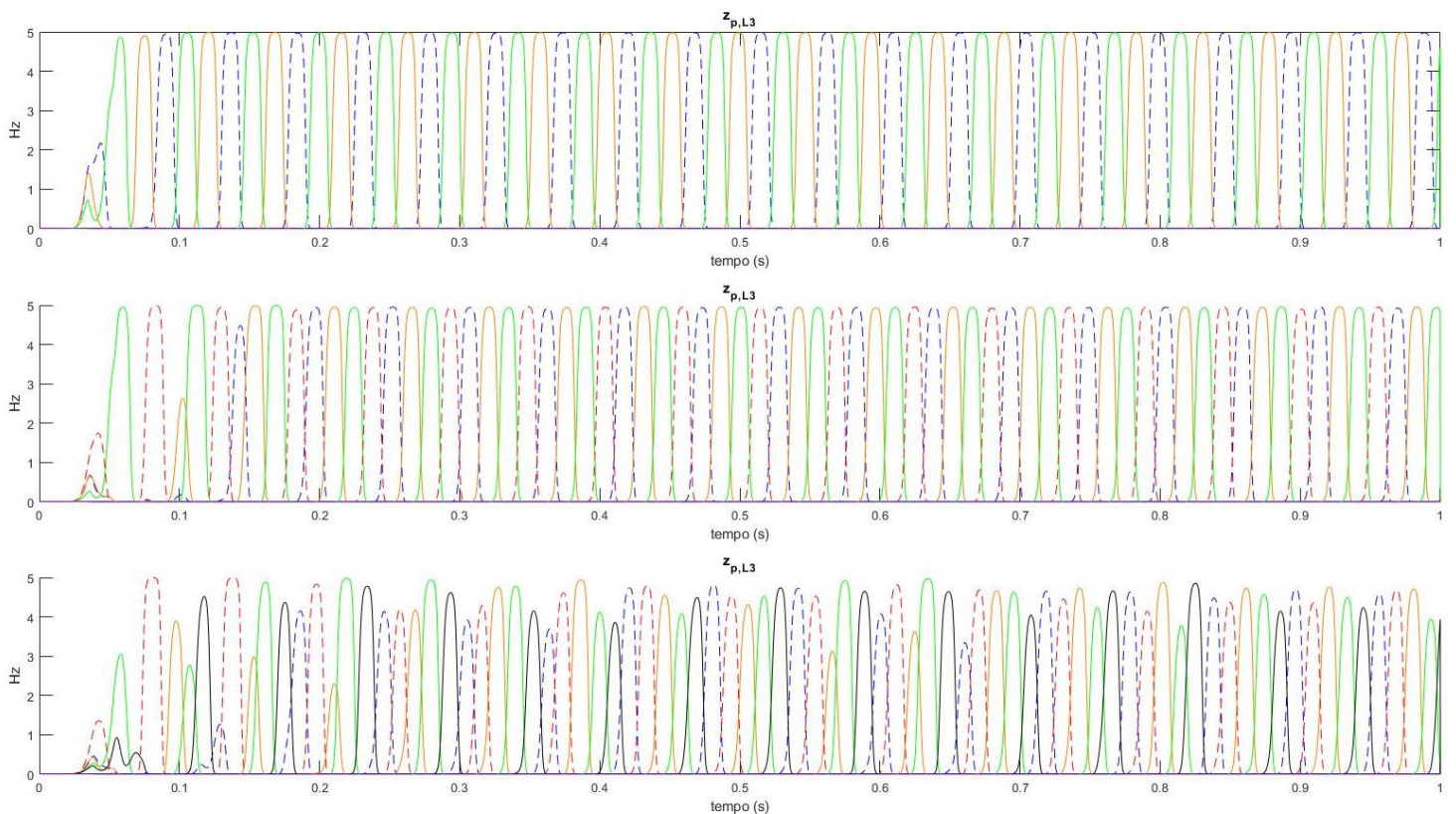


Figura 18 - Risultati di alcune prove in modalità desynchronize, con sinapsi A ridotte a due terzi del loro valore originale. È riportata la frequenza media istantanea in L_3 (considerando solo le colonne corticali eccitate dai rispettivi pattern). Rispettivamente, dall'alto in basso, sono stati posti in ingresso: 3 pattern; 4 pattern; 5 pattern. Sono distinte con stili diversi le tracce relative a pattern differenti. Si nota un ordinamento intrinseco degli oggetti segmentati. (SET1)

È immediato notare che la rete, a valle di un breve transitorio, inizia a segmentare i pattern in un ordine ben preciso; ordine che nessuno le ha imposto o

richiesto, ma che si presenta con consistenza ad ogni ripetizione, malgrado il codice contenga come già menzionato una componente aleatoria. Questo comportamento scompare se il numero di pattern da segmentare diventa elevato (già per nP pari a 5 si nota minor robustezza nella segmentazione; per nP pari a 6 o più l'ordinamento è assente). Il modello suggerisce dunque che, in particolari condizioni di lavoro, esista un meccanismo di “ordinamento intrinseco” nella working memory; un concetto che è espresso in letteratura [11]. In quanto poi alla definizione di tali “condizioni di lavoro” – si può immaginare che esistano meccanismi di controllo di tipo top-down, affini ai meccanismi attentivi, che consentono di regolare opportunamente la forza delle sinapsi a seconda del tipo di compito che il soggetto deve svolgere. A tale tipo di controlli attribuiamo dunque la riduzione delle sinapsi A ora implementata.

Test in condizioni alterate

Osservate le performance della rete in condizioni standard – cioè, in condizioni che mirano a riprodurre il comportamento, durante un certo task, delle corrispondenti regioni cerebrali in un soggetto sano – si è ritenuto opportuno investigare la capacità della rete di rispecchiare (almeno qualitativamente) peculiarità e anomalie di funzionamento che caratterizzano condizioni “alterate”. Con tale termine ci si riferisce al comportamento delle reti neurali naturali in 1) soggetti patologici e 2) particolari condizioni di stimolazione.

Il primo test concerne la malattia di Alzheimer, la quale determina, tra gli altri effetti, la perdita di alcune capacità di memoria. Numerose ricerche sperimentali hanno riportato che, in soggetti affetti da tale patologia, comportamenti anomali delle reti neurali relative alla memoria (ippocampali) sono influenzati da disfunzioni degli interneuroni – in particolare, degli interneuroni inibitori coinvolti nella generazione di oscillazioni gamma ($GABA_{fast}$) [27-28]. Esistono anche risultati che provano che, a partire dalla condizione patologica, correggere l'attività di questi neuroni può impattare lo stato della rete ippocampale e rendere l'attività di quest'ultima più simile a quella riscontrata in soggetti sani [29].

Per simulare la condizione patologica, si è dunque pensato di impiegare la rete in modalità *recall* ed agire sul coefficiente sinaptico intra-colonna che più impatta la genesi del ritmo gamma in L_2 e L_3 : C_{ff} . Riducendo progressivamente C_{ff} , gli strati WM e L_1 continuano ad operare normalmente, mentre L_2 e L_3 mostrano una crescente difficoltà nella rievocazione della sequenza appresa durante il training. Con C_{ff} ridotto ad un terzo del suo valore originale, L_2 ed L_3 iniziano a “perdere colpi” – evocano i pattern della sequenza in modo incompleto (frequenza media istantanea di picco in un certo ciclo gamma inferiore a 5 Hz) e il numero di pattern evocati in un tratto ON del ritmo theta scende da 5-6 a 4-5. Con C_{ff} ridotto ad un quarto del suo valore originale, l’andamento appena descritto è esacerbato: L_2 e L_3 continuano a “perdere colpi” e falliscono nel richiamare sequenza completa dopo l’evocazione di appena 2-3 pattern (fig. 19). Ciò si perpetua per ulteriori riduzioni di C_{ff} .

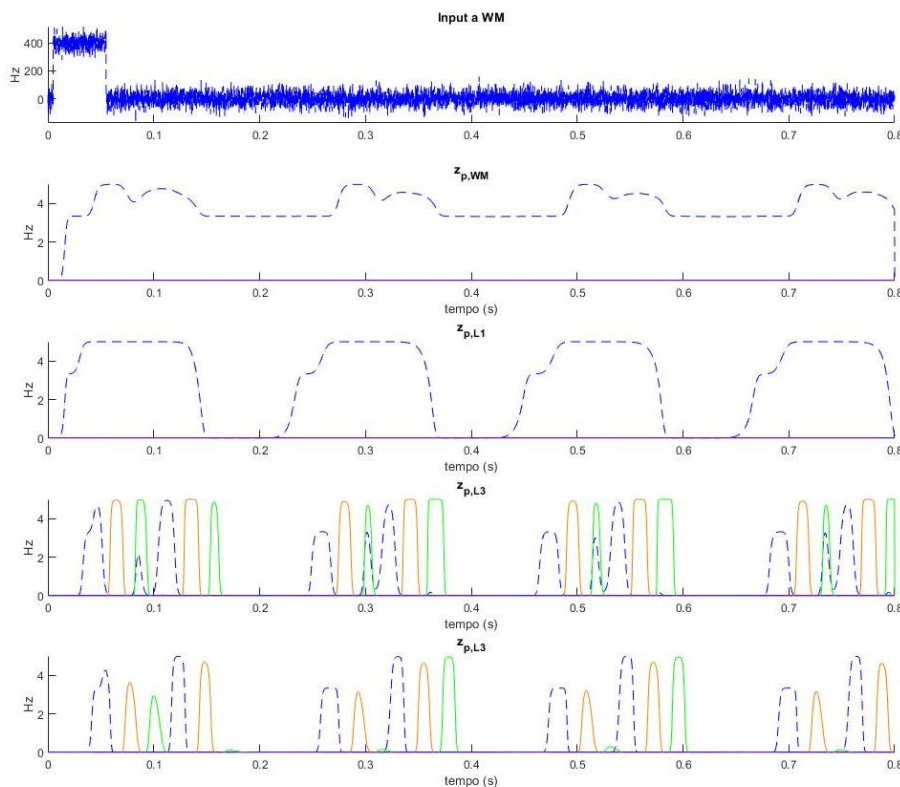


Figura 19 – Simulazione di condizione patologica: malattia di Alzheimer - risultati di una prova in modalità recall, con C_{ff} ridotto ad un quarto del suo valore originale (input: pattern1 del set, corrotto). Dall’alto in basso: input a WM, frequenza media istantanea in WM, L_1 , L_2 , L_3 . Sono distinte con colori diverse tracce relative a pattern differenti. Si nota che le capacità mnemoniche della rete sono significativamente alterate: il richiamo della sequenza fallisce dopo 2-3 elementi, ed i singoli pattern (in un certo ciclo gamma) sono spesso evocati solo parzialmente. (SET2)

Complessivamente, l'output prodotto dal modello in queste condizioni risulta soddisfacente ed in linea con le aspettative: le capacità mnemoniche della rete sono compromesse come nel caso patologico. Si sottolinea, tuttavia, che per riprodurre più fedelmente i risultati sperimentali (disfunzione dei neuroni $GABA_{fast}$) sarebbe opportuno modificare altri parametri sinaptici intra-colonna oltre a C_{ff} , in particolare C_{pf} (in modo da simulare popolazioni di interneuroni inibitori veloci alterati, che non riescono a generare corretti segnali presinaptici). Purtroppo, questo tipo di modifica non è applicabile nel modello discusso, in quanto un'alterazione nella comunicazione tra neuroni piramidali e $GABA_{fast}$ ostacola il ruolo del termine *Inibitore* (blocco di equazioni X12) dando luogo a un output non rappresentativo della sola condizione patologica in esame.

Il secondo test concerne la schizofrenia. È stato riportato in letteratura che soggetti che soffrono di questa patologia mostrano alterazioni dei recettori AMPA [30] (recettori coinvolti in quelle che nel presente modello sono chiamate sinapsi A), nonché un anomalo accoppiamento tra i ritmi theta e gamma nelle aree prefrontali durante compiti che coinvolgono la working memory [31]. Per simulare le condizioni patologiche si è pensato dunque di considerare la rete in entrambe le sue modalità principali, *recall* e *desynchronize*, riducendo la forza delle sinapsi A – in particolare, si è ridotto il loro valore ad un quarto di quello originale.

I risultati della prova in modalità *recall* sono mostrati in fig. 20. Gli strati WM e L_1 si comportano in modo normale. La modifica alle sinapsi determina però una minore modulazione, da parte del ritmo theta in L_1 , del ritmo gamma in L_2 e L_3 . Fondamentalmente, il termine *Inibitore* non riesce più a svolgere il proprio compito e l'accoppiamento tra i due ritmi è di conseguenza severamente alterato. Si può inoltre osservare che, sia in L_2 sia in L_3 , compaiono evocazioni di pattern spurie: dal punto di vista matematico, questo è dovuto a una competizione meno aggressiva tra i diversi pattern (le sinapsi A controllano, difatti, l'inibizione reciproca), mentre dal punto di vista fisiopatologico il fenomeno può essere collegato alla disorganizzazione di pensiero tipica della malattia.

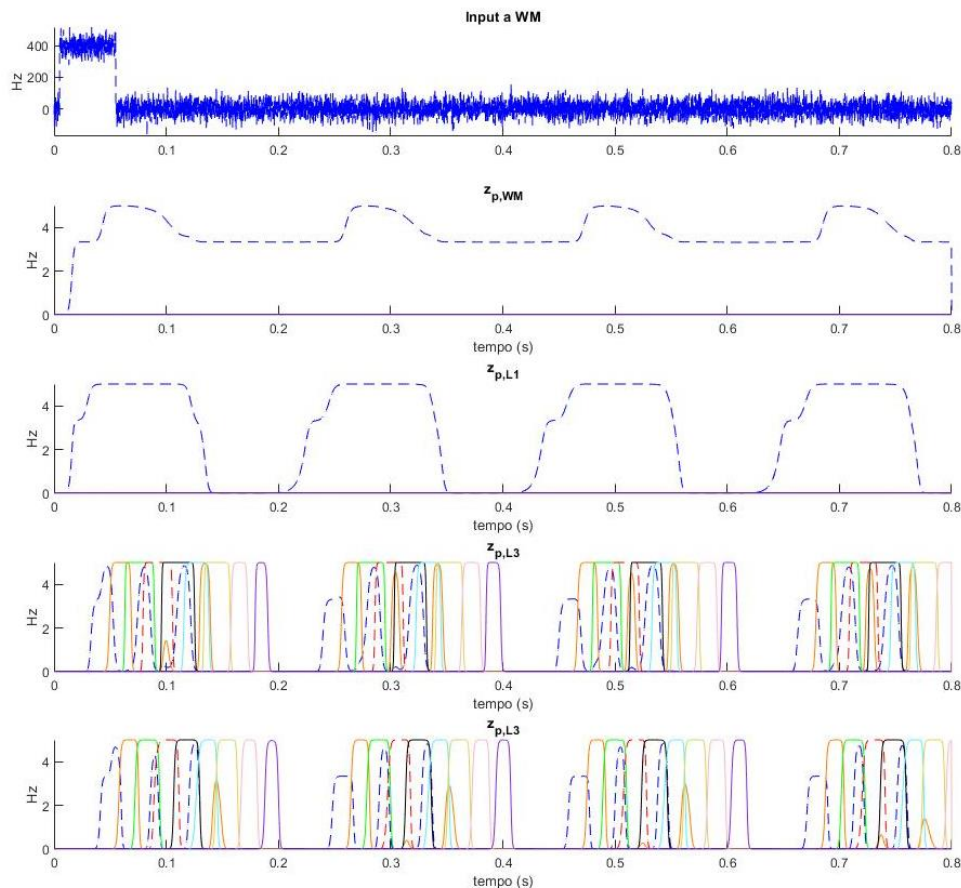


Figura 20 – Simulazione condizione patologica: schizofrenia - risultati di una prova in modalità recall, con sinapsi A ridotte ad un quarto del loro valore originale (input: pattern1 del set, corrotto). Dall'alto in basso: input a WM, frequenza media istantanea in WM, L₁, L₂, L₃. Sono distinte con colori diverse tracce relative a pattern differenti. Si evidenzia un alterato accoppiamento tra i ritmi theta e gamma, nonché la presenza di evocazioni spurie in L₂ e L₃. (SET2)

Il risultato della prova in modalità *desynchronize* è riportato nelle fig. 21 e 22. In particolare, la fig. 21 mostra gli andamenti dell'ingresso fornito alla rete e dell'attività in tutti gli strati quando nP (il numero di pattern correntemente presentati insieme in input) è pari a 3. In fig. 22, invece, sono per brevità riportate le sole frequenze medie istantanee delle colonne corticali di L₃, per nP pari a 5, 7, 9. Si sottolinea che i meccanismi attentivi che caratterizzano la modalità *desynchronize* (espressi nel blocco di equazioni X23) contribuiscono ai risultati mostrati in fig. 22 poiché sono normalmente implementati. Con ciò si intende che il potenziamento sinaptico dovuto all'attenzione è eseguito sulle sinapsi già ridotte al 25%: si è simulata l'alterazione delle sinapsi A, che di conseguenza rispondono diversamente ai meccanismi attentivi, ma non sono stati alterati i meccanismi attentivi stessi.

Ciò che si può osservare nelle figure è che, mentre WM e L₁ lavorano normalmente, in L₂ e L₃ (strati che in questa prova modellano regioni prefrontali ascritte alla memoria di lavoro) l'attività è significativamente più confusa rispetto al caso non patologico (mostrato in precedenza nelle fig. 16 e 17) – al punto che la segmentazione spesso fallisce. Colonne corticali che rappresentano le *features* di pattern diversi, difatti, risultano attive contemporaneamente. Come per la prova in modalità *recall*, ciò è spiegato matematicamente dal fatto che le sinapsi che controllano l'inibizione reciproca fra pattern (tipo A) sono diminuite, e dal punto di vista fisiopatologico il fenomeno può essere messo in relazione con il comportamento disorganizzato che caratterizza il paziente schizofrenico.

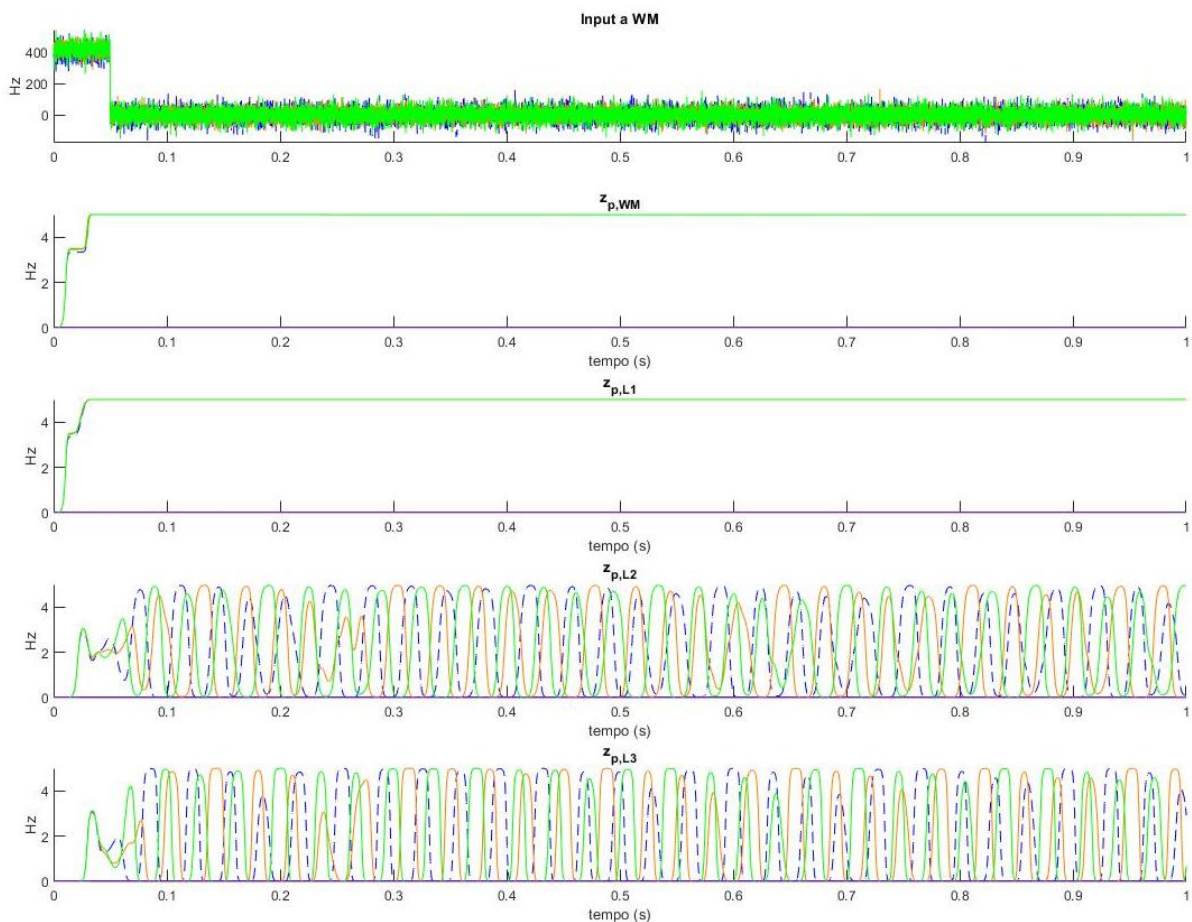


Figura 21 - Simulazione di una condizione patologica: schizofrenia – risultati di una prova in modalità desynchronize ($nP = 3$), con sinapsi A ridotte ad un quarto del loro valore originale. Dall'alto in basso: input a WM, frequenza media istantanea in WM, L₁, L₂, L₃. Sono distinte con colori diverse tracce relative a pattern differenti. Si evidenzia, rispetto alla condizione normale, maggiore disordine nella segmentazione dei pattern. (SET2)

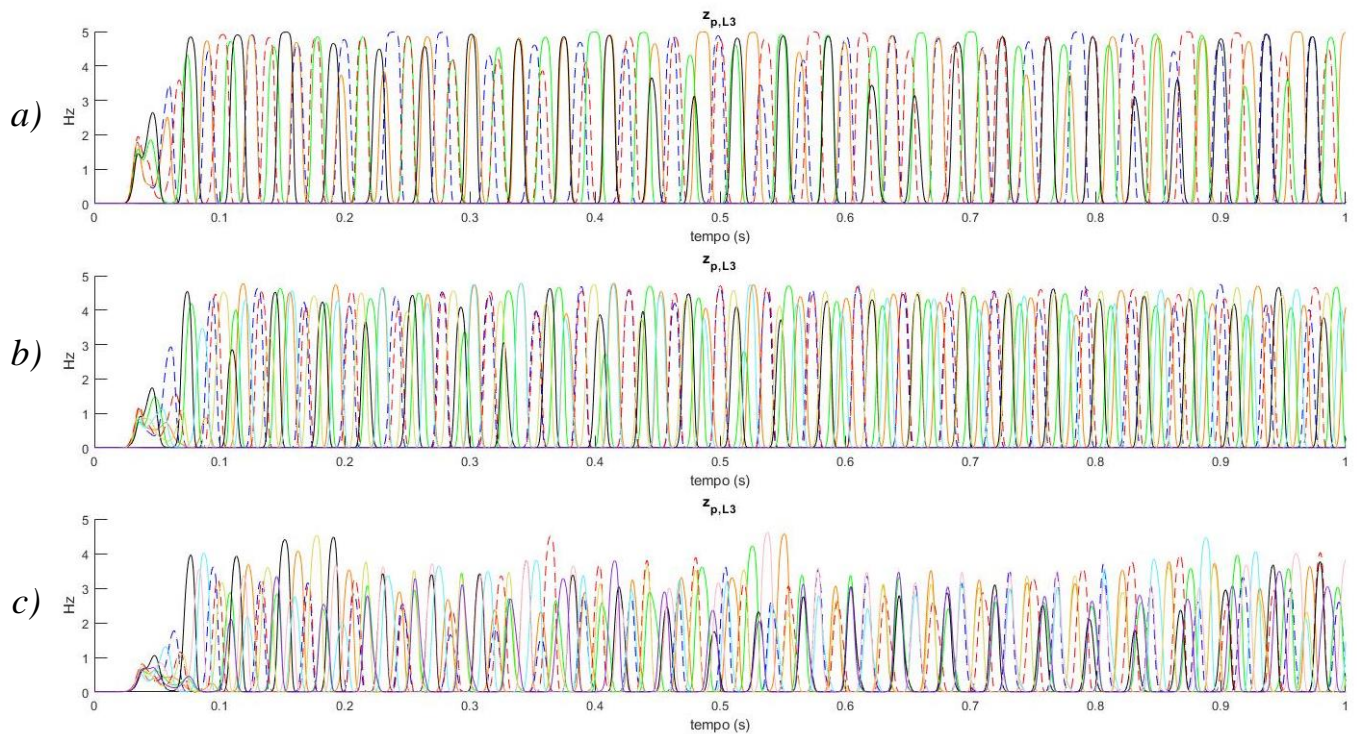


Figura 22 - Simulazione di una condizione patologica: schizofrenia – risultati di una prova in modalità desynchronize, con sinapsi A ridotte ad un quarto del loro valore originale. È riportata la frequenza media istantanea in L_3 , avendo posto in ingresso a) 5 pattern, b) 7 pattern, c) 9 pattern. Sono distinte con colori diverse tracce relative a pattern differenti. Si evidenzia molto disordine nell'attività, e la segmentazione dei pattern spesso fallisce. (SET2)

Complessivamente, l'output generato dalla rete nella modellazione del caso alterato in esame è soddisfacente: i risultati ottenuti consentono di confermare i dati sperimentali in merito all'importanza delle sinapsi che coinvolgono i recettori AMPA.

Il terzo test è relativo non ad una condizione patologica, ma alla condizione di esplorazione libera, o di sogno. Lo scopo è verificare se la rete è in grado di comportarsi creativamente, manipolando le sequenze apprese. Si suppone che il soggetto non abbia un particolare task da portare a termine, dunque la working memory non è coinvolta: lo strato WM è escluso dal modello. I layers L_1 , L_2 e L_3 (aventi l'architettura sinaptica della modalità *recall* e che si suppone modellino, in questa prova, regioni ippocampali) sono liberi di agire, in risposta ad un input (u_p , fornito a L_1) che consiste in una eccitazione non specifica variabile nel tempo. Tale input è cioè un segnale che viene fornito alle popolazioni piramidali di tutte

le colonne corticali che formano L_1 e, ad ogni passo temporale della simulazione, l'eccitazione che ogni colonna corticale riceve (m_p) cambia – in particolare, viene generata estraendo casualmente un numero reale dall'intervallo $75 \div 155$ (limiti dell'intervallo selezionati empiricamente) con distribuzione di probabilità uniforme. A tale numero è poi addizionato il rumore bianco a valor medio nullo (descritto da σ_p^2) che si sovrappone all'eccitazione esterna in tutte le prove.

In questo test è introdotta una variazione del parametro T , che compare nella definizione di *Inibitore* (riportata di seguito come riferimento). Si ricorda che T rappresenta la soglia con cui l'attività complessiva di L_1 viene confrontata – se tale attività risulta essere sottosoglia (tratti OFF del ritmo theta), gli strati L_2 e L_3 vengono silenziati.

$$Inibitore = R \cdot \frac{1}{2} \left(\left| T - \sum_j z_{p,j}^{L_1} \right| + \left(T - \sum_j z_{p,j}^{L_1} \right) \right) \quad (X24)$$

Nella condizione di esplorazione libera o sogno ora in esame, la soglia T è stata aumentata significativamente: da 20 a 100. Tale cambiamento è giustificato assumendo che T dipenda dal rapporto segnale-rumore (*signal-to-noise ratio*, SNR) in L_1 . Nelle normali modalità di funzionamento della rete, L_1 riceve un'eccitazione “strutturata” da WM ed ha di conseguenza un alto SNR; impiegando invece come ingresso allo strato il segnale u_p precedentemente descritto, l'attività in L_1 diviene più rumorosa e il SNR cala. La soglia T è adattata in modo che L_1 sia sempre in grado di modulare l'attività negli strati a valle (rendendo dunque il calcolo del termine *Inibitore* robusto alla presenza di rumore): se T non fosse stato variato da 20 a 100, L_2 non verrebbe mai appropriatamente inibito, neanche quando in L_1 non viene evocato alcun pattern.

I risultati ottenuti in questo test sono riportati in fig. 23. È immediato notare che in L_1 i pattern (tutti similmente eccitati dall'ingresso non specifico) compaiono in modo casuale, ricostruiti dalle connessioni laterali dello strato quando la frequenza media di spiking delle colonne corticali rispettive supera una certa soglia. Non c'è un particolare ordine di evocazione ed è possibile che più pattern emergano contemporaneamente. Questa peculiare attività guida gli strati

L_2 e L_3 , che oscillando in banda gamma evocano spezzoni della sequenza appresa (e a volte li concatenano) quando l'attività in L_1 è sopra soglia.

Considerate le condizioni di lavoro – una sola sequenza appresa, pattern ortogonali tra loro – la rete può mostrare un livello di creatività molto contenuto. Tuttavia, il risultato di questa prova è soddisfacente, in quanto dimostra che nel complesso il modello si presta ad esaminare il caso di esplorazione libera e illustra come la rete, lasciata a sé stessa, sia in grado di rievocare parti di quanto appreso.

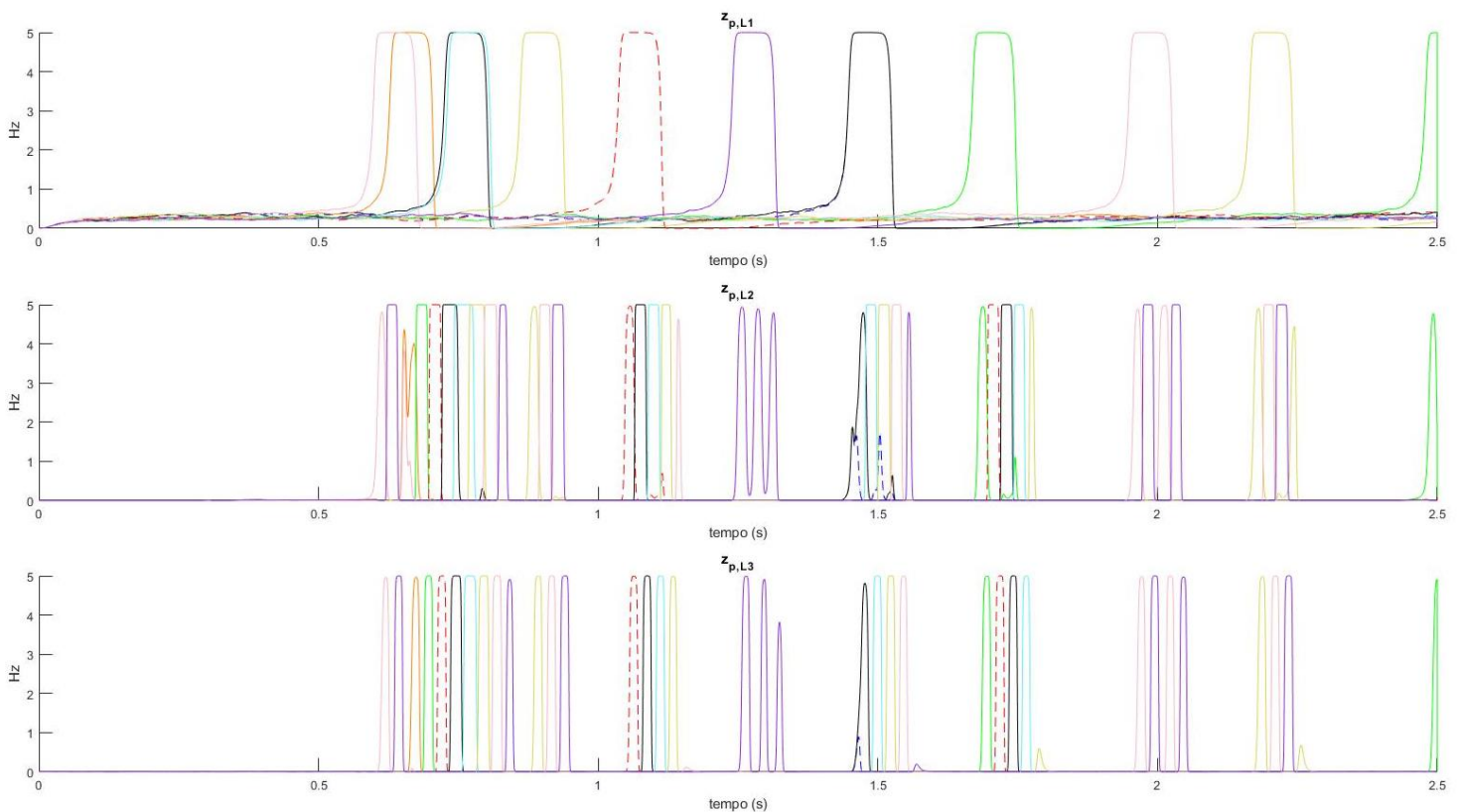


Figura 23 - Risultati del test in condizione di esplorazione libera/sogno. Dall'alto in basso: frequenza media istantanea in L_1 , L_2 , L_3 . Sono distinte con colori diverse tracce relative a pattern differenti. In L_1 emergono pattern in maniera casuale, ricostruiti dalle connessioni laterali quando la frequenza media di spiking delle colonne corticali rispettive supera una certa soglia. L_2 e L_3 richiamano spezzoni della sequenza appresa. (SET1)

Sviluppi futuri

Nei capitoli precedenti, sono state illustrate e discusse le prestazioni della rete sia nelle sue principali modalità di funzionamento, sia in condizioni alterate. Le simulazioni descritte delineano gli ambiti di applicazione del modello. Per ampliare tali ambiti di applicazione, o approfondire l'analisi in una particolare modalità di funzionamento, o affinare la specificità dei risultati (ad es., nell'ambito dello studio dei casi patologici o del caso di esplorazione libera), la rete beneficerebbe dall'implementazione di alcune modifiche. Alcune possibili modifiche saranno dunque esposte nei paragrafi seguenti, e potranno essere incluse in versioni future del modello.

Una prima proposta concerne le leggi di apprendimento. Si potrebbero impiegare regole simili a quelle descritte nel presente lavoro, sempre hebbiane e anti-hebbiane (si vedano i blocchi di equazioni X14, X18, X19 e X22), che però consentano di diminuire il valore dei pesi sinaptici, introducendo quindi un fattore di dimenticanza – fermo restando che un peso sinaptico non possa divenire negativo. Tale modifica consentirebbe alla rete di “dimenticare” nozioni apprese qualora non fossero ripresentate in ingresso per lungo tempo. Le regole di addestramento dovrebbero essere messe a punto in modo da garantire che l'indebolimento sinaptico sia relativamente più lento dell'incremento sinaptico – ad esempio, ponendo che le variazioni negative dei pesi sinaptici siano ridotte ad una loro frazione (come il 10%) prima di essere applicate. Questo garantirebbe che eventuali fluttuazioni dovute al rumore non impattino drasticamente il risultato finale.

Tale modifica delle leggi di apprendimento, di per sé, renderebbe il modello più fedele al caso fisiologico; le sinapsi naturali possono, difatti, indebolirsi nel tempo. Un ulteriore vantaggio è direttamente legato alla proposta successiva.

Si è menzionato in precedenza che l'impiego di pattern non tutti ortogonali tra loro è rilevante per modelli come quello descritto nel presente elaborato.

Questo poiché, in compiti che coinvolgono la memoria di lavoro, è più ragionevole e fisiologico avere a che fare con “oggetti” che abbiano *features* in comune (e siano dunque parzialmente sovrapposti) piuttosto che con “oggetti” tutti ortogonali tra loro. Sarebbe dunque opportuno impiegare uno o più set di pattern composti da “oggetti” che condividano (a coppie o a gruppi) alcune *features*. Il SET3, descritto in occasione della discussione dei risultati ottenuti in modalità *recall*, è un semplice esempio di set di pattern di questo tipo.

L’impiego di set di pattern non ortogonali, per necessità, dovrebbe accompagnarsi alla modifica alle leggi di apprendimento precedentemente proposta. Per garantire che la rete si comporti come desiderato, anche quando la percentuale di sovrapposizione tra le caratteristiche degli “oggetti” è relativamente alta (ad es., 50%), sarebbe difatti fondamentale consentire alle sinapsi di indebolirsi; la rete dovrebbe poter “dimenticare”. Per comprendere la ragione di ciò, si consideri il seguente esempio: siano i pattern P1 e P2 parzialmente sovrapposti; in particolare, siano: F1 l’insieme delle *features* di P1, F2 l’insieme delle *features* di P2, e F_s l’intersezione tra F1 ed F2 – dunque F_s rappresenta l’insieme delle *features* condivise dai pattern. Si consideri lo strato L₁ ed il suo compito di ricostruzione di pattern corrotti. Posto che in ingresso sia fornito P1 corrotto, per un corretto funzionamento del *layer*, è fondamentale che le caratteristiche nell’insieme F1 si richiamino tra loro – senza evocare *features* che non appartengono a P1. Dunque, le *features* che appartengono esclusivamente a P1 dovrebbero richiamarsi vicendevolmente e richiamare le *features* nell’insieme F_s; mentre le *features* nell’insieme F_s (poiché appartenenti anche a P2), non dovrebbero richiamare alcuna *feature*, onde evitare che evochino il pattern sbagliato. È allora evidente che, durante l’addestramento di L₁, le sinapsi che partono dalle *features* nell’insieme F_s e hanno come obiettivo le *features* esclusive di P1 – sinapsi che potrebbero in prima battuta formarsi – dovrebbero essere appropriatamente indebolite nel momento in cui è presentato il pattern P2. Tale indebolimento dei pesi sinaptici potrebbe essere ottenuto, ad esempio, impiegando (nella formulazione della regola hebbiana per l’aggiornamento delle sinapsi) soglie pre- e postsinaptiche diverse tra loro. In particolare, scegliendo una soglia postsinaptica significativamente maggiore rispetto a quella presinaptica. Sarebbe inoltre necessario, in fase di training, presentare alla rete ciascun pattern

più volte. Così facendo, la rete avrebbe infatti occasione di indebolire tutte le sinapsi che vanno dalle *features* condivise a quelle non condivise di ciascun pattern. Procedure di addestramento di questo genere sono già state impiegate, ad esempio, nell'ambito della modellazione della memoria semantica [32].

La possibilità di formare sinapsi unilaterali come quelle appena descritte, e di conseguenza la possibilità di impiegare largamente, anche in più modalità, set di pattern non ortogonali, è il secondo vantaggio, prima menzionato, legato alla modifica delle leggi sinaptiche. Si sottolinea che, se implementate correttamente, leggi di training come quelle proposte non altererebbero, rispetto al modello da noi realizzato e descritto, la formazione delle sinapsi (né, dunque, il comportamento della rete) nel caso in cui fossero impiegati set di pattern tutti ortogonali tra loro. Si avrebbe cioè a che fare con un modello più generalizzato che include al proprio interno quello discusso nel presente lavoro.

Un terzo suggerimento concerne le sequenze di pattern; esse sono definite, per ogni set di lavoro, durante la fase di training e vengono rievocate durante alcune simulazioni. Supponendo di avere a disposizione set di pattern non ortogonali – si potrebbero definire molteplici sequenze per ogni set, tali per cui gli elementi di una sequenza siano parzialmente sovrapposti a quelli di un'altra (come accade nel SET3 da noi impiegato). Con percentuali di sovrapposizione tra pattern sufficientemente elevate e appropriate condizioni di lavoro (ad es., elevato rumore in ingresso agli strati della rete), nell'ambito del test che analizza la condizione di esplorazione libera o di sogno si potrebbero ottenere interessanti comportamenti “creativi”. La rete potrebbe “confondersi” e ricombinare le sequenze apprese creandone di nuove.

Un'ulteriore modifica implementabile concerne i modelli di massa neuronale che formano le unità di base di ciascun *layer*. I NMMs da noi utilizzati impiegano tutti gli stessi parametri, a priori dallo strato di appartenenza; si potrebbe pensare di modificare tali parametri in modo da renderli differenti da strato a strato, così che le unità di calcolo di ciascun *layer* siano in grado di generare (senza necessità di connessioni con altre colonne corticali) oscillazioni rientranti in un certo range di frequenze. Si ricorda che, nel modello descritto in questo lavoro, tutte le unità, nelle condizioni ora specificate, oscillano in banda alfa (si veda la fig. 4). Agendo opportunamente sui parametri si potrebbero

ottenere oscillazioni in range theta per L_1 , e range gamma per L_2 e L_3 . Le caratteristiche e le performance di un network così costruito si presterebbero poi ad un confronto con il modello da noi realizzato.

Infine, un importante aspetto da esplorare, impiegando la rete con la sua architettura attuale o sue versioni future, è il confronto con dati sperimentali. Ci si potrebbe porre l'obiettivo di simulare particolari condizioni sperimentali, ad esempio alcuni test che coinvolgano la working memory; oppure si potrebbe approfondire l'analisi delle condizioni patologiche, implementando modifiche ai parametri ulteriori (ed in accordo con la letteratura) a quelle da noi prese in esame.

Conclusioni

Nel presente elaborato è stata descritta l'architettura di una rete neurale composta da quattro strati, che include: 1) meccanismi di tipo auto-associativo (nel *layer* L_1) per la ricostruzione di ingressi parzialmente corrotti; 2) meccanismi di tipo etero-associativo (tra i *layers* L_2 e L_3) per la rievocazione di sequenze di oggetti precedentemente apprese; 3) meccanismi che consentono di mantenere online l'informazione sull'ingresso (nel *layer* WM), dopo che esso è stato fornito per un breve intervallo di tempo e poi rimosso. La rete è stata addestrata con successo, impiegando due set di pattern, ciascuno composto da oggetti ortogonali tra loro (un set con oggetti a dimensione fissa, l'altro con oggetti a dimensione variabile), e facendo uso di regole di apprendimento hebbiane e anti-hebbiane. Sono state individuate due principali modalità di lavoro: *recall* e *desynchronize*.

In modalità *recall*, la rete modella verosimilmente il comportamento della memoria di lavoro in un compito che richiede il reclutamento delle regioni ippocampali. A valle della presentazione di un certo oggetto in input, ne sono, difatti, richiamati altri: è rievocata una sequenza consolidata durante il training. In questa modalità, emergono i ritmi theta (in L_1) e gamma (in L_2 e L_3) e la rete consente di simularne l'accoppiamento: il ritmo theta individua "finestre di lavoro" entro le quali emerge l'attività in banda gamma, in accordo con i dati sperimentali [3-9]. Il numero di periodi gamma all'interno di ciascun periodo theta (variabile tra 4 e 6) è anch'esso compatibile con le nozioni disponibili in letteratura [24-25]. È stato inoltre verificato che impiegare pattern che non siano tutti perfettamente ortogonali (ma siano piuttosto sovrapposti parzialmente) non ostacola il funzionamento del modello, purché il grado di sovrapposizione sia limitato.

In modalità *desynchronize*, la rete modella verosimilmente l'attività in aree prefrontali durante task che richiedono il mantenimento, nella memoria di lavoro, di diversi oggetti presentati in input simultaneamente. Le operazioni principali che devono essere eseguite sono quelle di *binding* e *segmentation*. Sono stati

inclusi nel modello anche meccanismi attentivi che entrano in gioco quando in ingresso sono posti 4 o più oggetti; tali meccanismi potenziano le sinapsi che consentono di effettuare la segmentazione e dunque migliorano le prestazioni della rete. In questa modalità, non si evidenzia accoppiamento tra i ritmi theta e gamma (il ritmo theta è del tutto assente). Le oscillazioni in banda gamma (presenti negli strati L_2 e L_3) possono mostrare comportamenti estremamente interessanti – in particolare, supponendo che sia applicata una modulazione delle sinapsi laterali desincronizzanti tale da ridurre la forza a due terzi del valore originale, si verifica un fenomeno di “ordinamento intrinseco” nella memoria di lavoro: se il numero di oggetti in input è tutto sommato ridotto (pari o inferiore a 5), essi sono automaticamente segmentati in un ordine ben preciso, che la rete crea da sé. Questa nozione – l’ordinamento intrinseco – è in accordo con teorie proposte in letteratura [11]. Se non è applicata la riduzione della forza delle sinapsi laterali desincronizzanti, e impiegando i meccanismi attentivi sopracitati, la rete riesce a segmentare con successo anche numerosi oggetti (ad es., 9). In tal caso il fenomeno dell’ordinamento automatico però svanisce, ed il ritmo gamma rallenta progressivamente fino a portarsi in banda beta; ritmo anch’esso pertinente secondo i dati disponibili in letteratura [14].

La modifica di alcuni parametri del modello consente alla rete di replicare, almeno qualitativamente, le caratteristiche di alcuni casi patologici. In particolare, sono state ricreate con successo: 1) la perdita di capacità mnemoniche (in modalità *recall*) evidenziata in soggetti affetti dalla malattia di Alzheimer [27-29]; 2) condizioni caratterizzate da elevato disordine (in entrambe le principali modalità di lavoro) e con alterato accoppiamento theta-gamma (in modalità *recall*), compatibili con il caso schizofrenico [30-31]. La rete si è anche mostrata capace, quando eccitata in maniera non specifica e privata dello strato WM, di evocare liberamente porzioni di quanto appreso (simulando dunque la condizione di esplorazione libera o di sogno).

Sono state, infine, presentate e discusse alcune possibili modifiche che potrebbero essere implementate in future versioni della rete. Tali proposte includono: 1) variazioni alle leggi di apprendimento; 2) l’impiego di set di pattern non ortogonali tra loro; 3) la definizione di molteplici sequenze che si intersechino fra loro (caso che è stato esplorato solo in modo contenuto nel presente lavoro) –

ciò, in particolare, in congiunzione con la modifica precedente; 4) l'impiego di unità di calcolo di base (modelli di massa neuronale) diverse rispetto a quelle da noi scelte.

Complessivamente, si può affermare che il modello realizzato è un buono strumento per la simulazione e l'analisi dei ritmi theta e gamma – e del loro accoppiamento – nelle regioni ippocampali e prefrontali, nell'ambito di task che coinvolgono la memoria di lavoro. La rete è in grado, infatti, di riprodurre risultati compatibili con i dati disponibili in letteratura, e consente inoltre di fare osservazioni che supportano alcune teorie proposte relativamente all'attività nelle regioni cerebrali sopracitate. Ulteriori sviluppi del modello potrebbero renderlo ancor più efficace ed impiegabile in più ambiti di applicazione.

Bibliografia

- [1] Alan Baddeley. *The fractionation of Working Memory*. Proc. Natl. Acad. Sci. USA Vol. 93, pp. 13468–13472, November 1996.
- [2] Mauro Cataldi, Chiara Vigliotti. *The evolving concept of the intrinsic hippocampal theta/gamma oscillator*. Front. Biosci. (Schol Ed), 2018 Jan 1;10:143-165.
- [3] O. Jensen, J. E. Lisman. *Hippocampal CA3 region predicts memory sequences: accounting for the phase precession of place cells*. Learn. Mem. 1996 3:279-287.
- [4] J. E. Lisman. *Working memory: the importance of theta and gamma oscillations*. Curr. Biol. 20, pp. 490-492.
- [5] J. E. Lisman, O. Jensen. *The theta-gamma neural code*. Neuron. 2013 March 20; 77(6): 1002–1016.
- [6] Adam Johnson, A. David Redish. *Neural Ensembles in CA3 Transiently Encode Paths Forward of the Animal at a Decision Point*. The Journal of Neuroscience, 2007 27(45):12176 –12189.
- [7] Moritz Köster, Holger Finger, Sebastian Graetz, Maren Kater, Thomas Gruber. *Theta-gamma coupling binds visual perceptual features in an associative memory task*. Sci. Rep. 2018 Dec 6;8(1):17688.
- [8] Ali Bahramisharif, Ole Jensen, Joshua Jacobs, John Lisman. *Serial representation of items during working memory maintenance at letter-selective cortical sites*. PLoS Biol. 2018 Aug 15;16(8):e2003805.
- [9] Liang-Tien Hsieh, Charan Ranganath. *Frontal midline theta oscillations during working memory maintenance and episodic encoding and retrieval*. Neuroimage. 2014 Jan 15;85 Pt 2(02):721-9.
- [10] André M Bastos, Roman Loonis, Simon Kornblith, Mikael Lundqvist, Earl K Miller. *Laminar recordings in frontal cortex suggest distinct layers for maintenance and control of working memory*. Proc. Nat. Acad. Sci. USA, 2018 Jan 30;115(5):1117-1122
- [11] Saul Sternberg. *High-Speed Scanning in Human Memory*. Science, August 1966, 153, 652-654.

- [12] Makoto Tamura, Timothy J. Spellman, Andrew M. Rosen, Joseph A. Gogos. *Hippocampal-prefrontal theta-gamma coupling during performance of a spatial working memory task*. Nat. Commun. 2017 Dec 19;8(1):2182.
- [13] Nikolai Axmacher, Melanie M. Henseler, Ole Jensen, Ilona Weinreich, Christian E. Elger, Juergen Fell. *Cross-frequency coupling supports multi-item working memory in the human hippocampus*. PNAS, February 16, 2010 vol. 107, pp. 3228–3233.
- [14] Leila Chaieb, Marcin Leszczynski, Nikolai Axmacher, Marlene Höhne, Christian E. Elger, Juergen Fell. *Theta-gamma phase-phase coupling during working memory maintenance in the human hippocampus*. Cogn. Neurosci. 2015;6(4):149-57.
- [15] Paul Sauseng, Birgit Griesmayr, Roman Freunberger, Wolfgang Klimesch. *Control mechanisms in working memory: a possible function of EEG theta oscillations*. Neurosci. Biobehav. Rev. 2010 Jun;34(7):1015-22.
- [16] Filippo Cona, Melissa Zavaglia, Mauro Ursino. *Binding and segmentation via a neural mass model trained with hebbian and anti-hebbian mechanisms*. International Journal of Neural Systems, Vol. 22, No. 2, 2012.
- [17] Filippo Cona, Mauro Ursino. *A multi-layer neural mass model for learning sequences using theta/gamma oscillations*. International Journal of Neural Systems, Vol. 23, No. 2, 2013.
- [18] Filippo Cona, Mauro Ursino. *A neural mass model of place cell activity: theta phase precession, replay and imagination of never experienced paths*. J Comput Neurosci, 2015, 38:105–127.
- [19] Mauro Ursino, Filippo Cona, Melissa Zavaglia. *The generation of rhythms within a cortical region: Analysis of a neural mass model*. NeuroImage 52, 2010, pp. 1080–1094.
- [20] Andreas K. Engel, Pascal Fries, Peter König, Michael Brecht, Wolf Singer. *Temporal Binding, Binocular Rivalry, and Consciousness*. Consciousness and Cognition 8, 128–151, 1999.
- [21] Roberto C. Sotero, Nelson J. Trujillo-Barreto, Yasser Iturria-Medina, Felix Carbonell, Juan C. Jimenez. *Realistically Coupled Neural Mass Models Can Generate EEG rhythms*. Neural Computation 19, 478–512, 2007.
- [22] Halely Balaban, Trafton Drew, Roy Luria. *Delineating resetting and updating in visual working memory based on the object-to-representation correspondence*. Neuropsychologia 113, 2018, pp. 85–94.
- [23] Halely Balaban, Roy Luria. *Neural and Behavioral Evidence for an Online Resetting Process in Visual Working Memory*. J. Neurosci. 2017 Feb 1;37(5):1225-1239.

- [24] George A. Miller. *The Magical Number Seven, Plus or Minus Two. Some Limits on Our Capacity for Processing Information*. Psychological Review, Vol. 101, No. 2, 343-352, 1956.
- [25] Nelson Cowan. *The magical number 4 in short-term memory: A reconsideration of mental storage capacity*. Behavioral and Brain Sciences 24, pp. 87–185, 2000.
- [26] Tarek K. Rajji, Reza Zomorodi, Mera S. Barr, Daniel M. Blumberger, Benoit H. Mulsant, Zafiris J. Daskalakis. *Ordering Information in Working Memory and Modulation of Gamma by Theta Oscillations in Humans*. Cereb Cortex. 2017 Feb 1;27(2):1482-1490.
- [27] Laure Verret, Edward O. Mann, Gao B. Hang, Albert M. I. Barth, Inma Cobos, Kaitlyn Ho, Nino Devidze, Eliezer Masliah, Anatol C. Kreitzer, Istvan Mody, Lennart Mucke, Jorge J. Palop. *Inhibitory interneuron deficit links altered network activity and cognitive dysfunction in Alzheimer model*. Cell. 2012 April 27; 149(3): 708–721.
- [28] Jorge J. Palop, Lennart Mucke. *Network abnormalities and interneuron dysfunction in alzheimer disease*. Neuroscience, 2016, vol. 17, pp. 777-792.
- [29] Kyerl Park, Jaedong Lee, Hyun Jae Jang, Blake A. Richards, Michael M. Kohl, Jeehyun Kwag. *Optogenetic activation of parvalbumin and somatostatin interneurons selectively restores theta-nested gamma oscillations and oscillation-induced spike timing dependent long-term potentiation impaired by amyloid β oligomers*. BMC Biology, 2020, 18:7.
- [30] Tommaso Zeppillo, Anton Schulmann, Fabio Macciardi, Brooke E. Hjelm, Melanie Föcking, P. Adolfo Sequeira, Ilaria Guella, David Cotter, William E. Bunney, Agenor Limon, Marquis P. Vawter. *Functional impairment of cortical AMPA receptors in schizophrenia*. Schizophr. Res., 2020.
- [31] Mera S. Barr, Tarek K. Rajji, Reza Zomorodi, Natasha Radhu, Tony P. George, Daniel M. Blumberger, Zafiris J. Daskalakis. *Impaired theta-gamma coupling during working memory performance in schizophrenia*. Schizophr. Res., 2017 Nov;189:104-110.
- [32] Mauro Ursino, Cristiano Cuppini, Stefano F. Cappa, Eleonora Catricalà. *A feature-based neurocomputational model of semantic memory*. Cognitive Neurodynamics 12, pp. 525–547, 2018.