APPRENDIMENTO PER RINFORZO E CODIFICA TRAMITE POPOLAZIONE NEURALE: UN MODELLO PER IL REACHING APPLICATO A DUE TASK

Dimitri Ognibene, Gianluca Baldassarre
Laboratorio di Robotica Autonoma e Vita Artificiale, Istituto di Scienze e Tecnologie della cognizione, Consiglio Nazionale delle Ricerche (LARAL-ISTC-CNR)
dimitri.ognibene@istc.cnr.it, gianluca.baldassarre@istc.cnr.it

Introduzione

Attualmente nell'ambito della modellistica per le neuroscienze si tende a costruire modelli capaci di replicare il funzionamento di parte dell'apparto neurale nel contesto di un piccolo, singolo task. In realtà una delle caratteristiche più interessanti, e attualmente irreplicabili, del nostro cervello è la capacità di affrontare una quantità elevatissima di situazioni diverse. L'approccio attuale alla modellazione neuroscientifica rischia quindi di perdere di vista alcune delle fondamentali caratteristiche e meccanismi che rendono il cervello umano così unico.

La versatilità del cervello umano è anche una delle caratteristiche di maggiore interesse per la robotica autonoma, dove tuttora si lamenta la difficoltà dei robot nell'adattarsi a compiti e situazioni diverse.

In questo articolo illustriamo il funzionamento di un modello per il reaching ispirato alle neuroscienze. Il modello è stato realizzato allo scopo di potere riprodurre un vasto insieme di esperimenti psicofisici senza essere modificato. Attualmente solo due diversi task sono stati testati: il sequence learning task (Hikosaka, Sakai et al., 2000; Ognibene, Rega et al., 2006) e il discrimination and reaching task (Cisek e Kalaska, 2005; Ognibene, Mannella et al., 2006). Altro scopo perseguito durante lo sviluppo del modello è stato quello di testare varie teorie cognitive e neuroscientifiche combinandole insieme in un'unica architettura organica e indipendente e non come pezzi a se stanti in un non ben determinato 'resto del cervello'.

Le caratteristiche di interesse del modello di reaching presentato sono:

1) Rappresentazione spaziale basata sull'ipotesi delle primitive motorie goal oriented (Giszter, Mussa-Ivaldi, Bizzi 1993; Graziano, Taylor, Moore 2002) codificate tramite una popolazione di neuroni nella corteccia celebrale;
2) Apprendimento per rinforzo (Sutton, Barto, 1995) su un insieme continuo di stati e azioni per simulare l'apprendimento in vita, riproducendo così il comportamento dei gangli della base (Houk, Adams, Barto, 1995);
3) Un meccanismo di action selection distribuito, dove tutte le possibili azioni competono per potere essere eseguite (Usher, McClelland, 2001), riproducendo così alcuni dati neuroscientifici sul processo di scelta ed esecuzione (Schall, 2001)(Cisek, Kalaska, 2005);
4) Acquisizione graduale della capacità motorie, simulando così un approccio costruttivista ossia il processo di sviluppo motorio e cognitivo a cui vanno incontro i bambini durante la crescita (Lungarella, Metta et al., 2004).

Le primitive motorie nella spina dorsale delle rane (Gistzer, Mussa-Ivaldi, Bizzi, 1993) o nella corteccia motoria dei primati (Graziano, Taylor, Moore, 2002), corrispondono a punti che se stimolati portano gli arti dell'animale in posizioni specifiche indipendentemente dalla loro posizione iniziale. Permettono così un livello di astrazione dai complessi problemi del controllo della dinamica e della cinematica degli arti.

La codifica per popolazione è un particolare tipo di codifica neurale, presente in varie zone

del cervello, dove delle grandezze di interesse vengono espresse non tramite il livello di attivazione di un singolo neurone ma dalla distribuzione dell'attivazione su un vasto gruppo di neuroni, questo modo di rappresentare le grandezze permette di risolvere i problemi riguardanti il limitato firing-rate dei neuroni biologici, di ottenere maggiore tolleranza al rumore e di avere anche maggiori capacità di generalizzazione rispetto a codifiche localistiche.

L'apprendimento per rinforzo permette ad un sistema di apprendere comportamenti anche abbastanza complessi tramite prova ed errore, senza la necessità di avere esempi di addestramento del tipo situazione-azione. In genere questi modelli utilizzano un insieme limitato di possibili azioni per problemi di efficienza, ma il modello presentato può scegliere efficacemente da un'insieme infinito di azioni. Questo è possibili grazie alla codifica finalizzata al goal delle primitive motorie e alle fasi di preaddestramento, corrispondenti allo sviluppo sensomotorio dei bambini, che consentono al sistema di sfruttare delle invarianze dell'ambiente durante la selezione dell'azione.

Nei paragrafi successivi tratterò:

1) le basi neuroscientifiche e di intelligenza artificiale su cui si basa questo lavoro;
2) I task a cui è sottoposto il sistema;
3) La struttura e le fasi di addestramento a cui è sottoposto il sistema;
4) I risultati ottenuti;
5) Considerazioni e sviluppi futuri.

Codifica per popolazione neurale

Il modo in cui sono codificate e processate le informazioni nell'architettura neurale del cervello è una delle quespotioni fondamentali della neuroscienza computazionale. Il modo in cui i dati vengono codificati nell'architettura neurale è indispensabile per capire veramente come facciamo ad agire, ad imparare ed a generalizzare le esperienze precedenti. La conoscenza della codifica neurale è fondamentale per capire le metodologie da usare nell'analisi dei dati sperimentali. Inoltre le varie funzionalità di alto livello come imparare e decidere, che sono realizzate dall'intera architettura neurale, impongono dei vincoli diversi sui meccanismi (ad esempio LTP, sincronizzazione) che devono essere presenti a livello della singola cellula in base alla diversa codifica utilizzata.

In generale esistono due casi estremi ed opposti di codifica neurale: localistica e distribuita. Nel caso di una rappresentazione localistica un singolo neurone si attiverebbe ad esempio per codificare la presenza di un particolare stimolo, come vedere il volto della nonna, mentre una codifica distribuita rappresenterebbe questo stimolo tramite un pattern globale di attivazione della rete neurale.

Ci sono evidenze sperimentali che entrambi i tipi di codifica sono presenti nel cervello. Ad esempio alcuni neuroni della corteccia temporale inferiore IT rispondono selettivamente a visi specifici (Perrett, Mistlin e Chitty, 1987; Young, Yamane, 1992). Altri neuroni rispondono ad un insieme molto vasto di stimoli e si è supposto che vi sia una codifica fortemente sparsa (Rolls e Deco,2002).

Il tipo di codifica utilizzato influisce sulla velocità dell'apprendimento ottenibile, sulla capacità di generalizzare e di rappresentare: una codifica totalmente localistica può apprendere hebbianamente ogni associazione di input-output con un singolo esempio, e il problema della separabilità lineare (Minksy e Papert, 1969) non può manifestarsi perché ogni situazione è codificata da una sola unità attivata e le altre tutte passive. Inoltre una rappresentazione localistica non subirà interferenze tra apprendimenti successivi su stati diversi. Ma una rappresentazione localistica non supporterà alcun tipo di generalizzazione ne

potrà rappresentare più situazioni di quante unità contiene la rete neurale (Foldiak, 2002).

Una rappresentazione distribuita potrà rappresentare molti più input, esattamente m^n, dove m è il numero di stati che ogni neurone può assumere ed n il numero di neuroni. Inoltre potrà generalizzare in quanto un neurone sarà attivo per più stimoli, ma non sempre la generalizzazione sarà corretta e un fenomeno di interferenza potrebbe manifestarsi. Inoltre gli algoritmi di apprendimento, per superare il problema della separabilità lineare, sono molto più lenti richiedendo un gran numero di esempi di addestramento.

La codifica semi localistica sparsa, dove solo una bassa percentuale di neuroni è attivo in corrispondenza di ogni stimolo permette di ottenere un buon compromesso tra i due estremi, permettendo di evitare interferenze ma consentendo comunque la generalizzazione e la rappresentazione di un ampio insieme di stimoli diversi. Affinché queste capacità siano verificate la codifica deve avere una struttura adeguata in modo da evitare interferenze tra i pattern di attivazioni corrispondenti a risposte diverse (Foldiak, 2002).

Nello studio della codifica neurale nel cervello bisogna comunque ricordare i limiti fisici imposti dai neuroni con frequenze di trasmissione dell'ordine delle decine di hertz e dalla presenza del rumore.

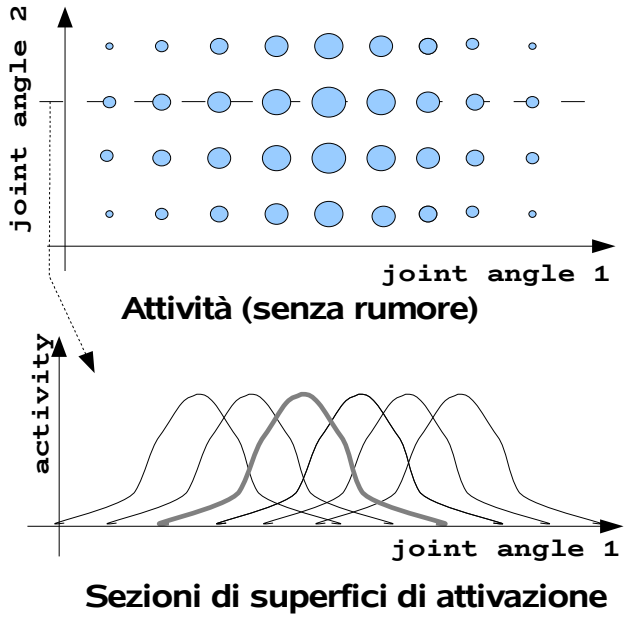

Figura 1: Codifica per popolazione utilizzata per rappresentare le posture del braccio e i dati della retina

Un particolare tipo di rappresentazione semi localistica, anche se non necessariamente sparsa, è quella per popolazione dove l'informazione viene codificata tramite l'attivazione di svariati neuroni aventi curve di attivazione simili e parzialmente sovrapposte (Pouget,Zemel e Dayan, 2000). Queste codifiche sono particolarmente robuste rispetto al rumore permettendo di definire con precisione uno o più valori su una sola popolazione, grazie alla ridondanza della codifica stessa. Inoltre permettono di apprendere facilmente complesse mappature non lineari, come ad esempio le trasformazioni dalla rappresentazione spaziale della retina a quella delle articolazioni (trasformazioni sensomotorie). La generalizzazione in una codifica per popolazione è presente ma limitata alle zone vicine allo stimolo rappresentato. Infine le popolazione possono, in alcuni casi rappresentare non solo il valore di una data variabile ma una distribuzione di probabilità (Zemel,Dayan, Pouget, 1998).

Primitive Motorie

Una ipotesi interessante suggerita dalla letteratura neuroscientifica è che il sistema motorio dei vertebrati sia organizzato sulla base di "primitive motorie". Queste primitive motorie sono poi "reclutate ed assemblate" da altre parti del sistema di controllo, come quelle deputate all'apprendimento per prove ed errori, per costruire comportamenti più complessi.

Evidenze dell'esistenza di tali primitive sono date da vari esperimenti tra cui si devono ricordare quelli di Giszter,Mussa-Ivaldi & Bizzi (1993). Questi autori mostrarono, attraverso la stimolazione elettrica di rane despinalizzate, che quando zone distinte della spina dorsale sono stimolate gli arti inferiori della rana tendono a raggiungere determinate posture indipendentemente dalla posizione in cui si trovano inizialmente.

Altri esperimenti interessanti sono quelli di Graziano, Taylor & Moore (2002) dove delle scimmie venivano sottoposte a microstimolazioni della corteccia premotoria, in dipendenza delle zone microstimolate gli arti tendevano a raggiungere delle posizioni, indipendentemente dalle posizioni iniziali, che sembravano orientate al raggiungimento di scopi specifici come portare del cibo alla bocca.

Mentre le primitive motorie nelle rane sono probabilmente di origine filogenetica, negli animali superiori, e in particolare nell'uomo alcune abilità motorie di basso livello, a cui possono ricondursi le primitive, sono acquisite nei primi anni di vita (von Hofsten, 1982).

Dal lato della robotica, l'interesse per l'ipotesi delle primitive motorie è notevole, ed è giustificato dal divario tra le capacità motorie dei robot e quelle degli animali. Questi ultimi sono capaci di un ampissimo insieme di movimenti e li adattano in maniera immediata e robusta all'ambiente e al contesto in cui si trovano.

Metodi classici di pianificazione dei movimenti, che debbano occuparsi di decidere ogni singolo passo per ogni attuatore in ogni condizione e in ogni ambiente (Nilsson, 1984), anche inaccessibile e non deterministico, non sono assolutamente in grado di raggiungere le stesse prestazioni. L'utilizzo invece di primitive motorie sia innate che apprese permetterebbe di alzare il livello di astrazione a cui la maggior parte dell'elaborazione avrà luogo.

Lo studio di architetture gerarchiche in robotica che sfruttino questo principio è presente da tempo però spesso i comportamenti di livello inferiore sono totalmente definiti dal progettista (Nilsson, 1984). Questo tende a definire a priori il livello di dettaglio a cui il sistema può arrivare, e a ridurre anche il livello di robustezza che il sistema può raggiungere. Inoltre raramente la definizione di comportamenti ha avuto ispirazione biologica.

Un aspetto interessante delle primitive motorie è che la loro creazione potrebbe avvenire tramite semplici meccanismi di clustering di ingressi sensoriali e comandi motori. Questo processo potrebbe protrarsi per l'intera vita del sistema inglobando nuovi elementi man mano che le abilità motorie si sviluppano.

Composizione di primitive motorie e Apprendimento per rinforzo

L'assemblaggio di primitive motorie per la costruzione di azioni complesse è il passo successivo da considerare. Sul fronte neurobiologico vari esperimenti indicano che un ruolo fondamentale in questa funzione è giocato dai gangli della base (Kandel, Schwartz e Jessell, 2000; Nakahara, Doya e Hikosaka, 2001). Questi nuclei subcorticali si trovano in una condizione ideale per svolgere questo compito poiché ricevono segnali dall'intera corteccia e inviano segnali alla corteccia pre-motoria e motoria.

Di grande interesse è il ruolo che i neuroni dopaminergici dei gangli della base giocano nel condizionamento classico. Durante l'apprendimento essi diventano gradualmente predittori di rinforzi primari (Shultz, Dayan e Montague, 1997).

Il processo del condizionamento classico è stato studiato con successo utilizzando il paradigma attore critico, una variante dell'apprendimento per rinforzo alle differenze

temporali TD (Barto, Sutton, Anderson, 1983; Barto, Sutton, 1998). Il modello riproduce fedelmente alcuni aspetti della fisiologia dei neuroni dopaminergici (Houk, Davis, Beister, 1995).

Lo scopo dell'apprendimento per rinforzo è di trovare, per prova ed errore, una politica d'azione che massimizzi la quantità di premi o rinforzi ricevuti dal sistema. Questo tipo di apprendimento è molto diverso da quello supervisionato in quanto non necessita di un "maestro" che indichi quale è il giusto comportamento che il sistema deve mantenere. Invece, attraverso un meccanismo di valutazione delle risposte dell'ambiente, dando un gradiente di desiderabilità ad ogni stato esperito, il sistema può apprendere quali azioni compiere in quali circostanze. Quindi rispetto all'apprendimento supervisionato il progettista necessita solo di una conoscenza del dominio limitata e così come limitati saranno i vincoli che verrano imposti al sistema. Questo potrà quindi adattarsi ad un più ampio insieme di ambienti.

Ai fini della modellazione del comportamento di animali inferiori o di funzionalità di basso livello l'apprendimento per rinforzo è maggiormente plausibile a livello biologico rispetto all'apprendimento supervisionato, in quanto la presenza di un "maestro" interno all'animale non è plausibile mentre la capacità di riconoscere i comportamenti altrui e imitarli richiede capacità cognitive superiori.

L'architettura attore critico distribuisce il compito di apprendere a due componenti:

1. Un critico che valuta il valore dello stato raggiunto basandosi sul rinforzo ottenuto in questo stato e in quelli raggiungibili da questo (si noti che non è necessario sapere quali sono gli stati accessibili dallo stato in cui il sistema si trova). In genere il segnale di rinforzo si può pensare come generato da una componente innata dell'animale che valuti come positivi alcuni stati o alcune azioni, come mangiare o bruciarsi.
2. Un attore che sceglie l'azione da compiere con una determinata politica. Ad esempio una politica di azione greedy è quella che fa scegliere all'attore l'azione che porta allo stato a più alta valutazione. L'adattamento della politica d'azione dipende da quanto viene valutato l'esito di un'azione dal critco.

Processo di scelta e reti di accumulatori a mutua inibizione

Uno dei temi più complessi e controversi nell'ambito delle neuroscienze è la distinzione tra il movimento di un arto ed un'azione e come questa differenza si manifesti a livello neurale (Schall, 2001). La distinzione tra fondamentale è che un'azione viene realizzata allo scopo di raggiungere un determinato obiettivo, scelto fra altri.

Senza scendere nel dettaglio riportiamo semplicemente che ci sono varie evidenze sperimentali che il processo di scelta tra varie azioni produca nelle zone motorie dell'architettura celebrale una preattivazione relativa alle varie possibili opzioni(Cisek e Kalaska, 2005; Riehle e Requin 1989). Solo una di queste preattivazioni darebbe poi vita all'attivazione necessaria all'esecuzione effettiva dell'azione. La scelta sarebbe fatta accumulando via via evidenze a favore delle varie opzioni. Questo permette al sistema nervoso di funzionare in un ambiente non totalmente osservabile, in cui le varie evidenze si manifestano in tempi successivi e di essere robusto rispetto al rumore.

Alcuni modelli come quelli di Usher e McClelland (2001;Bolgacz, Brown et al. 2005; Usher, Olami e McClelland, 2002), permettono di modellare il processo di scelta tramite l'utilizzo di una semplice rete di accumulatori a mutua inibizione. In questi modelli ad ogni accumulatore corrisponde una particolare scelta ed essi accumulano le evidenze a loro favore e inibiscono le altre possibili scelte. Alla fine una sola scelta, un solo accumulatore sarà attivo. Limite di questi modelli è però che l'insieme di possibili scelte deve essere predefinito. Nel modello proposto invece la scelta avviene tra un insieme infinito e acquisito durante la vita di opzioni (Ognibene, Rega e Baldassarre, 2006).

Task

- **Discrimination and Reaching** (Ognibene, Manella et al., 2006; Cisek e Kalaska, 2005) dove è stata testata la capacità del modello di imparare a integrare informazioni nel tempo e si è confrontato il comportamento del modello con quello del sistema nervoso delle scimmie sottoposte agli esperimenti.

  Questo task è stato pensato per studiare il fenomeno della decisione all'interno del sistema nervoso, seguendo l'evolversi delle attivazioni durante la progressiva acquisizione delle informazioni necessarie a compiere un'azione. L'esperimento consiste nel presentare ad una scimmia, addestrata a compiere il task, una sequenza di immagini, ognuna delle immagini contiene solo parte delle informazioni necessarie a determinare la posizione dello schermo da toccare(Figura 2, in alto). La sequenza era costituita da:

  1. segnale di attesa (pallino verde al centro);
  2. informazione spaziale: la seconda immagine mostra due posizioni contrassegnate con un colore diverso (pallino rosso in alto a destra, pallino blu in basso a sinistra blu);
  3. segnale di attesa, pallino verde al centro;
  4. informazione di colore: un pallino del colore di uno dei due target dell'immagine 2, che sarà la posizione da toccare (pallino rosso al centro);

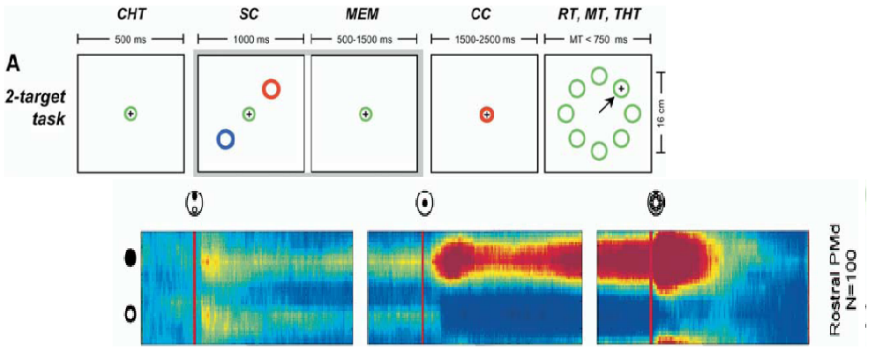

Figura 2: Sequenza di immagini del task e attivazioni della corteccia premotoria

  5. segnale di go, vari pallini verdi disposti in cerchio segnalano alla scimmia che è il momento giusto per agire.

  La scimmia riceve il rinforzo se tocca la posizione del target mostrato nell'immagine 2 che ha lo stesso colore mostrato nell'immagine 4 (in questo caso in altro a destra).

  Durante l'esperimento Cisek e Kalaska (2005) hanno misurato l'attivazione delle zone premotorie della corteccia relative ai due possibili target (Figura 2, in basso). Le letture hanno mostrato che prima della quarta immagine entrambe le zone sono attive, quando viene mostrata la quarta immagine l'attivazione della zona relativa alla posizione sbagliata veniva inibita, infine quando viene mostrata l'immagine 5 l'attivazione della zona legata alla posizione giusta raggiunge in breve il massimo livello di attivazione e la scimmia compie l'azione.

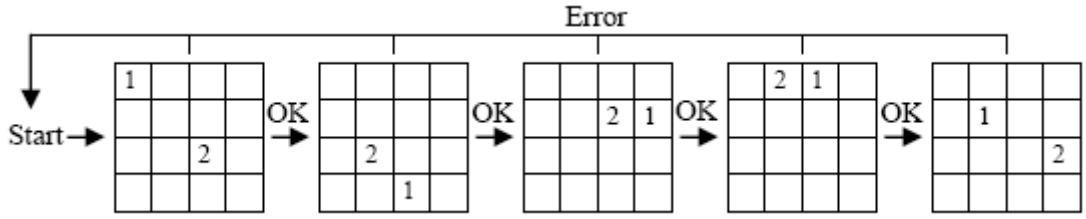

Figura 3: Sequenza appresa nel task 2, la sequenza è formata da cinque configurazioni di due bottoni che vanno selezionati nell'ordine mostrato altrimenti la sequenza ricomincia dalla prima configurazione.

- **Sequence learning**, (Ognibene, Rega et al., 2006) questo task è simile a quello utilizzato da Hikosaka e dai suoi collaboratori (Hikosaka, Sakai et al., 2000; Rand, Hikosaka et al., 1998) per studiare la fisiologia di varie aree del cervello durante l'apprendimento di alcune sequenze di reaching. In questo task la scimmia è posta davanti ad un pannello con 16 bottoni LED(Figura 3) disposti su quattro file. Sul pannello viene mostrata una sequenza di cinque configurazioni con soli due bottoni LED accesi. Per ottenere il rinforzo ad ogni configurazione la scimmia dovrà selezionare i due bottoni LED in un ordine che potrà scoprire tramite prova ed errore (indicato in figura dai numeri); se invece la scimmia dovesse sbagliare la sequenza o compiere una qualsiasi azione errata il task ricomincerà dalla prima configurazione.

Architettura e funzionamento del sistema

Il sistema di controllo senso-motorio per la coordinazione di un braccio simulato sviluppa tramite fasi successive di addestramento la sua configurazione finale, che potremmo definire la fase "adulta" del soggetto artificiale. Di seguito descriverò il funzionamento, l'architettura e i componenti del sistema in quest'ultima fase.

Due processi contemporanei hanno luogo nel sistema durante la sua fase adulta: il primo processo elabora le percezioni provenienti dai sensori e realizza in base ad esse un'azione; nel secondo processo il sistema valuta le proprie azioni, apprende quali di esse siano più appropriate nelle varie situazioni e modifica in conseguenza il proprio comportamento.

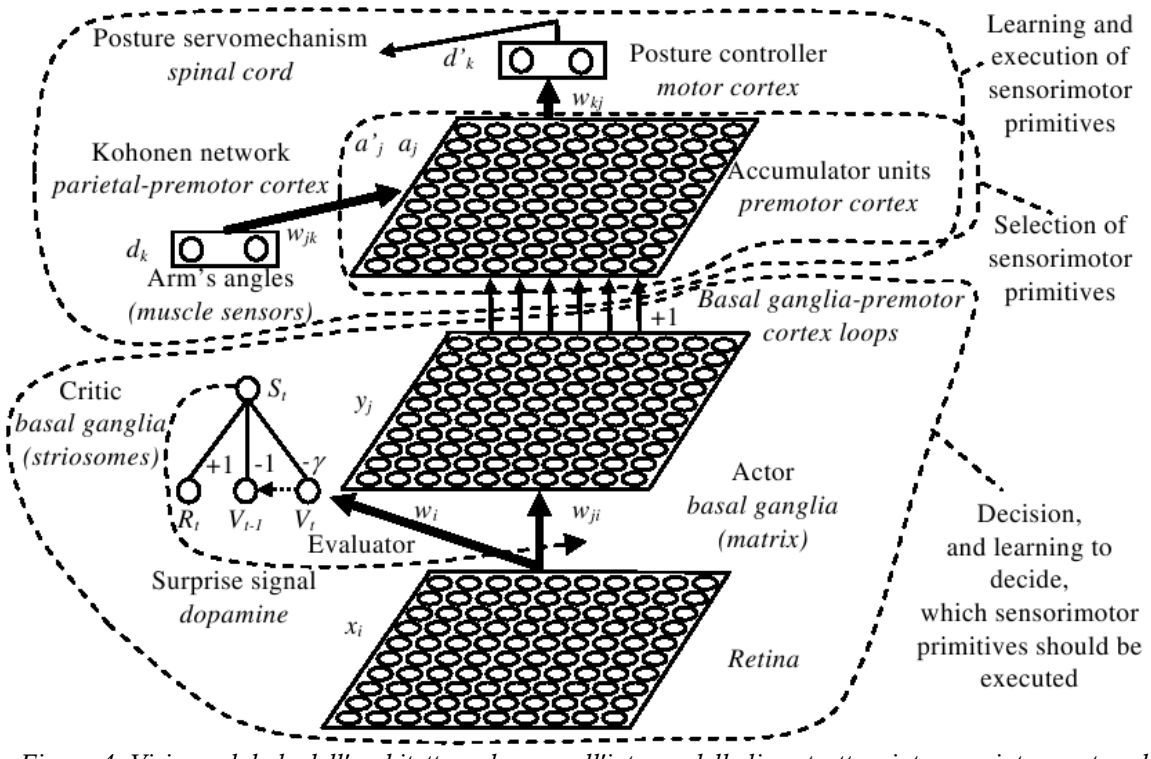

Figura 4: Visione globale dell'architettura, le zone all'interno delle linee tratteggiate sono interessate nella fase di aprrendimento indicata

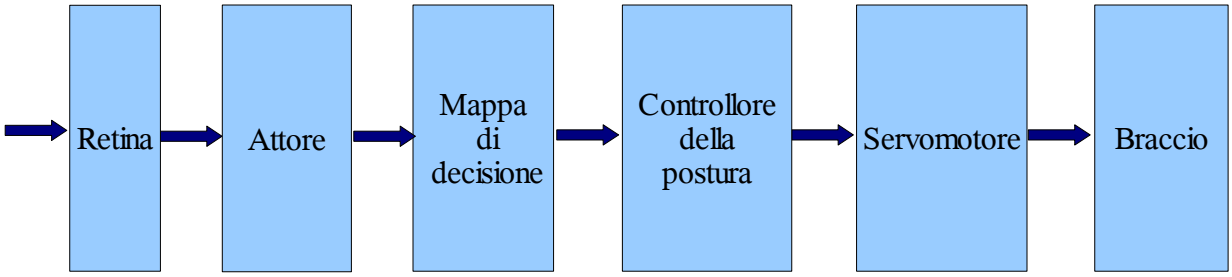

Figura 5: Processo che trasforma la percezione in azione

Il primo processo (Figura 5) parte dalla retina che trasforma l'input ottico in una rappresentazione semi-localistica. Questa rappresentazione è utilizzata dall'attore per attribuire un voto ad ogni singola primitiva motoria. I voti sono accumulati dalla mappa di decisione, che raggiunto un livello di attivazione adeguato produce un'azione codificata in maniera distribuita sulle primitive, questa rappresentazione vie utilizzata dal controllore della postura e fusione che la trasforma in un comando motorio per il servomotore che infine sposta il braccio.

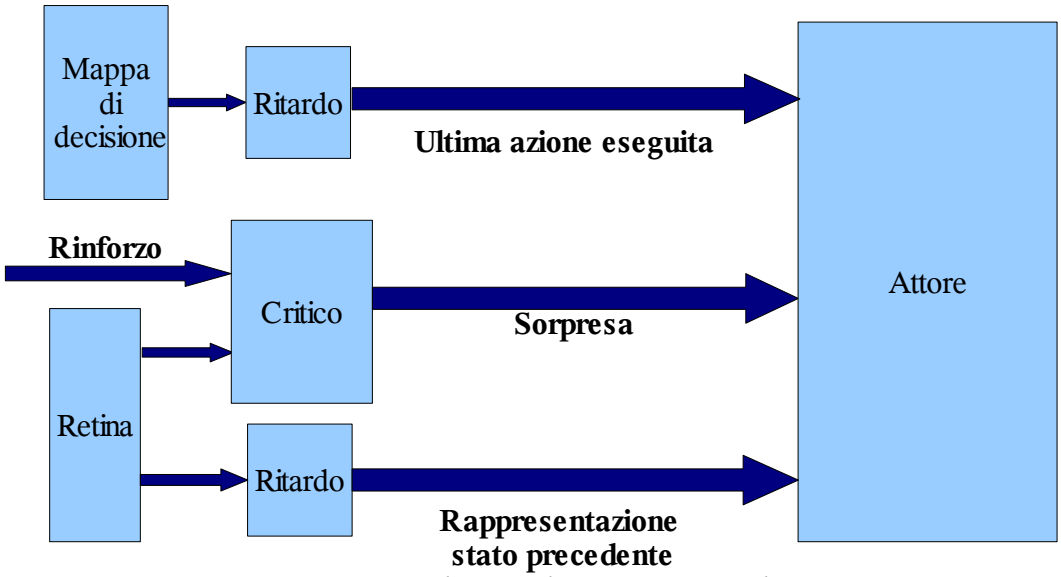

Figura 6: Processo di apprendimento per prova ed errore

Il secondo processo (Figura 6) modifica la capacità dell'attore di produrre voti come quelli che hanno portato all'ultima azione qualora l'attore nel processo 1 dovesse ricevere una rappresentazione di stato come quella ricevuta prima della precedente azione. La rappresentazione utilizzata dell'ultima azione è quella generata dalla mappa di decisione. L'attore verrà modificato in modo da generare più facilmente l'azione precedente se è positiva la sorpresa ottenuta dal critico al raggiungimento del nuovo stato, se invece la sorpresa è negativa l'azione sarà inibita.

- **Retina**
  Input visivo elaborato da un modello semplicistico di **retina,** costituito da un insieme di unità distribuite come un reticolo ed equidistanti. L'input della retina è un insieme di punti caratterizzati da posizione e colore. Ogni elemento della retina ha un area di sensibilità circolare che si sovrappone con quella degli elementi adiacenti. Nessuno stimolo interesserà solo un elemento della retina e la sua rappresentazione sarà quindi distribuita tra i suoi elementi della retina. Ogni elemento reagisce ad uno stimolo che cade dentro la sua area di sensibilità producendo una tripletta di valori corrispondenti ai tre colori fondamentali. I valori della reazione dipendono da due fattori: il primo è identico per i tre valori ed è la gaussiana della distanza dello stimolo dal centro

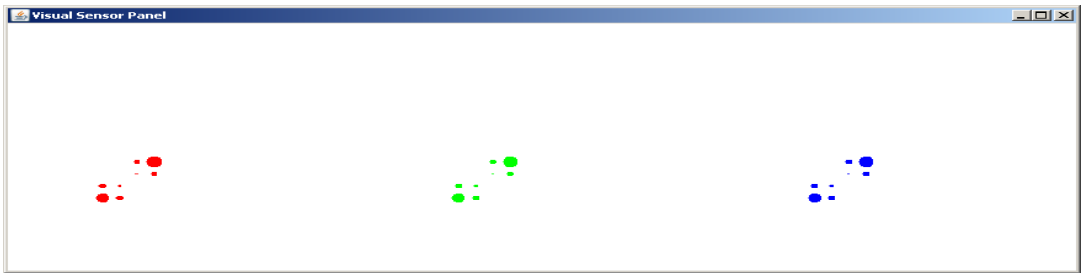

Figura 7: Output della retina, con i tre componenti di colore separati

  dell'area di sensibilità; il secondo fattore è l'intensità di uno dei tre componenti RGB dello stimolo, distinto per i tre valori della reazione. Nel caso in cui più stimoli cadano nell'area di sensibilità dell'elemento esso produrrà una risposta pari alla somma delle risposte corrispondente ai singoli stimoli.

- **Memoria Percettiva**
  Una memoria a decadimento riceve l'input dalla retina. La memoria ha la stessa struttura reticolare della retina ed ha un nodo associato ad ogni nodo della retina. L'uscita di ogni nodo è, come nella retina, un insieme di 3 valori. ed il valore è pari al valore corrispondente dell'uscita del nodo dalla retina se questo è non nullo altrimenti sarà pari al valore prodotto all'ultimo passo con un decadimento λ.

  Questo componente era presente solo nella prima versione del modello (Ognibene, Manella et al. 2006) sarà probabilmente sostituito da un'implementazione delle Liquid State Machine (Maass et al., 2002) o da un componente che possa simulare in maniera biologicamente plausibile un meccanismo di memoria. Nelle versione attuale del modello (Ognibene, Rega et al., 2006), non essendo necessaria alcuna memoria per il task scelto, essa non è stata utilizzata e l'input per i restanti componenti del sistema è fornito direttamente dall'uscita istantanea della retina.

- **Attore**
  L'attore nel processo 1, ha come input lo stato della retina, o della memoria percettiva, e genera dei voti da attribuire ad ogni primitiva motoria che vengono inviati alla mappa di decisione. Attualmente l'attore è costituito da una rete neurale feed-forward avente tanti ingressi quanti sono gli elementi della retina e tante uscite quante sono le primitive motorie.

  Grazie alla terza fase di addestramento preliminare (vedi sezione successiva) l'attore produrrà delle votazioni corrispondenti alla combinazione pesata di una azione di reaching per ogni oggetto presente sulla retina. Questo riduce lo spazio di esplorazione da quello infinito di tutte le posture assumibili a quello finito di tutti gli oggetti raggiungibili.

  Nel secondo processo (Figura 6) l'attore è addestrato tramite back-propagation, il

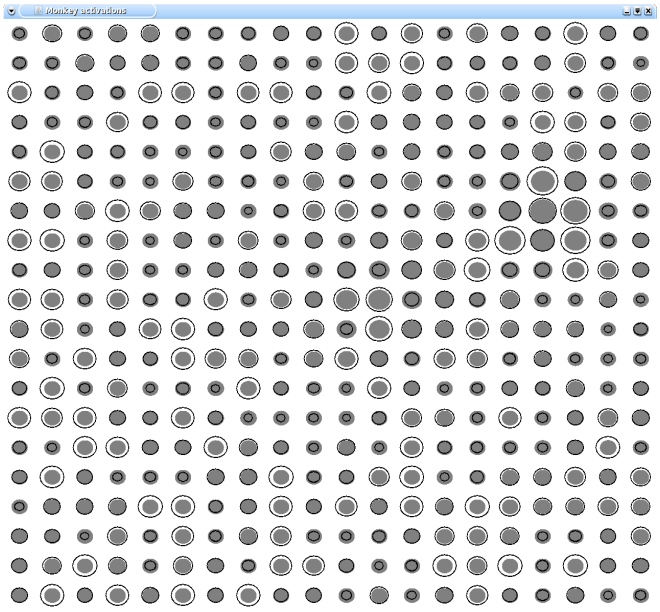

Figura 8: Voti dell'attore e rumore (si vedono 2 cluster di attivazione)

segnale utilizzato come valore desiderato (y_{ij}) è una funzione della sorpresa (s) calcolata dal critico, dell'azione selezionata dalla mappa di decisione (a_{ij}) e dei voti (v_{ij}) generati dall'attore stesso. Alcune varianti della regola sono state testate ma sempre con lo stesso principio di rinforzare le azioni che hanno avuto successo. Ad esempio la regola che ha portato i migliori risultati è stata:

$y_{ij} = v_{ij} + s * a_{ij}$

L'apprendimento dell'attore avviene solo una volta conclusa un'azione. Per azione si intende il raggiungimento della postura finale, dopo che il servomotore ha terminato il suo lavoro. Questo è uno dei punti più deboli del modello, infatti il sistema non ha modo di capire se l'azione è stata eseguita come previsto o se, ad esempio, il braccio ha incontrato ostacoli[1]

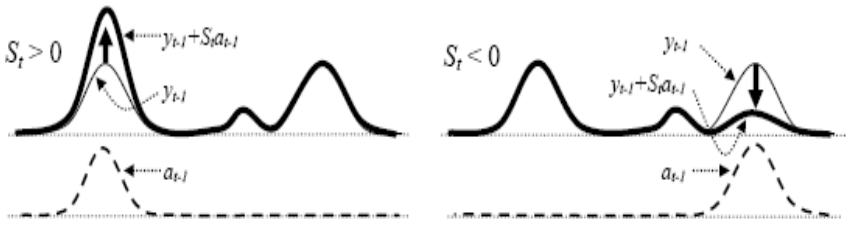

Figura 9: Addestramento dell'attore

- **Mappa di decisione**

  Il processo di decisione (Schall, 2001) è implementato da una mappa bidimensionale di unità che accumulano voti provenienti dall'attore in favore delle varie primitive motorie

  Il processo di decisione avviene con un progressivo accumulo di eccitazione negli elementi della mappa e termina quando uno di questi raggiunge il livello di saturazione, a questo punto tutte le unità inviano un segnale al livello successivo ossia il controllore della postura (Usher e McClelland, 2001).

  L'attivazione di un'unità accumulatrice al momento del rilascio, corrisponderà al

1 L'attore viene aggiornato quando il servomotore porta il braccio nella posizione desiderata.

livello di attivazione della primitiva motoria connessa all'unità stessa e codificata nel controllore della postura.

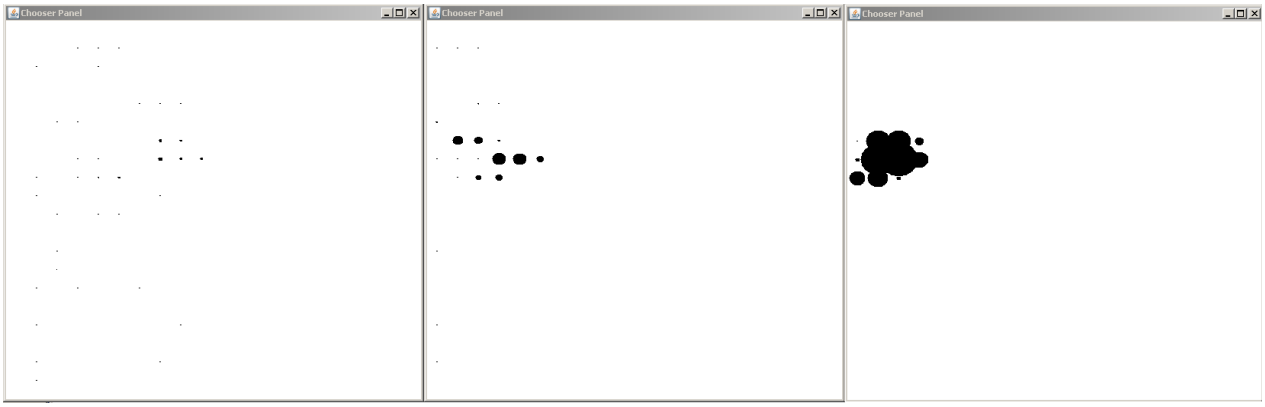

Figura 10: Step successivi del processo di decisione nella mappa a mutua inibizione

Gli elementi della mappa, sono accumulatori con dispersione e ricevono delle eccitazioni sia dall'attore che dagli elementi vicini (le eccitazioni decrescono con la distanza sulla mappa tra gli elementi). Oltre alla dispersione gli elementi subiscono anche una inibizione proveniente da tutti gli altri elementi della mappa, indipendentemente dalla loro posizione. Inoltre ogni nodo è sottoposto a due tipi di rumore, che si differenziano per la frequenza, uno ad alta frequenza che viene filtrato dalla mappa di accumulatori ed uno a bassa frequenza, che può restare praticamente invariato per lunghi periodi e quindi influisce maggiormente sull'esito delle decisioni, favorendo quindi l'esplorazione rispetto all'utilizzo della politica ritenuta ottimale in base alle esperienze precedenti.

Le inibizioni ed eccitazioni dalle unità vicine permettono di ottenere a fine corsa una forma delle attivazioni localizzata, in genere a campana. Lo scopo è quello di favorire la località delle attivazioni che rappresentano la postura desiderata, in quanto questo permette di limitare le interferenze tra azioni e situazioni simili. Si è visto inoltre che queste reti possono, con opportuni parametri, dare luogo a decisioni in tempo ottimo integrando le informazioni nel tempo (Bolgacz, Brown et al. 2005; Usher, Olami e McClelland,2002).

Nella Figura 10 viene mostrato un processo di decisione in varie fasi, l'inizio, quando l'attivazione è totalmente frutto del rumore, poi entrano in gioco i voti provenienti dall'attore, che corrispondono a due target, questi restano attivi finchè non entrano in gioco le inibizioni e le eccitazioni che fanno collassare in un unica zona le attivazioni prima del raggiungimento del livello di saturazione.

- **Controllore della postura**

  Il controllore della postura trasforma le attivazioni della mappa di decisione

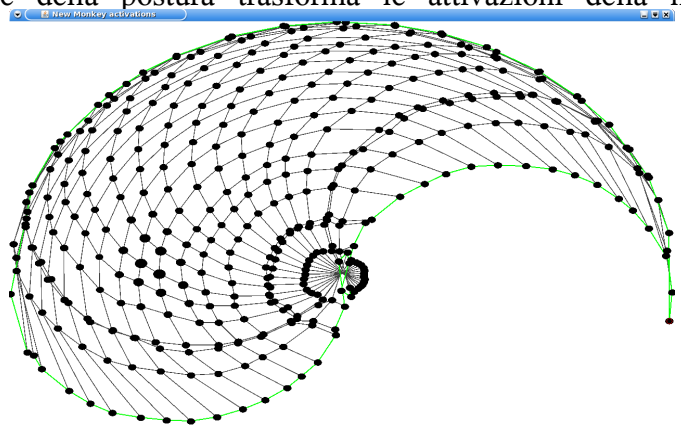

Figura 11: Posizioni generate dalle primitive motorie

nell'attivazione dei neuroni motori. Questi secondo il modello attuale codificano direttamente la postura finale che il braccio dovrà ottenere. Essa sarà la postura che permetterà di ottenere l'equilibrio tra le tensione dei muscoli agonisti e quella dei muscoli antagonisti.

Il controllore è costituito da una rete neurale feed-forward che ha per ingressi i nodi della mappa di decisione e per uscite due neuroni, uno per ogni grado di libertà del braccio.

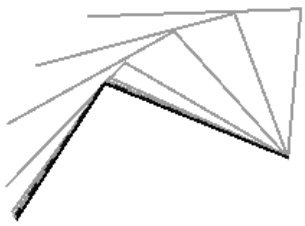

Figura 12: Spostamento in più passi operato dal servomotore

- **Servomotore**
  Non tutti gli spostamenti richiesti al braccio possono essere compiuti in un solo passo. Per raggiungere la postura desiderata potranno essere necessari un certo numero di passi, questo processo viene realizzato dal servomotore, che riceve in ingresso le posture desiderate e sposta ad ogni passo il braccio di un angolo limitato (Figura 12).

  Vari problemi sono stati trascurati nell'implementazione di questo componente e dovranno essere affrontati in futuro perché di interesse anche teorico:

  1) Notifica alle parti superiori della fine dell'azione, fondamentale nella fase di apprendimento delle primitive;
  2) Possibilità di combinare nel tempo più comandi, se il sistema resta basato sulle posture, per ottenere traiettorie di velocità arbitraria si dovranno creare successioni di posture.

- **Critico**
  L'ultimo componente è il critico che ha il ruolo di valutare l'attività dell'attore in base all'esito delle sue azioni e alle esperienze precedenti, consentendo così al sistema di risolvere compiti di apprendimento per rinforzo. Il critico valuterà l'azione attuale come positiva se permette al sistema di ottenere un rinforzo oppure se il sistema raggiunge uno stato ritenuto migliore delle posizioni raggiunte fino a quel momento a partire dallo stato in cui l'azione era iniziata (Sutton and Barto, 1998; Barto, Sutton e Anderson, 1983). Attualmente il sistema non tiene conto della durata delle azioni.(McGovern, 1998)

  Il critico è implementato come una rete neurale, seguendo il modello di Houk, Adams e Barto (1995) per la modellazione del ruolo dei gangli della base negli esperimenti di

condizionamento.

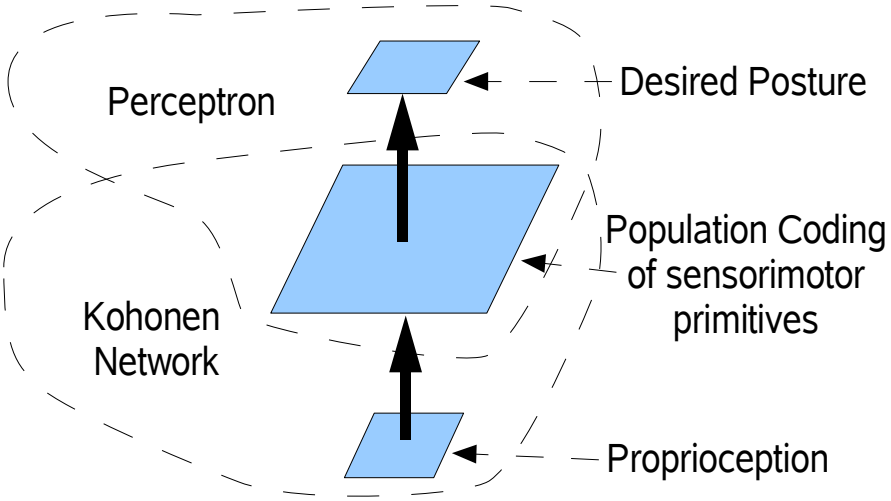

Figura 13: zone interessate dalle prime due fasi di addestramento preliminare

A differenza dell'attore il critico viene addestrato ad ogni passo dell'esperimento, dato che lo stato dell'ambiente potrà variare indipendentemente dal fatto che il modello esegua un' azione, come nel task dell'accumulazione di evidenze necessaria a decidere quale azione generare.

Fasi di addestramento preliminari

Il sistema viene sottoposto a 3 fasi preliminari di addestramento:

1) Preparazione dei nodi della mappa di decisione. Con la generazione casuale di spostamenti del braccio, gli input propriocettivi ottenuti sono clusterizzati tramite una SOM. La locazione dei nodi nella SOM è importante perché sarà la stessa che avranno sulla mappa di decisione quindi le eccitazioni e inibizioni laterali dipenderanno da come la SOM strutturerà le proprie unità(Figura 11).

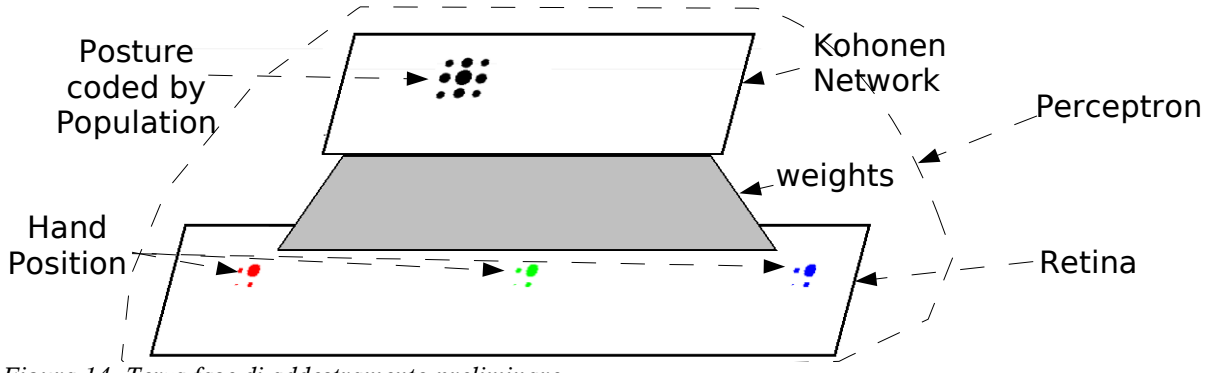

Figura 14: Terza fase di addestramento preliminare

2) Addestramento controllore della postura: sempre tramite movimenti casuali il modello impara ad associare un livello di attivazione dei neuroni motori del controllore della postura ad ogni nodo della SOM. La rete neurale del controllore di postura ha per teaching input il livello di attivazione muscolare della postura assunta e per input l'attivazione dei nodi della SOM[2].

3) Pre-addestramento dell'attore con il mapping visione-postura(Figura 14): ancora

2 Il nodo della SOM più vicino alla postura raggiunta avrà attivazione pari ad uno, gli altri nodi avranno un attivazione come funzione gaussiana della distanza dal nodo più vicino. La distanza è calcolata in termini di unità sulla mappa.

movimenti casuali permettono all'attore di apprendere che primitive motorie corrispondono allo stimolo visivo della mano. Il teaching input è l'attivazione dei nodi della SOM in corrispondenza ai valori propriocettivi. L'input è dato dall'attivazione della retina corrispondente alla posizione raggiunta dalla mano. Il sistema impara quindi un'invariante indipendente dal task che permette di ridurre in maniera sensibile il tempo di addestramento per rinforzo, sfrutterà quindi dei modelli delle proprie azioni che ha generato autonomanente. Sarebbe interessante studiare la possibilità di generalizzare questo approccio e renderlo parte del processo online di vita del sistema.

Risultati Ottenuti

1. Il modello ha appreso il task di (Figura 2)(Ognibene, Mannella et al., 2006), imparando a non generare nessuna azione durante la sequenza di immagini e a generare l'azione giusta dopo la quinta immagine. Da notare che questo non è apprendimento per rinforzo classico, in quanto l'azione di non fare nulla non è rappresentata esplicitamente, quindi il sistema non valuta ad ogni passo il valore della sua attesa, piuttosto resta in attesa di informazioni più significative.
   Varie problematiche sono emerse durante la realizzazione di questo task, ad esempio si è visto che i parametri di rumore, del coefficiente di sconto e di apprendimento sono fortemente correlati e se non vengono adeguatamente scelti potrebbe non essere possibile individuare la giusta finestra temporale in cui agire.

   Inoltre i tempi di addestramento sono molto lunghi e dovuti alla mancanza di strutturazione dei dati di input, infatti tutti i dati provenienti dalla retina sono semplicemente una sequenza di bit e il modello non ha alcuna concezione di colore o posizione. Aggiungendo delle feature di alto livello, come un neurone per ogni posizione, attivo se uno qualunque dei colori è attivo, si è avuto un miglioramento dei tempi di apprendimento di un ordine di grandezza, anche se questa configurazione non è stata provata del tutto e comunque manca di generalità e scalabilità.

   Un risultato interessante, anche se in parte atteso, è che la similitudine tra i pattern in memoria con quello finale, ottenuto dopo aver visto l'intera sequenza, non è sufficiente a generare le preattivazioni che si sono osservate nella scimmia, anzi il modello impara ad inibire del tutto l'attività della primitive motorie ossia degli accumulatori finché non vede il segnale di via.

Figura 15: errore di posizione su un insieme continuo di target

Infine si è visto che il tempo di reazione dipende fortemente dal rumore che si accumula anche nelle primitive sbagliate e deve essere "scaricato" prima che l'azione giusta possa vincere. Basare l'esplorazione su un semplice rumore, è distruttivo per le prestazioni. Un meccanismo biologicamente plausibile di esplorazione è indispensabile.

2. Sequence learning, (Ognibene, Rega et al., 2006). Il sistema ha appreso l'intero task di cinque configurazioni (Figura 3). L'apprendimento della sequenza di azioni ha mostrato empiricamente che il numero di prove necessarie cresce esponenzialmente con il numero della configurazione, ossia più una configurazione è distante dalla prima più è difficile che venga appresa. Una motivazione è che la prima volta che viene compiuta l'azione giusta solo l'ultima azione prende il rinforzo. Un'altra è che se continua a scegliere per prima la seconda azione, cioè quella sbagliata il sistema si ritrova all'inizio della sequenza dove ha una valutazione molto alta e quindi non è portato ad esplorare. Algoritmi più avanzati di esplorazione e di aggiornamento della valutazione permetterebbero di eliminare questa crescita esponenziale.

   La novità fondamentale del modello utilizzato per questo secondo task è stata la possibilità di generare un insieme infinito di azioni, mantenendo grazie alla terza fase di preaddestramento, uno bias che rende l'esplorazione dello spazio delle azioni molto veloce.

   Abbiamo verificato la capacità del sistema di generare un insieme continuo di azioni di reaching su un oggetto che si muoveva lungo una traiettoria circolare, i risultati hanno mostrato una buona precisione anche se ci sono delle aree dove la precisione diminuisce notevolmente(Figura 15).

Conclusioni sullo stato attuale del sistema

Elenco di seguito i punti di forza del sistema:

1. **Integrazione di informazione nel tempo**. Il sistema implementa un modello di scelte o decisioni (Usher e McClelland, 2001), ossia produce risposte discrete, istantanee rispetto ad un continuo temporale nell'input, non utilizzato nella gran parte dei sistemi

basati su reti neurali di mia conoscenza. Vari esperimenti dimostrano la presenza nel cervello di tali processi (Schall, 2001). Implementarle permette lo studio dei problemi di scelta e dei fenomeni di integrazione delle informazioni nel tempo che permettono ad un sistema di mantenere un comportamento performante anche quando non ha istantaneamente informazioni sufficienti per valutare lo stato del mondo ed agire di conseguenza.

2. **Uso di una rappresentazione distribuita dello spazio** basata sulle posture finali secondo la teoria delle primitive motorie di Graziano (Graziano, Taylor e Moore, 2002). Tale rappresentazione consente di ottenere migliori prestazioni dell'apprendimento per rinforzo e maggiore plausibilità biologica.
3. **Varie fasi di apprendimento**, le quali rispecchiano un approccio costruttivista allo sviluppo di capacità motorie di livello superiore sulla base di quelli più semplici. Tale approccio permette, almeno in teoria, maggiore generalità e la possibilità di applicare lo stesso modello a più tipi di attuatori.
4. **Apprendimento per rinforzo con spazio continuo delle azioni**, ottenute combinando un numero finito di primitive motorie
5. **Generazione di sole azioni significative tramite apprendimento del mapping posture-visione.** Questa capacità è stata ottenuta tramite un semplice addestramento ed è quindi totalmente estratto dai dati, con nessuno sforzo nella modellazione del dominio. Potrebbe riapplicarsi altrove, magari modellando primitive motorie che controllino più arti in maniera sinergica.
6. **Apprendimento per rinforzo dell'attesa e del momento in cui eseguire un'azione senza modellare l'azione nulla**. Questo è stato ottenuto solo in parte nel lavoro riguardante il compito di Discrimination and Reaching (Ognibene, Mannella et al. 2006). Un classico apprendimento per rinforzo dovrebbe valutare anche l'azione di non fare nulla per apprendere ad attendere il momento in cui agire. Il sistema piuttosto genera semplicemente attivazioni basse. Penso che sia interessante pensare questa capacità come un ulteriore grado di flessibilità del sistema che permette di non modellare tutte le azioni, esse emergono dal comportamento del sistema.

1 Direzioni future

Elenco qui le possibili direzioni di ricerca future, tra cui ne saranno selezionate alcune. L'elenco dei problemi seguenti non dovrà pensarsi come una semplice serie di caratteristiche che si vogliono aggiungere a un sistema piuttosto come varie facce di uno stesso complesso problema: lo sviluppo di un sistema di affrontare un numero di molti task, non solo alcuni task ad hoc.

1. **Ulteriore analisi dell'ipotesi delle primitive motorie** basate sulla postura finale. L'ipotesi della codifica di primitive motorie nella corteccia motoria è una delle più affermate in letteratura ma non l'unica, alcuni esperimenti non sono del tutto giustificabili sulla base di tali ipotesi (Aflalo & Graziano, 2005). Problematiche correlate sono la capacità espressiva di cui le primitive sono dotate, e la possibilità di apprenderle e utilizzarle in vari modi, tutte correlate alla codifica dell'azione nella rappresentazione interna, attualmente per postura finale. Altri tipi di primitive richiederebbero una codifica di variabili dinamiche e non di valori statici, e si potrebbe quindi avere un notevole aumento di complessità.
2. **Studio dell'utilità biologica delle macro-azioni.** La modularizzazione in tutte le sue forme è spesso una suddivisione astratta di un sistema operata da chi lo studia piuttosto che una proprietà intrinseca del problema stesso. Questo è anche il caso dei

sistemi biologici, che noi tendiamo a suddividere in componenti funzionali: non sempre questo è rispecchiato nella struttura di questi componenti. Tuttavia la modularità è importante: ad esempio è noto che moduli distinti possono apprendere più facilmente e con meno esempi, compiti specifici grazie alla loro struttura. Anche azioni specifiche possono favorire l'apprendimento di particolari problemi, aumentando quindi la capacità del soggetto di adattarsi (McGovern, 1998). Ma di per se la modularità introduce dei vincoli interni, dovuti alla necessità di coordinare i vari moduli, che non sempre è facile spiegare nei modelli ed estrapolare dai comportamenti degli animali (Fodor, 2000). Potrebbe essere interessante affrontare il problema dal punto di vista evolutivo, studiando quindi l'ipotesi della modularità come utile durante l'evoluzione e ottenere degli insight su caratteristiche che i moduli e l'architettura in genere dovrebbero possedere.

3. **Studio dell'utilità biologica delle decisioni**. L'assunzione che internamente al cervello avvengano delle decisioni discrete e istantanee come quelle prodotte dal nostro modello, per un sistema di forte ispirazione connessionista non è usuale. Alcune evidenze neurobiologiche di tale processo discreto di decisione sono state trovate (Schall, 2001), ma una motivazione per cui il sistema debba procedere tramite fasi distaccate di elaborazione piuttosto che come un processo continuo non sono chiare. Un'ipotesi contraria è che la percezione della “scelta” sia solo illusoria e legata all'emergere del linguaggio. Una giustificazione evolutiva per l'utilità di un meccanismo di decisione sarebbe molto interessante. Un semplice scenario che potrebbe richiedere tale tipo di processo è dovere intercettare un oggetto mobile, dove il momento in cui iniziare il movimento dipende sia dalla velocità del target che dalla propria velocità.
4. **Ripianificazione**. Un altro problema da tenere in considerazione in un sistema che fa uso di macroazioni e quindi di modularità è quanto questa modularità vincoli il sistema. Un problema che nasce dall'utilizzo di queste macroazioni di durata ed estensione elevata è ad esempio la possibilità di annullarne l'esecuzione. Un possibile task su cui confrontare il modello con un soggetto reale sarebbe il target switching, dove il soggetto è costretto ad annullare l'azione decisa in quanto le sue premesse falliscono durante l'esecuzione.
5. **Analisi della precisione del sistema di controllo nel continuo e della presenza di fenomeni di interferenza dovuti a più oggetti nell'ambiente**.

   La precisione del sistema di controllo è influenzata dalla grana della retina e quindi alla difficoltà di separare gli stimoli visivi degli oggetti (il caso di oggetti parzialmente sovrapposti richiederebbe indubbiamente un meccanismo tipo gestalt). Il problema viene ingrandito dalla distribuzione della rappresentazione, che può fare interferire oggetti vicini. Altra cosa che potrebbe influire sulla precisione durante l'apprendimento per rinforzo è il dominio in cui viene presa la decisione che viene poi ricordata come eseguita e premiata. E' diverso pensare che è stato produttivo generare una postura con un dato ingresso visivo, piuttosto che con lo stesso ingresso avere toccato un oggetto. Ci sono moltissime posture che permettono di raggiungere lo stesso oggetto, aumentando lo spazio di ricerca.

   Inoltre potrebbero esserci interferenze tra rappresentazioni sovrapposte delle posture che vengono rafforzate o inibite. Nella Figura 17 questo problema viene spiegato più chiaramente: nel modello attuale ogni postura viene rappresentata in maniera distribuita attivando diversi punti della mappa. Supponendo che durante un task di apprendimento per rinforzo il sistema veda due oggetti che preattivano la rappresentazione di un movimenti di reaching per ognuno dei due oggetti. Se i due oggetti sono troppo vicini avremo una rappresentazione distribuita di una postura (A)

che si sovrappone all'altra (B). Nel task, e solo l'azione rappresentata da A permette di ottenere il rinforzo mentre B non lo consente, quindi rappresenta una postura sbagliata. Quando il sistema impara che A è giusta e B sbagliata, quindi quando rinforza e inibisce le rispettive rappresentazioni per prova ed errore, le rappresentazioni A e B interferiscono e la postura che il sistema apprende come giusta è una versione distorta di quella rappresentata da A. Se invece il sistema apprendesse che deve prendere un oggetto piuttosto che un altro e poi generasse la postura corrispondente le interferenze non ci sarebbero.

Questo corrisponderebbe allo spostamento della decisione ad un livello diverso, dove si la decisione avviene tra oggetti generando poi un'azione specifica di reaching per il dato oggetto. Questo processo d'altronde non è banale e richiede proprietà tipiche dei sistemi simbolici piuttosto che delle reti neurali, come il binding del parametro di posizione necessario all'azione e da estrarre dalla rappresentazione dell'oggetto (Balkenius, 1994; Smolensky, 1990).

Altro problema relativo alla precisione dei movimenti ottenibili dal sistema è connesso alla regola di apprendimento per rinforzo utilizzata e alla non linearità delle uscite dell'attore (Figura 16): anche con un solo obiettivo potrebbe causare lo spostamento del target da raggiungere.

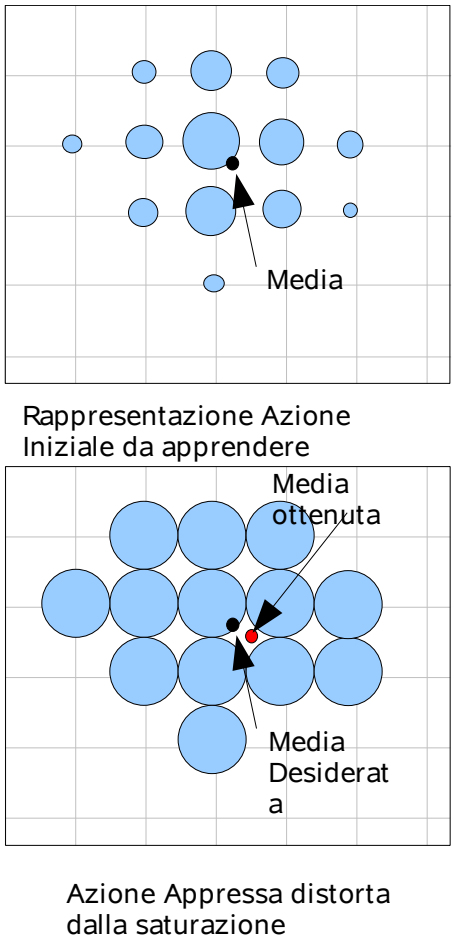

Figura 16: Distorsione dell'azione causata dalla saturazione delle unità dell'attore

La regola utilizzata per la costruzione dell'uscita desiderata dell'attore è: $y_{ij} = v_{ij} + s * a_{ij}$ (v: uscita dell'attore; s: sorpresa; a: attivazione del nodo della mappa di decisione).

Questa regola potrebbe rivelarsi semplicistica per la non linearità della funzione di attivazione scelta (sigmoidale). Le attivazioni delle primitive motorie $\{a_{ij}\}$ avranno una distribuzione dalla forma tipicamente circolare, dovuta alla terza fase di pre-addestramento, con le attivazioni vicine al centro molto attive e quelle distanti invece appena diverse da zero, con somma pesata pari al punto raggiunto dall'azione. Nel caso in cui l'azione compiuta sia valutata positivamente, questa stessa distribuzione sarà utilizzata per addestrare l'attore, che dovrebbe replicarla. In questo caso, però, data la saturazione delle unità dell'attore si potrebbe verificare uno spostamento del centro delle attivazioni e quindi una postura sbagliata. Per fortuna il critico apprende abbastanza velocemente quindi il fattore s tende velocemente a 0 e le deformazioni sono limitate. Dato che i due task a cui è stato sottoposto il modello non necessitavano di particolare precisione, non è stata indagata la capacità di queste regole di mantenere delle azioni precise. Però un processo di analisi più approfondito è necessario.

6. **Ridondanza.** Jordan e Rumhelart (1992) hanno messo in luce che il Direct Inverse Modeling (Kuperstain, 1988), procedura utilizzata per l'addestramento del controllore della postura, può soffrire del problema della convessità quando si ha ridondanza nei gradi di libertà disponibili. In pratica è necessario un modo per decidere tra tante possibili posture del braccio che portano la mano nella posizione scelta. Avendo

lavorato con un braccio con soli due giunti e avendo modellato le primitive solo tramite la posizione finale il problema della ridondanza non è stato assolutamente affrontato, ma esso si presenterà quando passeremo ad un braccio 3D. Inoltre non conta solo la ridondanza introdotta dai giunti ma anche quella introdotta dalle traiettorie e dalla dinamica che devono essere tenute in conto in ambienti dinamici, e necessarie per particolari movimenti come colpire o intercettare un oggetto (Grush,2005)

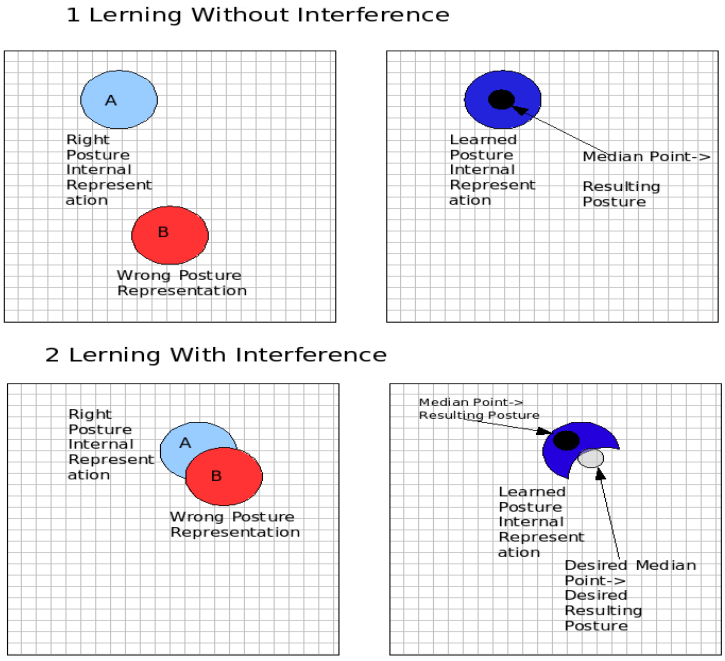

Figura 17: Interferenza tra più rappresentazioni di scelte nell'apprendimento

7. **Parziale osservabilità dello stato e modelli interni**. L'apprendimento per rinforzo generalmente lavora su problemi il cui stato è osservabile, in alcuni casi invece è necessario che il sistema impari a compiere l'azione giusta anche se non ha ad ogni istante tutte le informazioni necessarie direttamente dall'ambiente.

   E' necessario, talvolta, che il sistema tenga traccia di eventi precedenti (Ognibene, Mannella et al., 2006), dello scorrere del tempo (Ognibene, Mannella et al., 2006) e delle proprie azioni precedenti, come, ad esempio, quando deve realizzare diverse azioni che di per se non modificano le percezioni del sistema (Ognibene, Rega et al., 2006).

   La memoria percettiva utilizzata in (Ognibene, Mannella et al., 2006) non è biologicamente plausibile e non ha la capacità di rappresentare le azioni precedenti. Il modo in cui le azioni verranno ricordate non può prescindere dal modo in cui esse vengono rappresentate, e quindi dalla loro relazione con la semplice postura finale o con caratterizzazioni più complesse.

   Un cambiamento radicale sarebbe spostare il punto del modello in cui avviene la decisione, che come già detto, può prescindere dalla rappresentazione delle posture e basarsi su un livello più alto come quello degli oggetti percepiti e disambiguare le percezioni utilizzando dati estratti da esperienze precedenti, quindi si necessiterà un altro modo per ricordare le azioni fatte precedentemente.

8. **Azioni e decisioni a durata variabile**. Un problema che è stato trascurato è

l'apprendimento del modello inverso per le azioni che hanno durate diverse, non semplicemente limitate alla durata di un passo di simulazione, e anche che cosa implica il loro utilizzo in un paradigma come quello dell'apprendimento per rinforzo. Ad esempio, come fa il sistema ad imparare a compiere più velocemente un'azione? Inoltre una delle caratteristiche desiderate delle politiche di azione è che portino alla soluzione nel più breve tempo possibile, quindi non valutare la durata di un'azione (Mc Govern, 1998) e anche quello di scelta impedisce al sistema di compiere un task nel più breve tempo possibile. Questo porta anche a riflettere sul motivo per cui tale caratteristica è richiesta: in un mondo dinamico non sempre le precondizioni per raggiungere uno scopo restano invariate durante i tempi morti. Infatti se un agente perde troppo tempo per raggiungere un obiettivo quest'ultimo potrebbe anche scomparire.

9. **Integrazione della predizione**. Anche se le primitive motorie basate su posture finali si possono pensare orientate al raggiungimento di un goal e quindi di per se predittive, alcuni tipi di predizione, che sembrano avere luogo nel cervello (Cisek, 2005), non sono spiegabili con tali semplici meccanismi, come visto nell'articolo presentato ad ICCM06 (Ognibene, Manella et al., 2006). Anche la "predizione" che deriva dall'apprendimento per rinforzo non è sufficiente poiché non ha la potenza espressiva per rappresentare le variazioni dell'ambiente ma solo la possibilità di ottenere ulteriori rinforzi. La capacità di prevedere è necessaria in svariate situazioni dove viene utilizzata per generare un segnale di errore del comportamento prodotto rispetto al comportamento previsto e desiderato, oppure per interagire con oggetti in movimento o per migliorare i tempi di risposta. La predizione deve non solo prevedere lo stato successivo dell'ambiente o del robot ma dovrebbe anche potere scalare nel tempo per potere prevedere la possibilità del verificarsi di un evento. Ad esempio per prendere al volo un oggetto il sistema deve prevedere la traiettoria dell'oggetto e una traiettoria del braccio, e il fatto che vorrà prendere l'oggetto. Il problema è parecchio complesso ed è correlato sia alla modellazione delle primitive, all'apprendimento di modelli dell'ambiente da potere collegare all'apprendimento per rinforzo, alla soluzione di problemi con ridondanza, poiché ci saranno varie possibili traiettorie degli arti che permetteranno di intercettare l'oggetto, infine al problema dell'estrazione di oggetti dalla scena necessaria alla costruzione dei modelli predittivi.
10. **Apprendimento per rinforzo, tempi di decisione, sicurezza**. E' plausibile che un animale diminuisca i tempi di decisione e si comporti quindi con più sicurezza in un ambiente che conosce bene, mentre cerchi di accumulare informazioni prima di rischiare un passo falso in un ambiente che non conosce. Attualmente il modello realizzato riesce a diminuire i tempi di decisione fino ad essere limitati dal solo tempo di salita dell'accumulatore, e dal rumore che si era accumulato negli altri, perché il sistema impara qual'è l'azione giusta e il rinforzo le permette di ottenere un alto livello di attivazione. Però si è visto che il rumore che si accumula può essere notevole ma è necessario in questo modello per favorire l'esplorazione necessaria ad un sistema di apprendimento per rinforzo. Ma l'esplorazione non è necessaria in un ambiente conosciuto. Quindi sarà necessario trovare un modello più plausibile e utile del rumore per far compiere al sistema un'adeguata esplorazione dell'ambiente. Inoltre si è visto che le scimmie imparano a preattivare le azioni da eseguire prevedendo che saranno utili nelle prossime situazioni (Cisek, 2005), permettendo di avere risposte più pronte. Per ottenere questo genere di prestazioni sarà necessario integrare il modello con la predizione e con la capacità di predire che cresce gradualmente con l'esperienza dell'animale.
11. **Estrazione di features di alto livello da interazioni con l'ambiente.** Un altro problema è che l'attuale rappresentazione dell'esperienza è totalmente localistica, ossia

le proprietà che il sistema può imparare dell'ambiente sono totalmente legate alla posizione in cui egli percepisce gli stimoli: non abbiamo alcuna possibilità di generalizzare nello spazio. Le percezioni sono indistinte e scorrelate descrizioni nello spazio. Così lo sono le azioni: le primitive motorie così come sono state implementate permettono di apprendere una sola invarianza, quella della mappatura tra visione e propriocezione. La nascita di concetti, come le forme, i colori, gli oggetti, la distanza, e funzioni cognitive di più alto livello necessitano di potere generalizzare e di riconoscere nuove categorie. Non avendo quasi nessuna possibilità di generalizzare le prestazioni dell'apprendimento per rinforzo sono troppo basse. Per migliorare l'apprendimento sarebbe necessario essere capaci di estrapolare delle similitudini tra stimoli diversi. Queste possono aversi tramite la creazione di rappresentazioni con componenti comuni a diversi stimoli (Balkenius, 1996). Ma è necessario anche l'apprendimento tramite l'azione della similitudini tra gli stimoli (Balkenius,1994). Si pensi ad esempio che gli ingressi RGB della retina non permettono di per se di inferire che i colori presenti in posizioni diverse abbiano qualche relazione tra loro. Lo stesso vale per le forme. Anche se per feature di questo tipo si può pensare ad un sistema innato, per altre è necessario un apprendimento in vita. Il problema è riuscire a memorizzare l'esperienza per creare queste rappresentazioni in modo che possano essere usate per selezionare l'azione adeguata durante l'apprendimento per rinforzo.

12. **Apprendimento per rinforzo di sequenze e tempi di addestramento**, un ulteriore problema è il numero di interazioni con l'ambiente necessarie affinché il sistema apprenda il task. La velocità di adattarsi all'ambiente è un fattore evolutivo fondamentale e non riuscire a modellarlo adeguatamente renderebbe il modello molto limitato. Non avendo trovato documenti sui tempi necessari all'addestramento degli animali non è possibile paragonarli a quelli del sistema. Questi sono però troppo elevati per vari motivi, dovuti ad esempio ai problemi che ha l'apprendimento per rinforzo quando viene applicato a approssimatori di funzioni (Sutton e Barto, 1998), dalla politica di esplorazione totalmente affidata al rumore e dal meccanismo di rinforzo semplice, ossia ad ogni stato viene associato una valutazione dipendente dall'esito dell'azione e alla stato successivo passo per passo. Ad esempio si pensi ad una sequenza di due azioni, il sistema per ottenere il rinforzo deve eseguire (A,B) e otterrà il rinforzo ogni volta che esegue l'azione B. B guadagnerà per prima una valutazione superiore, alla prima esecuzione esatta del task, e verrà scelta più spesso di quanto dovuto in quanto B non avrà ottenuto alcun rinforzo. Alcuni algoritmi permettono di mantenere traccia delle ultime azioni fatte e così espandere in maniera più adeguata la valutazione con meno interazioni con l'ambiente. Questi dovrebbero essere implementati da una rete neurale che possa mantenere una rappresentazione delle ultime azioni fatte.
13. **Integrare i vari tipi di apprendimento e riprodurre gli esperimenti sull'adattamento.** Un passo necessario sarà quello di trasformare gli attuali passi di apprendimento, distinti e successivi, in un processo unico e continuo. Questo è necessario per aumentare la plausibilità biologica dell'architettura. Indubbiamente alcune capacità di apprendimento del livello più basso devono restare attive anche nelle fasi successive per permettere al sistema di adattarsi a disturbi esterni, come negli esperimenti di Polit e Bizzi (1979) e di Shadmehr e Mussa-Ivaldi (1994).
14. **Migliorare il sistema di visione per distinguere vari oggetti e soprattutto la propria mano.** Uno degli ostacoli maggiori all'integrazione dei vari tipi di apprendimento è che al modello attuale non viene fatta vedere la propria mano durante le fasi dell'apprendimento successive all'apprendimento del mapping tra visione e azione. Un meccanismo attenzionale potrebbe essere utilizzato per annullare la tendenza del sistema a generare l'azione nulla di portare la mano nella posizione in cui

si trova. Questo meccanismo potrebbe essere basato su un meccanismo di predizione necessario per collegare l'ultima azione svolta alla posizione raggiunta dalla mano. Il meccanismo attenzionale lavorerebbe quindi come un filtro che toglie rilevanza agli stimoli molto prevedibili e sarebbe quindi utilizzabile non solo per riconoscere la propria mano ma per impostare un valore di interesse indipendente dall'apprendimento per rinforzo ma legato al fattore novità (Balkenius, 2000).

15. **Maggiore corrispondenza del modello alle attuali conoscenze neurobiologiche.** Attualmente il nostro sistema non modella alcuni aspetti delle parti del sistema nervoso che sono ormai parte della letteratura neuro-scientifica, ossia:
    1. L'apprendimento tramite back-propagation è criticato come non biologicamente plausibile anche se alcuni modelli permettono di realizzare la back-propagation in maniera plausibile (Van Ooyen, e Roelfsema, 2003).
    2. Il modello dei gangli della base non tiene conto delle ultime evidenze sul processo di apprendimento (Pasupathy e Miller, 2005) e dei percorsi diretti e indiretti (Kandel, Schwartz e Jessell, 2000; Houk, Davis e Beiser, 1995; Hikosaka, Sakai et al., 2000).

Bibliografia

Aflalo, Tyson N., Graziano, Michael S.A.(2006),"Partial tuning of motor cortex neurons to final posture in a free-moving paradigm",*Proceedings of the National Academy of Sciences*,Vol.103,No.8,pp. 2909-2914

Balkenius, C. (1994). Some properties of neural representations. In Bodén, M. B., and Niklasson, L. F. (Eds.), *Connectionism in a Broad Perspective,* pp. 79-88, New York, Ellis Horwood.

Balkenius, C. (1996). Generalization in instrumental learning. In Maes, P., Mataric, M., Meyer, J.-A. , Pollack, J., and Wilson, S. W. (Eds.), *From Animals to Animats 4: Proceedings of the Fourth International Conference on Simulation of Adaptive Behavior*. Cambridge, MA, MIT Press/Bradford Books.

Balkenius C. (2000), Attention, habituation and conditioning: toward a computational model, *Cognitive Science Quarterly, 1, 2*, pp. 171-214.

Bernstein N. (1967), *The Coordination and Regulation of Movements*, Oxford, Pergamon Press.

Barto A.,Sutton R., Anderson C. (1983), Neuron-like Elements that can Solve Difficult Learning Control Problems, *IEEE Trans. on Systems, Man and Cybernetics*, vol. 13, pp. 835-846.

Bogacz R., Brown E., Moehlis J., Holmes, P., Cohen J.D. (2005), How a biological decision network can implement a statistically optimal test,*An International Workshop on Modelling Natural Action Selection*, pp 3-8, Edinburgh,AISB Press.

Cisek P., Kalaska J.F.(2005), Neural correlates of reaching decisions in dorsal premotor cortex: specification of multiple direction choices and final selection of action, *Neuron.*, 45(5), pp. 801-814.

Fodor, J. (2000) The Mind Doesn?t Work That Way: The Scope and Limits of Computational Psychology. Cambridge Mass. : MIT Press.

Foldiak P. (2002), Sparse coding in the primate cortex, *The Handbook of Brain Theory and Neural Networks, Second Edition*, pp. 1064 - 1068, ed. Michael A. Arbib, MIT Press,

Giszter S.F., Mussa-Ivaldi F.A., Bizzi E. (1993), Convergent force fields organized in the frog's spinal cord, *Journal of Neuroscience*, Vol 13, pp. 467-491.

Graziano M. S., Taylor C. S., Moore T. (2002). Complex movements evoked by microstimulation of precentral cortex. *Neuron, 34*, pp. 841-851.

Grush, Rick (2005).”Internal models and the construction of time: generalizing from state estimation to trajectory estimation to address temporal features of perception, including temporal illusions”, J. Neural Eng. 2 S209-S218

Gurney, K., Prescott, T. J., Redgrave, P.(2001) “A Computational Model of Action Selection in the Basal Ganglia”, A New Functional Anatomy. Biological Cybernetics, 84,401-410

Hikosaka O., Sakai K., Nakahara H., Lu X., Miyachi S., Nakamura K., Rand M. K.(2000), Neural Mechanisms for Learning of Sequential Procedures. In Gazzaniga M.S. (ed.): *The New Cognitive Neurosciences*. pp. 553--572 ,Cambridge MA,MIT Press.

Houk J. C., Adams J. L., Barto A. G. (1995), A model of how the basal ganglia generate and use neural signals that predict reinforcement In J. C. Houk, J. L. Davis, D. G. Beiser (Eds.), *Models of information processing in the basal ganglia,* pp. 249-270, Cambridge, MA, MIT Press.

Houk, J.C., Davis, J.L., Beiser, D.G. (eds.)(1995): Models of Information Processing in the Basal Ganglia. MIT Press, Cambridge MA

Jordan, M., Rumelhart, D.: Forward models: supervised learning with a distal teacher. Cognitive Science 16 (1992) 307-354

Kandel, E. R., Schwartz,J.H.,Jessell,T. M. (2000)”Principles of Neural Science”Mc Graw Hill, 2000.

Kuperstein, M.: A Neural Model of Adaptive Hand-Eye Coordination for Single Postures. Science 239 (1988) 1308-1311

Lungarella M., Metta G., Pfeifer R. , Sandini G. (2004), Developmental Robotics: A

Survey. *Connection Science*. Vol. 15 Issue 4, pp. 151-190.
Maass, W.; Natschläger, T. & Markram, H.(2002) “A Model for Real-Time Computation in Generic Neural Microcircuits”, Proc. of NIPS 2002, MIT Press, 15, 229-236

McGovern, Amy , Sutton, Richard S. (1998)”Macro-Actions in Reinforcement Learning: An Empirical Analysis”, Master's thesis and University of Massachusetts, Amherst Technical Report 98-70

Minsky Marvin L., Papert Seymour A. (1969) *Perceptrons: An Introduction to Computational Geometry*. MIT Press, Cambridge, MA.

Nakahara H., Doya K.,Hikosaka O.(2001),Parallel Cortico-Basal Ganglia Mechanisms for Acquisition and Execution of Visuomotor Sequences—A Computational Approach,*Journal of Cognitive Neuroscience*,13:5, 626-647.

Nilsson, N.J.(1984), *Shakey the robot*, Nota tecninca 323, SRI International, Menlo Park, California.

Ognibene D., Mannella F., Pezzullo G., Baldassarre G. (2006), Integrating Reinforcement-Learning, Accumulator Models, and Motor-Primitives to Study Action Selection and Reaching in Monkeys,*Proceedings of the Seventh International Congerence on Cognitive Modeling*(ICCM06)

Ognibene D., Rega A., Baldassarre G. (2006), A Model of Reaching That Integrates Reinforcement Learning and Population Encoding of Postures,accepted for publication in the *Proceedings of The Ninth International Conference on the Simulation of Adaptive Behavior* (SAB06)

Pasupathy, A., Miller E.K.(2005), Different Time Courses of Learning-Related Activity In the

Prefrontal Cortex and Striatum,*Nature* 433, pp. 873-876

Perrett, D. I., Mistlin, A. J., Chitty, A. J. (1987), Visual cells responsive to faces. *Trends in Neuroscience* , 10, pp. 358-364.

Polit, A., Bizzi, E. (1979),"Characteristics of motor programs underlying arm movements in monkeys", Journal of Neurophysiology,42:183-194

Pouget A., Dayan P., Zemel R. (2000), Information processing with population codes, *Nat Rev Neurosci*, 1, pp. 125-132

Rand M.K., Hikosaka O., Miyachi S., Lu X., Miyashita K.(1998) Characteristics of a Long-Term Procedural Skill in the Monkey, *Exp Brain Res,* 118, 293-297

Riehle A.,Requin J. (1989), Monkey primary motor and premotor cortex: single-cell activity related to prior information about direction and extent of an intended movement, *J.Neurophysiol,*61, pp. 534–54

Rolls E. T., Deco G. (2002), *Computational Neuroscience of Vision,* Oxford, Oxford University Press

Sanger Terence David (2000) Human Arm Movements Described by a Low-Dimensional Superposition of Principal Components,*J. Neurosci*. 20,pp. 1066-1072

Schall J.(2001),Neural basis of deciding, choosing and acting, *Nature Reviews Neuroscience*, 2, pp. 33-42

Schultz W., Dayan P., Montague P. R. (1997),A neural substrate of prediction and reward, *Science*, 275, pp. 1593-1599.

Shadmehr R., Wise S.P.(2005),*The Computational Neurobiology of Reaching and Pointing, A Foundation for Motor Learning*,Mit Press.

Shadmher R., Mussa-Ivaldi F. A.(1994)Adaptative representation of dynamics during learning of motor task,*Journal of Neuroscience*,5(14),3208-3224

Smolensky P. (1990),Tensor product variable binding and the representation of symbolic structures in connectionist systems, *Artificial Intelligence*, 46, 159-216.

Sutton Richard S., Barto Andrew G. (1998), *Reinforcement Learning: An Introduction*, Cambridge, MIT Press

Usher M., McClelland J.L. (2001), The time course of perceptual choice: the leaky, competing accumulator model, *Psychological Review*, 108, pp.550-592

Usher, M., Olav, Z. & McClelland, J.L. (2002), Hick's Law in a Stochastic Race Model with Speed-Accuracy Tradoff, *Journal of Mathematical Psycology,46*, pp. 704-175

van Ooyen A., Roelfsema P.R.(2003) A Biologically Plausible Implementation of Error-Backpropagation for Classification Tasks, *Artificial Neural Networks and Neural Information Processing Supplementary Proceedings ICANN/ICONIP 2003*

von Hofsten C. (1982), Eye-hand coordination in newborns,*Developmental Psycology*,18,pp 450-461.

Young. M. P., Yamane, S. (1992), Sparse population coding of faces in the inferotemporal cortex. *Science* , 256, pp.1327-1331.

Zemel R.S., Dayan P., Pouget, A. (1998), Probabilistic interpretation of population codes, *Neural Comput*, 10, pp. 403-430