

Studio ed implementazione robotica di un
modello neurofisiologico dei movimenti oculari
di inseguimento basato su predizione e
apprendimento

Zambrano Davide

Indice

1	Sistemi percettivi di tipo anticipativo	3
1.1	Introduzione	3
1.2	Architetture dei sistemi robotici	4
1.3	Modelli interni di tipo anticipativo	8
1.4	Predizione nel movimento oculare di inseguimento	12
1.5	Sviluppo della capacità di inseguimento di target in movimento	14
2	Neurofisiologia del sistema visivo	16
2.1	Introduzione	16
2.2	Analisi delle informazioni visive nella retina	17
2.3	Le vie visive centrali	21
2.4	La percezione del movimento	25
2.5	Apprendimento visivo	28
2.6	Controllo dello sguardo	32
3	Modello proposto di movimento oculare di inseguimento	35
3.1	Introduzione	35
3.2	Modello di inseguimento classico	36
3.3	Modello di Schaal e Shibata	38
3.4	Simulazione in Matlab Simulink	41
3.5	Miglioramento del modello di Schaal	48
4	Implementazione sulla piattaforma robotica iCub	57
4.1	Introduzione	57
4.2	Piattaforma iCub	58
4.3	YARP e l'ambiente di simulazione	61
4.4	Scelte implementative	63
5	Risultati e conclusioni	65
5.1	Risultati	65
5.2	Conclusioni	70

Introduzione

Il significato che attribuiamo al senso della visione è riscontrabile nell'importanza che occupa nelle attività quotidiane. Poter osservare quello che ci circonda permette l'interazione con il mondo, distinguendone e riconoscendone oggetti, persone, distanze e profondità. Tra le proprietà che tale sistema possiede c'è la capacità di riconoscere il movimento di tutto ciò che è all'interno nel nostro campo visivo. Occorre tenere ben presente che questa capacità entra in funzione anche quando siamo noi stessi a muoverci rispetto al mondo. Per esempio, camminare all'interno di una stanza è un'operazione che svolgiamo con naturalezza, ma che in realtà richiede l'attività di diverse aree cerebrali, le quali, tutt'oggi, sono oggetto di numerosi studi per comprendere integralmente la loro funzione. Il movimento d'altronde è legato strettamente al tempo, la grandezza fisica che scandisce le nostre giornate. Riconoscere che un oggetto si muove più velocemente di un altro, significa dire che compie più spazio a parità di tempo. Vedere il movimento, dunque, significa essere presenti nel tempo reale, significa poter interagire con il mondo. Ma elaborare i segnali che provengono dagli occhi implica l'impiego di un determinato tempo, seppur breve, per poter eseguire le operazioni di distinzione e riconoscimento citate. E questo significa essere sempre in ritardo rispetto al mondo. E' lecito dunque domandarsi come sia possibile seguire perfettamente un oggetto che si muove nel nostro campo visivo. La risposta risiede in un'altra capacità che il cervello possiede legata alla grandezza tempo, ovvero, la capacità di prevedere il comportamento ed i cambiamenti dell'ambiente esterno.

Utilizzando le tecniche d'indagine tipiche della bioingegneria questo progetto di tesi ha lo scopo di comprendere la natura della capacità di riconoscere e seguire correttamente un oggetto che si muove all'interno del campo visivo. Lo studio approfondito della neurofisiologia di questo sistema, permette lo sviluppo di un modello matematico che diviene sia spiegazione del comportamento biologico e sia utile strumento per lo sviluppo tecnologico dei sistemi percettivi artificiali.

Capitolo 1

Sistemi percettivi di tipo anticipativo

1.1 Introduzione

Questo progetto di tesi si inserisce nel campo della neuro-robotica. Il termine si riferisce ad una parte della robotica che applica il metodo di indagine scientifica tipico della bioingegneria al settore delle neuroscienze [1]. Gli studi provenienti dal settore neuroscientifico vengono, in una prima fase, usati per l'analisi, la caratterizzazione e la modellizzazione del sistema biologico di riferimento o di alcune sue componenti usando strumenti e metodi tipici della bioingegneria. Successivamente, in base al modello ottenuto si progettano artefatti intelligenti che permettono, da un lato, la validazione sperimentale delle componenti, delle funzioni o del sistema modellato, e quindi in definitiva al miglioramento delle conoscenze di tale sistema biologico e dall'altro di ottenere strumenti utili per lo sviluppo tecnologico ispirandosi alle soluzioni adottate in natura. Durante la progettazione di un sistema robotico, infatti, si giunge spesso a problemi dovuti ai limiti fisici imposti dall'hardware utilizzato come la qualità dei segnali provenienti dal sistema di acquisizione, la velocità di elaborazione dei dati o il ritardo delle risposte, il cui superamento necessita di soluzioni ingegneristiche sempre più avanzate. La bioingegneria insegna come lo studio e l'osservazione dei sistemi biologici possa essere utile per ottenere nuove tecnologie. Infatti, la teoria di Charles Darwin insegna che l'evoluzione è uno strumento molto efficiente che ha permesso lo sviluppo di organismi biologici complessi in grado di adattarsi alle variazioni dell'ambiente al fine della sopravvivenza delle specie. Volendo entrare nell'argomento di questo lavoro, un esempio è dato dal sistema visivo che nel corso di milioni di anni di evoluzione è diventato molto complesso in parte per l'anatomia degli organi coinvolti, ma soprattutto per le prestazioni che tale sistema riesce ad ottenere e per i compiti che è in grado di svolgere. L'analogia tra l'occhio umano e una normale fotocamera è, infatti, divenuta un esempio classico di come la natura riesca, con un hardware rela-

tivamente semplice, ad eseguire compiti straordinariamente complessi quale il riconoscimento di un volto o di oggetti parzialmente coperti oppure di adattarsi rapidamente alle variazioni di condizioni ambientali, come l'intensità luminosa, mentre le nostre fotocamere seppure estremamente complesse dal punto di vista tecnologico, non riescono a imitare tali comportamenti. Per studiare dispositivi artificiali con queste caratteristiche è indispensabile comprendere a fondo le soluzioni biologiche che sono alla base delle prestazioni degli esseri umani ed inoltre, costruire sistemi umanoidi per trovare soluzioni confrontabili e basate su tecnologie realizzabili [2].

1.2 Architetture dei sistemi robotici

L'obiettivo di riprodurre sistemi artificiali che simulino alcune delle funzionalità tipiche degli esseri umani ha spinto gli scienziati a proporre metodi di progettazione robotica che superino la classica visione del robot industriale. Quest'ultima definisce il robot come un manipolatore multifunzionale riprogrammabile progettato per svolgere un compito ripetitivo al posto dell'uomo, in un ambiente di lavoro strutturato e noto a priori, utilizzato per rendere la produzione più accurata ed efficiente. Sulla base di questo, infatti, si sono formate le prime idee sull'intelligenza artificiale, nella quale si ipotizzava di riprodurre la cognizione umana attraverso una computazione definita su rappresentazioni simboliche, in un processo nel quale l'informazione sul mondo è astratta dalla percezione e viene rappresentata utilizzando alcuni insiemi di simboli appropriati, sulla quale si compiono dei ragionamenti, e alla fine utilizzata per pianificare come agire nel mondo. Tale approccio è generalmente detto *cognitivista*. Il tipico sistema robotico può essere rappresentato da tre blocchi funzionali: SENSE, PLAN, ACT. Nella categoria SENSE si trovano tutti i moduli che accettano input dal mondo esterno, i sensori che forniscono le informazioni utili agli altri blocchi; la categoria PLAN invece si occupa di elaborare le informazioni provenienti dai moduli SENSE e restituisce un elenco di azioni che il sistema deve eseguire; la categoria ACT è composta dai moduli che traducono le istruzioni provenienti dal modulo PLAN in una sequenza di azioni da compiere, quindi di comandi per gli attuatori. Murphy [3] definisce come Paradigma Gerarchico l'architettura nella quale le relazioni tra i moduli vanno da SENSE a PLAN e da PLAN a ACT (figura 1.1). Il robot così costituito dovrà essere accuratamente programmato in modo da produrre azioni adeguate in risposta ad input specifici. Questo è proprio il limite al quale è giunta la visione dell'intelligenza artificiale classica, in quanto il sistema non è in grado di reagire correttamente se sottoposto ad input non previsti, quali quelli provenienti da un ambiente non strutturato. Infatti per ottenere una buona rappresentazione simbolica dell'ambiente esterno occorre un'alta complessità computazionale e questo rallenta necessariamente i tempi di elaborazione e di azione. L'approccio cognitivista e la sua tipica rappresentazione nel Paradigma Gerarchico, risulta molto efficiente in ambienti ben noti e definiti e nello svolgimento di compiti ripetitivi comuni nelle applicazioni industriali.

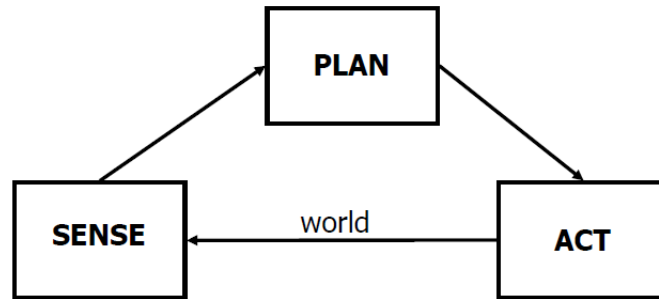


Figura 1.1: Paradigma Gerarchico: dal blocco SENSE è possibile ricavare una rappresentazione del mondo, il blocco PLAN pianifica la sequenza di atti motori da inviare al blocco ACT che riceve le direttive ed agisce sul mondo.

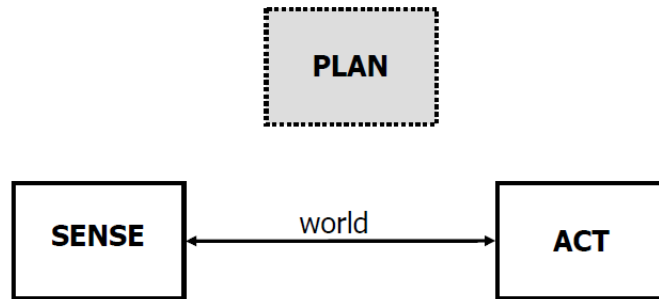


Figura 1.2: Paradigma Reattivo: la connessione dei blocchi SENSE ed ACT è diretta, il sistema risulta essere più veloce in quanto si elide la parte di pianificazione.

L'osservazione dei sistemi biologici, mostra come gli esseri viventi si siano adattati a vivere in condizioni ambientali variabili, siano in grado di gestire molti compiti differenti in modo autonomo e parallelo e possano imparare dall'esperienza. Queste sono le caratteristiche più ambite per un sistema robotico autonomo capace di pianificare dinamicamente i comportamenti in un ambiente di lavoro non noto a priori e variabile nel tempo. Per questo, negli ultimi anni, gli sforzi dei ricercatori è stato rivolto verso architetture differenti, ispirandosi anche ad organismi più semplici in grado di esibire comportamenti complessi senza una rappresentazione simbolica accurata dell'ambiente esterno e senza alcun sistema di controllo centralizzato. L'organizzazione semplificata conduce a paradigmi di controllo di tipo Reattivo in cui sono presenti esclusivamente i moduli SENSE e ACT (figura 1.2). In questo modo, la connessione di moduli di percezione ed azione è in grado di produrre comportamenti emergenti, anche complessi, senza una rappresentazione del mondo. Nelle architetture di control-

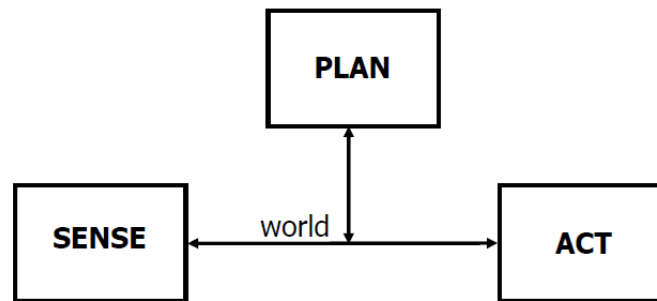


Figura 1.3: Paradigma Ibrido: permette di unire la velocità di risposta dei sistemi reattivi e la possibilità di intervento pianificato tipico delle architetture gerarchiche.

lo reattive l'esecuzione di un compito è suddivisa tra moduli ognuno dei quali ha assegnata una specifica competenza ed attua un determinato comportamento; il comportamento complessivo è determinato dall'insieme dei comportamenti presenti. Si ottiene una struttura semplice, veloce ed efficiente, ma anche imprevedibile e che comporta un'accurata gestione delle priorità dei vari moduli. L'agente mostra un comportamento che non deriva da un'attività di ragionamento e pianificazione, ma è frutto di una rete più o meno articolata di archi riflessi, di semplici collegamenti senso-motori: l'agente reagisce solo a ciò che percepisce in quel momento. In pratica si ottiene un sistema sicuramente meno complesso dal punto di vista computazionale e capace di risposte rapide, ma legato alle caratteristiche locali dell'ambiente e quindi incapace di trovare soluzioni a problemi per cui è necessario ricorrere alla conoscenza di informazioni memorizzate sull'ambiente, che riguardano proprietà dell'ambiente non correntemente percepite dai sensori. Occorre quindi ipotizzare architetture di tipo Ibrido che siano in grado di unire sia la reattività delle risposte dei sistemi Reattivi sia la pianificazione di comportamenti tipici dell'architettura Gerarchica (figura 1.3).

L'interesse dell'Ingegneria per le scienze che studiano la vita ha avuto una forte spinta a partire dagli anni Ottanta, con l'avvento delle reti neurali artificiali capaci da un lato, di simulare il comportamento dei neuroni biologici e dall'altro di fornire tecniche avanzate di elaborazione dati applicabili in molti settori, anche nella biologia stessa. Questo duplice scambio di conoscenze giunge a completa espansione nel campo delle Neuroscienze nel quale confluiscono anche altre discipline come Medicina, Psicologia, Filosofia. Nella sfida di comprendere il funzionamento del cervello umano è risultato chiaro che solo un continuo flusso di informazioni da tutti i settori scientifici è in grado di raggiungere lo scopo. In quest'ottica i comportamenti intelligenti superiori tipici dell'essere umano, che la robotica vorrebbe riprodurre, devono tener conto che l'uomo è il frutto dalla storia evolutiva della sua specie, *filogenesi*, ed al contempo, del proprio sviluppo come organismo singolo, come individuo, *ontogenesi*. Il nuovo approccio per lo studio della cognizione può, così, essere definito *emergente*. Nei sistemi emergen-

ti la cognizione è un processo di auto-organizzazione tramite il quale il sistema si ricostruisce e modifica continuamente in tempo reale attraverso interazioni reciproche sistema-ambiente. L'agente è in grado di costruirsi la propria realtà (il proprio mondo) come risultato del suo agire in quel mondo. Neuroscienze e scienze cognitive, infatti, concordano sull'importanza dell'attività motoria e sul suo ruolo sulla percezione. E' di esempio la scoperta nella corteccia premotoria di neuroni la cui attività è modulata dall'informazione visiva [4]. L'acquisizione dell'informazione sensoriale pertanto è quasi sempre un processo attivo. Nell'uomo e in tutti gli animali dotati di fovea, gli occhi sono in continuo movimento ed esplorano l'ambiente per acquisire l'informazione visiva in maniera efficiente. Altre modalità sensoriali, quali ad esempio il tatto, richiedono vere e proprie strategie di esplorazione che coinvolgono l'interazione tra corpo e ambiente. L'attività motoria ha pertanto un ruolo fondamentale per il corretto sviluppo percettivo dell'uomo. Tali considerazioni conducono alla visione di un paradigma emergente dotato di un punto di vista innovativo rispetto all'approccio cognitivista tipico dell'Architettura Gerarchica vista in precedenza (figura 1.1). Le differenze si possono ricercare in diversi aspetti, di seguito ne verranno mostrati alcuni al fine di evidenziare come le due tipologie di approccio allo studio della cognizione risultino essere completamente opposti.

- **Restrizioni temporali:** i sistemi cognitivisti non sono necessariamente sincroni con gli eventi del mondo esterno. Quelli emergenti sono invece legati al mondo esterno e operano in modo sincronico, in tempo reale, con gli eventi che accadono nel loro ambiente circostante.
- **Embodiment:** in linea di massima, i sistemi cognitivisti non hanno bisogno di possedere un corpo fisico in virtù delle proprie radici nel funzionalismo (che asserisce che la cognizione è indipendente dalla piattaforma fisica nella quale viene implementata). D'altra parte, i sistemi emergenti sono intrinsecamente legati all'apparato fisico che gestiscono e l'istanziamento fisico gioca un ruolo costitutivo diretto nel processo cognitivo ([5, 6, 7]).
- **Percezione:** Nei sistemi cognitivisti la percezione fornisce un'interfaccia tra il mondo esterno e la rappresentazione simbolica di quel mondo. La percezione astrae rappresentazioni spazio-temporali fedeli del mondo esterno a partire dai dati sensoriali. Nei sistemi emergenti la percezione è un cambiamento nello stato del sistema in risposta alle perturbazioni dell'ambiente esterno, al fine di mantenere la stabilità.
- **Azione.** Nei sistemi cognitivisti le azioni sono conseguenze causali dell'elaborazione simbolica di rappresentazioni interne. Nei sistemi emergenti le azioni sono perturbazioni dell'ambiente effettuate dal sistema.
- **Predizione.** Nei sistemi cognitivisti la predizione consiste, tipicamente, nella pianificazione, utilizzando qualche forma di ragionamento di tipo procedurale o probabilistico attraverso un modello a priori. La predizione nel paradigma emergente richiede che il sistema visiti un numero

di stati all'interno del proprio spazio degli stati percezione-azione, senza necessariamente eseguire le azioni associate.

1.3 Modelli interni di tipo anticipativo

Nel precedente paragrafo è stato mostrato come l'approccio emergente fornisca un punto di vista innovativo ed opposto rispetto a quello fornito da quello cognitivista. Al fine della costruzione di artefatti intelligenti, che simulino i comportamenti umani è dunque necessario fare riferimento alle soluzioni adottate in natura in quanto rappresentano la struttura dalla quale emergono tali comportamenti: anziché cercare di emulare l'intelligenza come concetto astratto superiore, si simula l'architettura nella quale essa prende forma. Nella maggior parte dei sistemi biologici si osservano molteplici comportamenti di tipo reattivo necessari per risposte rapide ed efficienti che non necessitano di un controllo superiore complesso (figura 1.4). La distribuzione di differenti architetture reattive e la loro connessione parallela in un unico sistema permette il manifestarsi di comportamenti complessi che risulterebbero difficili da effettuare per un unico controllore, come nelle architetture gerarchiche. La distribuzione del controllo in più elementi semplici, invece, riduce il carico computazionale e mantiene la struttura flessibile. Per esempio, nel tatto, i recettori sono sempre in numero elevato e ridondante in modo che il mal funzionamento di uno non compromette le prestazioni dell'intero sistema. Il controllo reattivo a feed-back però non permette la pianificazione del comportamento dell'intero sistema. L'architettura necessita di un elemento responsabile della pianificazione dei comportamenti di alto livello, lasciando il controllo di basso livello alle strutture reattive. Tale sistema è rappresentato, nell'uomo, dal Sistema Nervoso Centrale (SNC). La gestione delle informazioni da parte dell' SNC risulta comunque decisamente complessa, in quanto le informazioni che provengono dall'ambiente sono numerose e di differente natura, basti pensare alla quantità di fibre che compongono il Sistema Nervoso Periferico, distribuite non uniformemente in tutto il corpo e che trasmettono le informazioni direttamente all'SNC. Questo rende impossibile una gestione dei compiti di percezione complessi da parte di un controllo puramente reattivo, tipo feed-back negativo. Infatti nella trasduzione sensoriale, nel processamento centrale e nell'uscita del segnale verso gli attuatori si riscontrerebbero ritardi notevoli. Per esempio nel sistema visivo, la retina introduce dai 30 ai 60 ms di ritardo, occorre, inoltre, considerare i ritardi dovuti alla conduzione del segnale nervoso che risultano maggiormente evidenti nel sistema tattile, i ritardi dovuti al processamento centrale, al sistema decisionale, al collo di bottiglia dovuto al comando motorio, e ai ritardi presenti nell'attuatore stesso. La combinazione di essi fornisce un inevitabile ritardo all'interno del circuito di feed-back negativo che può essere stimato tra i 30 ms per un riflesso spinale fino ai 200-300 ms per una risposta visiva [8]. Movimenti veloci e coordinati non possono essere eseguiti sotto un controllo con un puro feedback perché essi sono, nei sistemi biologici, troppo lenti ed hanno piccoli guadagni [9]. La soluzione a questo problema risulta essere necessariamente la predizione delle

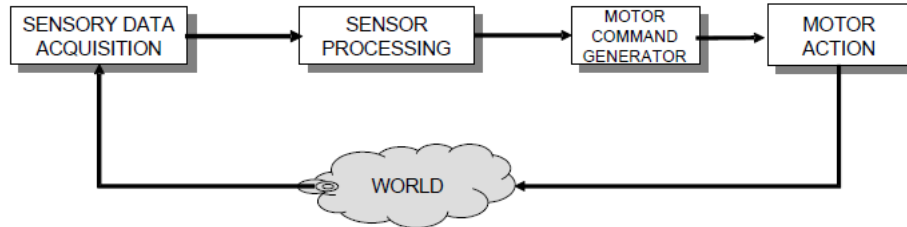


Figura 1.4: Il tipico sistema feedback riceve le informazioni dal mondo in cui agisce e restituisce un comando motorio per gli attuatori che modifica il mondo chiudendo il loop di controllo.

informazioni sensoriali e quindi della conseguenza dell'azione che il sistema sta per svolgere. Berthoz ritiene che in tale processo la memoria possiede un ruolo essenziale in quanto la memorizzazione del passato non è fine a se stessa, bensì può avere proprio lo scopo di anticipare le conseguenze delle proprie azioni o le risposte dell'ambiente [10]. I sistemi biologici complessi utilizzano il passato per predire il futuro e, successivamente, assimilano la conseguenza dell'azione al fine di adattare e migliorare la capacità di anticipazione. I sistemi che utilizzano l'anticipazione sono stati ampiamente studiati [8, 9, 11] e caratterizzati per modellizzare la capacità di coordinazione senso-motoria che contraddistinguono le capacità umane. Dato che, a causa dei ritardi sopra citati, non è possibile gestire tali attività con sistemi a feedback, è necessario ricorrere alla presenza di modelli interni che sfruttino l'anticipazione. I modelli interni sono meccanismi neurali che possono mimare la caratteristica input/output, o la sua inversa, dell'apparato motorio. Dalla copia efferente del comando motorio, il modello interno di tipo anticipativo può predire la conseguenza sensoriale del comando motorio. Dall'altro lato, un modello interno inverso può calcolare i comandi motori anticipativi dalle informazioni sulla traiettoria desiderata [9].

In generale i modelli interni possono essere classificati in tre gruppi distinti [12]. Il primo tipo è una rappresentazione causale dell'apparato motorio, spesso conosciuto come modello anticipativo (*feed-forward*). Esso è un modello che punta a mimare o rappresentare il normale comportamento del sistema motorio in risposta ai comandi che gli sono stati inviati. Per esempio, un modello anticipativo della dinamica di un braccio potrebbe avere come input lo stato attuale e i comandi motori che sono stati emanati dal controllore e produce come uscita una stima del nuovo stato. Questo modello quindi cattura il comportamento della transizione di stato del braccio in risposta ai comandi motori. Attraverso lo stato, riferito alla posizione e velocità corrente, è possibile conoscere il suo comportamento futuro. Lo stato può non essere accuratamente determinato dal controllore e quindi è necessario separare le variabili di stato dalle variabili ottenute dai sensori. Nei sistemi fisiologici infatti la posizione e la velocità di un braccio umano sono recepite dai fusi muscolari, i quali segnali

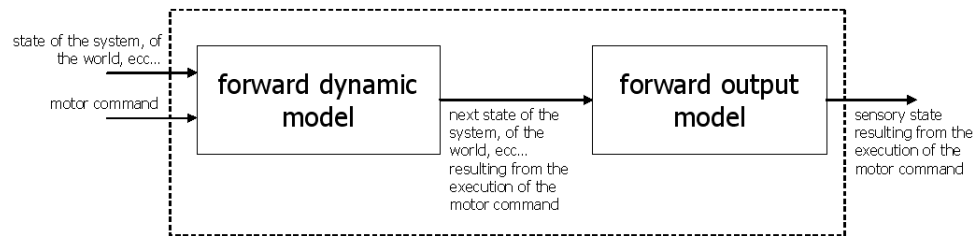


Figura 1.5: Modello *feed-forward*: in una prima fase si predice il prossimo stato del mondo come risultato dell'elaborazione dello stato attuale e dei comandi motori; in una seconda fase si utilizzano i comandi motori per predire la risposta sensoriale conseguente.



Figura 1.6: Un modello inverso è in grado di fornire i comandi motori necessari per ottenere la posizione desiderata.

cambiano in relazione alla lunghezza del muscolo piuttosto che agli angoli dei giunti. E' possibile, dunque proporre un modello anticipatorio dell'uscita sensoriale del braccio che potrebbe predire i segnali sensoriali conseguenti ad un particolare stato. Questo modello avrebbe in ingresso lo stato corrente e come uscita il feedback sensoriale predetto. Dalla cascata del modello anticipatorio della dinamica e del modello anticipatorio sensoriale si ottiene una stima delle conseguenze sensoriali di un determinato comando motorio (figura 1.5). Nel dominio cinematico il modello anticipatorio è preso come la mappatura tra le coordinate dei giunti e le coordinate finali della mano. Sebbene la stessa configurazione della mano può essere ottenuta con molte differenti configurazioni del braccio (a causa dei ridondanti gradi di libertà del braccio dei primati) non c'è ambiguità per la locazione della mano ottenuta da una particolare configurazione del braccio in quanto i modelli anticipatori sono causali. Il secondo gruppo di modelli interni sono concettualmente simili ai modelli anticipatori del sistema motorio, ma essi modellizzano il comportamento dell'ambiente esterno. Tali modelli cognitivi inglobano le conoscenze delle proprietà fisiche del mondo esterno. Noi comprendiamo il normale comportamento del mondo fisico, e per esempio, possiamo predire accuratamente la traiettoria di una palla che vorremmo catturare. Questi modelli sono comunque utili per la pianificazione dei comandi motori. Un terzo gruppo di modelli è conosciuto come modelli inversi, che invertono il causale flusso del sistema motorio (figura 1.6). Questi modelli

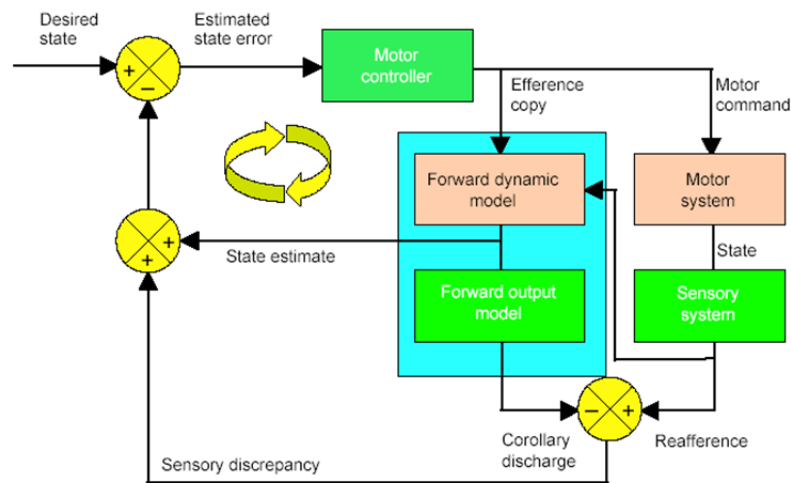


Figura 1.7: Nel predittore di Smith vengono utilizzati due modelli anticipativi all'interno di un loop tradizionale a feedback negativo. Il modello stima le conseguenze sensoriali delle azioni dei motori e tale stima serve a correggere l'azione. In questo modo è possibile superare alcuni dei ritardi dovuti al normale feedback negativo.

inglobano anch'essi le conoscenze del comportamento del sistema motorio, ma ricevono come input lo stato attuale e la transizione di stato e forniscono come output la rappresentazione degli eventi causali che producono quello stato. Per esempio, un modello di dinamica inversa del braccio stima il comando motorio che causa un particolare movimento. L'input potrebbe essere lo stato corrente e quello desiderato e l'output sarebbe il comando motorio che causa quel particolare cambiamento di stato. Un modello sensoriale inverso predice i cambiamenti di stato che corrispondono ad un cambiamento dell'afflusso sensoriale. Le applicazioni proposte per i modelli interni di tipo anticipativo sono diverse, significativa è quella conosciuta come predittore di Smith [8]. In questo caso vengono utilizzati due modelli anticipativi all'interno di un loop tradizionale a feedback negativo (figura 1.7). La traiettoria desiderata è lo stato da raggiungere ed è passato ad un controllore che calcola la differenza tra lo stato attuale e quello desiderato. La differenza è convertita in un comando motorio. Mentre il controllore sta eseguendo il comando una copia di esso è passata come input al modello anticipativo. Il modello stima le conseguenze sensoriali delle azioni dei motori e tale stima serve a correggere l'azione. In questo modo è possibile superare alcuni dei ritardi dovuti al normale feedback negativo. Un altro interessante schema basato su modelli anticipativi è quello proposto da Johansson [11]. In tale schema viene generata un'immagine attesa tattile e propriocettiva prima che il braccio afferrì un oggetto. Questa aspettazione sensoriale è usata per calibrare la presa e le forze di carico. Quando il braccio afferra l'oggetto,

di un oggetto di interesse (mira o target), è possibile solo se la sua immagine cade al centro della retina e vi resta stabile. I movimenti oculari hanno quindi la funzione di orientare lo sguardo verso una mira e di mantenerne stabile l'immagine sulla retina durante il movimento relativo tra osservatore e mira. Lo sguardo (*gaze*) è definito come la posizione dell'occhio nello spazio e corrisponde alla somma della posizione dell'occhio nell'orbita e della testa nello spazio. E' necessario, dunque, che l'obiettivo da seguire sia sempre centrato all'interno del campo visivo, anche quando è la testa stessa a muoversi. Il segnale che dà il via al processo di inseguimento è, infatti, il movimento dell'immagine sulla retina che può essere dovuto sia al movimento dell'oggetto stesso sia al movimento relativo della retina rispetto ad oggetti fermi, per esempio durante la camminata.

In un primo livello di classificazione dei movimenti oculari si distinguono movimenti di *verggenza* e di *versione* [16]. Nei primi gli occhi si muovono di uguale ampiezza ma in direzione opposta, per mantenere sulle due retine, in posizioni equivalenti l'immagine di una mira che si sposta in profondità rispetto all'osservatore. Nei movimenti di versione invece gli occhi si muovono nella stessa direzione. I movimenti oculari di versione possono essere a loro volta classificati in movimenti *lenti* e movimenti *rapidi*. I movimenti lenti possono raggiungere velocità dell'ordine di un centinaio di gradi al secondo. La velocità dei movimenti rapidi, caratterizzati da una forte accelerazione iniziale, è strettamente legata all'ampiezza del movimento, e nell'uomo può raggiungere valori di oltre 500°/sec. Esiste un solo movimento rapido, il movimento saccadico, mentre i movimenti lenti possono essere prodotti da meccanismi nervosi differenti tra loro e con finalità diverse: inseguimento di un target (smooth pursuit), stabilizzazione dello sguardo durante movimenti relativi tra il soggetto e l'ambiente visivo circostante (riflesso ottocinetico e riflesso vestibolo-oculomotore). I movimenti saccadici sono gli unici che possono essere eseguiti anche volontariamente, mentre i movimenti lenti sono sempre involontari. I movimenti di inseguimento, richiedono l'effettiva presenza di qualcosa in movimento nel campo visivo e quindi il controllo volontario dello smooth pursuit riguarda la possibilità di scegliere, *quale* target inseguire ed eventualmente *se* inseguirlo oppure no. La velocità di movimento dipende dai parametri di velocità e ampiezza del movimento del target e possono superare valori di circa 70°/sec. Quando la velocità dello smooth pursuit non è sufficiente per mantenere lo sguardo sul target compaiono saccadi di correzione. Il parametro più utilizzato per valutare le prestazioni del sistema è il *guadagno*, inteso come rapporto tra la velocità degli occhi e la velocità del target. Se l'operazione di inseguimento lento di un target fosse eseguita con un normale processo a feedback negativo, il sistema accumulerebbe eccessivi ritardi senza riuscire ad mantenere l'immagine al centro dell'occhio. I ritardi in questo caso sono dovuti principalmente alla frequenza di aggiornamento dell'immagine sulla retina, ai ritardi accumulati nella trasmissione dell'immagine dalla retina al SNC, alla sua elaborazione ed infine all'invio e all'esecuzione del comando motorio. E' stato accreditato da diversi studi che il cervello utilizza la predizione per eseguire questo compito [17, 18, 19] e che valori tipici di predizione si aggirano intorno ai 200 ms. E' interessante inoltre citare

gli studi di Claes von Hofsten che mostrano come tale abilità non sia innata, ma viene appresa gradualmente e diventa matura a circa 6 mesi dalla nascita del bambino [20, 21]. La pianificazione delle azioni deve anticipare gli stati del mondo che saranno ottenuti quando l'azione avrà luogo e quindi le azioni effettive dipendono dalla predizione. Gli esperimenti presentati da von Hofsten esplorano i principi sottesi alla predizione, nello sviluppo infantile, del movimento degli oggetti, focalizzandosi su l'inseguimento e il raggiungimento di oggetti in movimento da parte di bambini di varie fasce di età. Il ritardo nell'inseguimento decresce nel corso dello sviluppo dai 2 mesi fino a 6 mesi e viene compensato da movimenti saccadici. E' stato inoltre osservato che l'inseguimento degli oggetti lungo l'asse orizzontale viene appreso prima rispetto ai movimenti lungo l'asse verticale dimostrando che queste due componenti seguono sviluppi distinti e solo intorno ai 12 mesi diventano maturi per l'inseguimento di traiettorie che richiedono entrambe le componenti [22, 23].

1.5 Sviluppo della capacità di inseguimento di target in movimento

I bambini nascono con un sistema visivo immaturo [24]. Tra le abilità studiate in soggetti adulti e che non sono state trovate nei neonati c'è la capacità di inseguire oggetti in movimento con movimenti lenti degli occhi (smooth pursuit) che emerge nei primi sei mesi di vita. Nel primo mese tale capacità è particolarmente carente, avviene solo con stimoli lenti e generalmente è composta principalmente da movimenti saccadici. Tra le 4 e le 8 settimane di vita, i movimenti di smooth pursuit si manifestano su stimoli lenti e molto ampi [25], i movimenti saccadici di correzione sono molto frequenti. Tra i 2 e 6 mesi avviene un rapido sviluppo della capacità di inseguire con movimenti lenti [20, 21]. Questi studi mostrano che il controllo dello sguardo subisce due particolari cambiamenti nei primi sei mesi di vita. Un cambiamento è la velocità del target che lo smooth pursuit può inseguire, infatti i neonati sotto le 8 settimane possono seguire target che si muovono sotto i $10^\circ/\text{sec}$. L'inseguimento sopra tale velocità viene eseguito principalmente con movimenti saccadici. Dopo i 6 mesi di età le velocità di inseguimento raggiungono quelle adulte. Un secondo cambiamento è l'abilità dei bambini di usare lo smooth pursuit a velocità sostenute per la gran parte della durata dell'inseguimento. Prima delle sei settimane stimoli con velocità sopra ai $9^\circ/\text{sec}$ sono principalmente seguiti con movimenti saccadici. La velocità alla quale i target sono seguiti con smooth pursuit aumenta gradualmente fino ai 6 mesi di vita. Negli studi di von Hofsten [21] bambini di 9, 15 e 21 settimane (2, 3 e 5 mesi circa) inseguono stimoli sinusoidali e triangolari con velocità di 10, 20 e $40^\circ/\text{sec}$. La percentuale di smooth pursuit aumenta fino al quinto mese. Per esempio in figura (1.9) si mostra che per velocità di $10^\circ/\text{sec}$ la percentuale passa dall'80% all'87% rispettivamente dai 2 ai 5 mesi di vita. Per velocità di $20^\circ/\text{sec}$ si passa dal 65% all'80%. A $40^\circ/\text{sec}$ si passa invece dal 45% fino ad un massimo del 70%. Fattore importante per questo tipo di

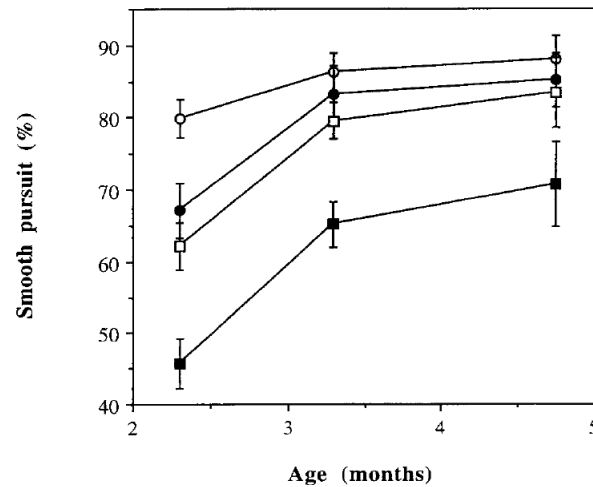


Figura 1.9: Media e deviazione standard della percentuale di guadagno di smooth pursuit per le differenti ampiezze e frequenze rispetto all'età del neonato. (cerchi = 0.2 Hz, quadrati = 0.4 Hz, vuoti = 10° di ampiezza, pieni = 20° di ampiezza).

ricerche risulta essere l'attenzione che il bambino rivolge all'oggetto da inseguire [24]. Stimoli più interessanti (come la faccia di personaggi dei cartoni animati) aumentano queste percentuali. Altri studi [26, 27, 28] mostrano come la pianificazione e il controllo delle azioni debba precedere la loro esecuzione. Occorre essere in grado di predire come gli eventi evolvano nell'imminente futuro. Per l'inseguimento di un oggetto in movimento, in particolare, è necessario predire la sua dinamica. Questa abilità è stata studiata facendo seguire ai bambini una traiettoria lineare. Se viene bloccato l'oggetto in modo improvviso, lo sguardo del bambino continua sulla traiettoria lineare per almeno 200 ms. In questo caso si considera il movimento della testa relazionato con quello degli occhi.

Nel prossimo capitolo verranno illustrati i meccanismi biologici alla base del riconoscimento del movimento da parte del SNC così da comprendere la natura dei modelli che riproducono questo comportamento.

Capitolo 2

Neurofisiologia del sistema visivo

2.1 Introduzione

Le ricerche sull'intelligenza artificiale hanno dimostrato che il cervello dell'Uomo analizza l'ambiente esterno e mette in atto le attività motorie, mediante meccanismi che, a tutt'oggi, qualsiasi computer è lungi dal poter emulare ([29] capitoli 25-26). Le nostre percezioni, quali la vista, l'udito, l'olfatto, il gusto ed il tatto richiedono una capacità di elaborazione prodigiosa. Analogamente, i nostri movimenti volontari richiedono capacità ingegneristiche altrettanto prodigiose. Il sistema nervoso centrale è in grado di eseguire questi compiti nel campo dell'elaborazione delle informazioni in quanto le cellule nervose che lo costituiscono sono connesse fra di loro in modo estremamente preciso. Le neuroscienze hanno dimostrato che le percezioni non sono affatto semplici copie del mondo che ci circonda. Le sensazioni sono astrazioni, non repliche, del mondo in cui viviamo. Il sistema nervoso centrale dopo aver preliminarmente analizzato diverse caratteristiche degli eventi fisici che hanno luogo nell'ambiente esterno se ne costruisce una sua rappresentazione interna. Ogni esperienza consiste in un'analisi simultanea e separata delle varie caratteristiche che la compongono ed i risultati di quest'analisi vengono integrate in un'esperienza cosciente. Per quanto riguarda il sistema visivo, infatti, le caratteristiche di forma, colore, orientamento e movimento sono analizzate in modo separato. Lo scopo di questo capitolo è mettere in evidenza come il cervello interpreta il movimento di un oggetto che si muove nel suo campo visivo, e come questa informazione possa condurre ad un consequenziale movimento degli occhi atto all'inseguimento dell'oggetto stesso.

2.2 Analisi delle informazioni visive nella retina

I processi che conducono alla percezione visiva prendono inizio nella retina e si svolgono in due stadi. La luce che penetra attraverso la cornea viene dapprima proiettata sul fondo dell'occhio, dove viene trasformata in segnali elettrici da un organo sensoriale specializzato detto *retina* (figura 2.1). Questi segnali vengono poi convogliati, attraverso il nervo ottico, verso centri cerebrali superiori, dove vengono elaborati ulteriormente fino a livello percettivo. A differenza di altre strutture sensoriali, come quelle dei recettori del sistema uditivo o tattile, la retina non è un organo periferico ma una parte del sistema nervoso centrale. La luce viene focalizzata dalla cornea e dal cristallino e deve attraversare l'umor vitreo, che occupa la cavità dell'occhio, prima di raggiungere i fotorecettori retinici. Quest'ultimi sono di due tipi: i coni, responsabili della visione diurna e dei colori; e i bastoncelli responsabili della visione notturna. La retina sta davanti all'epitelio pigmentato che ricopre la superficie posteriore dell'occhio e contiene una grande quantità di pigmento nero che assorbe tutta la luce che non è stata trattenuta dalla retina. I fotorecettori sono localizzati nella parte posteriore dell'occhio, immediatamente davanti all'epitelio pigmentato, mentre tutte le altre cellule retiniche sono disposte davanti ai fotorecettori, in posizione più prossima al cristallino. Queste cellule sono però abbastanza trasparenti e quindi la luce può raggiungere i recettori senza essere apprezzabilmente distorta. Esiste inoltre una zona della retina, la *fovea*, nella quale i corpi cellulari dei neuroni degli strati prossimali della retina sono spostati lateralmente, e ciò mette in grado i fotorecettori foveali di ricevere l'immagine visiva nella sua forma meno distorta possibile. Inoltre in questa zona, il numero di fotorecettori è estremamente alto se paragonato alle altre zone della retina, per questo motivo l'Uomo muove costantemente gli occhi in modo che le immagini che guidano la sua attenzione vadano a cadere sulla fovea per una migliore analisi visiva.

I neuroni di uscita della retina sono le *cellule gangliari*. A differenza dei fotorecettori che rispondono alla luce con modificazioni graduali del loro potenziale di membrana, le cellule gangliari trasmettono le proprie informazioni sotto forma di scariche di potenziali d'azione. I loro assoni vanno a formare il nervo ottico che proietta al talamo, nel corpo genicolato laterale, al collicolo superiore, al pretetto e ad altri nuclei bersaglio (figura 2.2). Fra i fotorecettori e le cellule gangliari sono interposte tre classi diverse di interneuroni: le cellule bipolari, le cellule orizzontali e le cellule amacrine. Queste tipologie di cellule combinano i segnali luminosi trasmessi da numerosi fotorecettori in modo tale che le risposte elettriche che nascono nelle cellule gangliari dipendono in modo critico dalle caratteristiche spaziali e temporali degli stimoli luminosi che hanno colpito la retina. Le cellule gangliari non sono mai silenti, ma posseggono un'attività spontanea che viene modulata dalle afferenze provenienti dai diversi tipi di interneuroni retinici. Le afferenze di ogni cellula gangliare originano da gruppi di fotorecettori circoscritti, localizzati in una ristretta area di retina che costituisce il campo recettivo della cellula gangliare stessa. Esso rappresenta quindi l'area di retina che è sotto il controllo di una certa cellula gangliare. I campi recettivi si possono suddividere in due parti: una zona circolare, al centro,

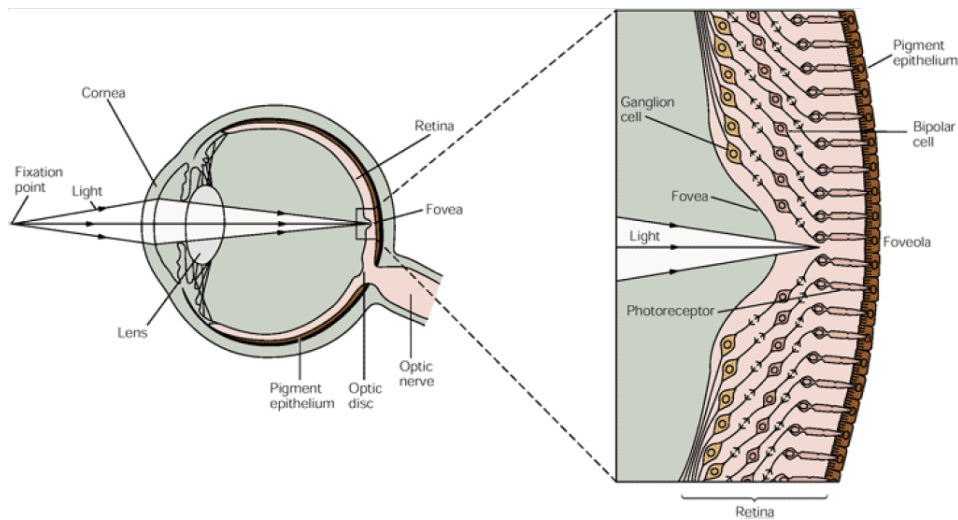


Figura 2.1: Nell'immagine a sinistra viene mostrata la localizzazione della retina all'interno dell'occhio. Il dettaglio mostra il centro della retina, la fovea dove i neuroni prossimali sono spostati lateralmente in modo da ottenere immagini poco distorte.

detta centro del campo recettivo, e la restante parte, detta periferia. Queste due zone rispondono alla luce in maniera antagonista ed in base a questo è possibile dividere le cellule gangliari in due classi. Le cellule gangliari *centro-on* vengono eccitate quando si proietta la luce sul centro del loro campo recettivo. La luce applicata alla periferia del campo recettivo inibisce l'effetto prodotto dall'illuminazione del centro. Le cellule gangliari *centro-off* sono invece inibite dalla luce proiettata nel centro del loro campo recettivo ed eccitate da stimoli luminosi che cadono sulla periferia del loro campo recettivo. In entrambi questi tipi di cellule, l'illuminazione diffusa di tutto il campo recettivo è in grado di evocare soltanto una risposta debole (figura 2.3). Il numero delle cellule gangliari centro-on e centro-off è approssimativamente uguale ed ogni fotorecettore invia i propri segnali d'uscita a cellule gangliari di entrambi i tipi. In tal modo, le cellule gangliari vanno a costituire due vie in parallelo per l'elaborazione dell'informazione visiva. Inoltre, le dimensioni dei campi recettivi variano notevolmente da una regione della retina all'altra. Nella fovea dei primati, dove massima è l'acuità visiva, i campi recettivi sono piccoli ed hanno un centro che si estende soltanto per pochi gradi (1° sulla retina corrisponde a circa 0.25 mm). Alla periferia della retina, dove l'acuità visiva è bassa, i campi sono molto più grandi ed hanno zone centrali che vanno da 3° a 5° . L'organizzazione delle cellule gangliari basata sull'antagonismo centro-periferia dei loro campi recettivi implica il fatto che tali cellule analizzano soprattutto i contrasti di luminosità presenti nello scenario visivo e non l'intensità assoluta dell'illuminazione stessa. Pertanto, la maggior parte delle informazioni utili, in uno scenario visivo, risiede nella distribuzione

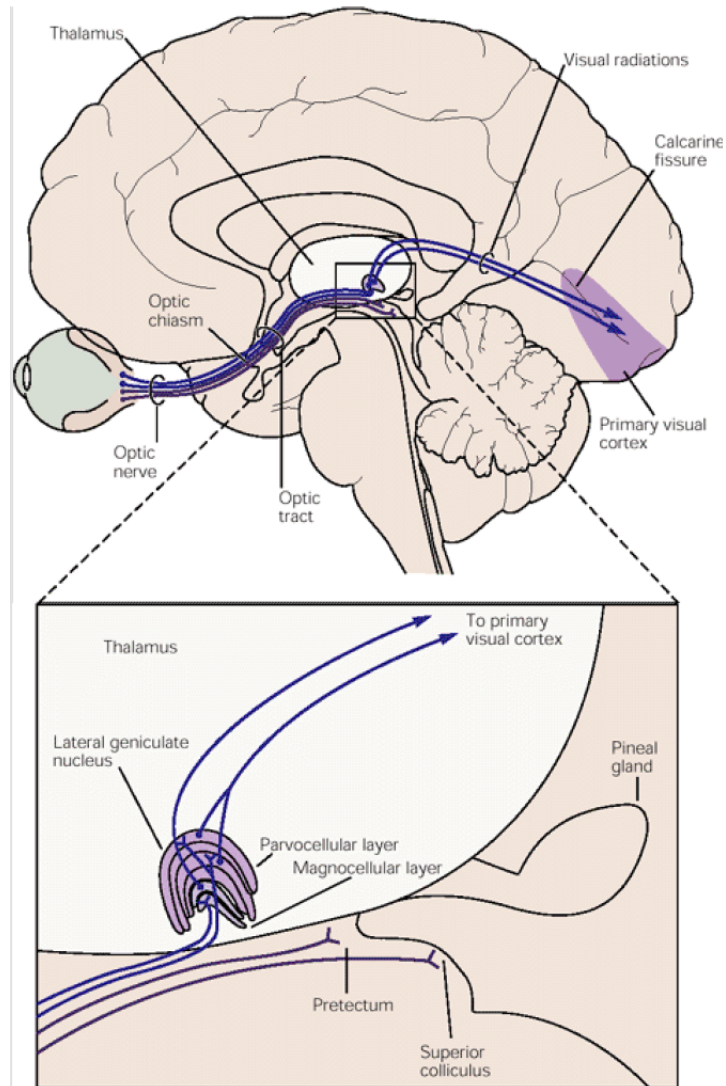


Figura 2.2: Diagramma semplificato delle proiezioni che partono dalla retina per le aree visive del talamo (corpo genicolato laterale), il collicolo superiore e il pretetto.

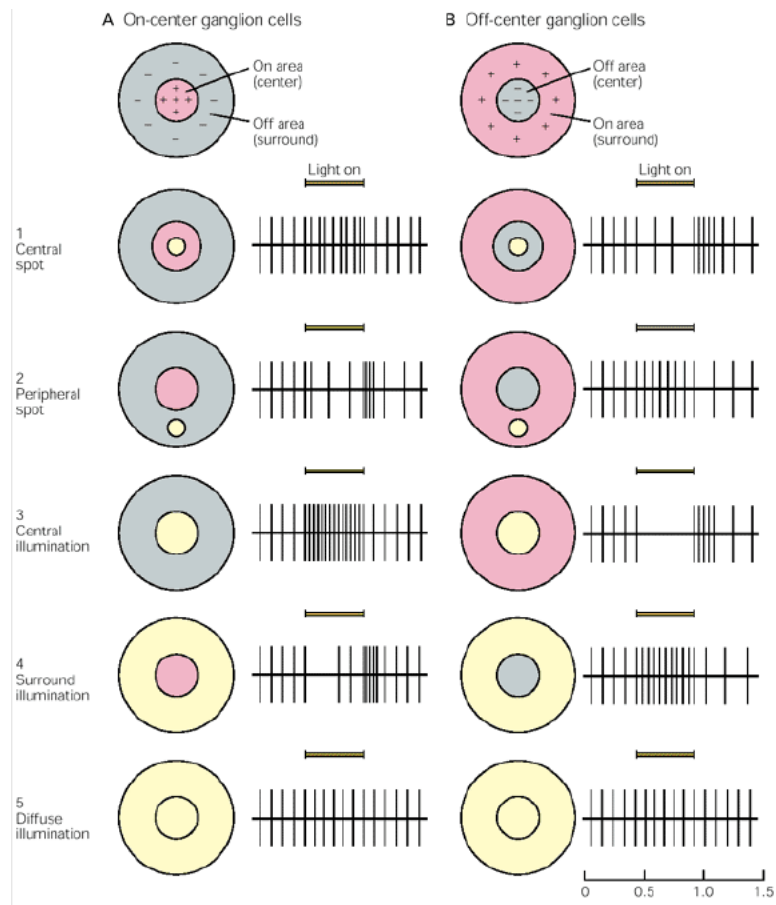


Figura 2.3: Le cellule gangliari rispondono ottimamente ai contrasti di luce nel loro campo recettivo. In A, una cellula centro on risponde ottimamente quando il suo campo centrale è completamente stimolato (3). In B, una cellula centro off aumenta la frequenza di scarica della risposta se viene illuminata la periferia del loro campo recettivo (4).

dei contrasti di luminosità. I valori assoluti di luminosità danno assai poche informazioni in quanto dipendono in gran parte dall'intensità della sorgente stessa; se si raddoppia la luminosità ambientale la quantità di luce riflessa degli oggetti raddoppierà, mentre i rispettivi contrasti rimarranno costanti. L'organizzazione centro-periferia costituisce un adattamento funzionale utile per mettere in evidenza le informazioni e minimizzare il rischio di errori nella trasmissione del segnale dalla retina alla corteccia. Anche il fatto che l'analisi dell'informazione sia separata in due vie parallele, rispettivamente centro-on e centro-off, migliora le prestazioni del sistema visivo in quanto ogni tipo di cellula gangliare segnala in maniera ottimale sia rapidi aumenti che rapide diminuzioni del livello di illuminazione. Un aumento della frequenza di scarica delle cellule centro on, infatti, segnala bene un aumento dell'intensità luminosa nel loro campo recettivo, mentre un aumento della frequenza di scarica delle cellule centro off segnala una diminuzione di illuminazione nel loro campo recettivo. Le diverse risposte delle cellule gangliari sono espressione dei differenti tipi di contatti sinaptici presenti nella retina. Fra i fotorecettori e le cellule gangliari vi sono interposti, come interneuroni, le cellule bipolari. Anche le cellule bipolari si possono distinguere in centro on e centro off. Ciascun cono ha contatti sinaptici con cellule bipolari di entrambi i tipi. I coni localizzati nel centro del campo recettivo di una cellula gangliare fanno sinapsi con cellule bipolari che entrano in diretto contatto con la cellula gangliare stessa. Gli stimoli provenienti dai coni localizzati alla periferia dei campi recettivi vengono invece convogliati lungo vie collaterali che passano attraverso le cellule orizzontali e le cellule amacrine.

Il sistema visivo analizza in vie poste in parallelo, anche altre informazioni oltre al contrasto e alle variazioni di illuminazione, come il colore, la forma e il movimento. Ogni zona della retina possiede sottogruppi diversi di cellule gangliari funzionalmente distinte. La maggior parte possono venir classificate in due gruppi, le cellule M (per *magnae* o grandi) e le cellule P (per *parvae* o piccole). Ciascuno di questi gruppi comprende sia cellule centro-on che centro-off. Le cellule M posseggono campi recettivi grandi (il che è coerente con la grande estensione del loro albero dendritico) e danno risposte relativamente fugaci durante l'esposizione protratta alla luce. Queste cellule rispondono bene a stimoli rappresentati da oggetti di grandi dimensioni e la loro frequenza di scarica è in grado di seguire fedelmente variazioni rapide dell'intensità degli stimoli. La loro funzione sembra pertanto in rapporto all'analisi delle caratteristiche grossolane degli stimoli luminosi e del loro movimento. Le cellule P sono più piccole e molto più numerose, hanno campi recettivi piccoli, danno risposte selettive per le diverse lunghezze d'onda e sono deputate alla percezione delle forme e dei colori. La loro funzione sembra quindi legata all'analisi dei minuti dettagli delle immagini visive.

2.3 Le vie visive centrali

Le immagini retiniche e l'immagine del campo visivo sono ben distinte ([29] capitolo 27). La superficie retinica può essere suddivisa in una *emiretina nasale*, che

sta medialmente alla fovea, ed in una *emiretina temporale*, situata lateralmente a questa. Ogni emiretina può essere, a sua volta, distinta in un quadrante *dorsale* (o superiore) e in un quadrante *ventrale* (o inferiore). Il campo visivo invece è la parte del mondo esterno che viene vista dai due occhi in assenza di movimento del capo. E' possibile definire la metà destra e sinistra del campo visivo quando le fovee di entrambi gli occhi fissano un sol punto nello spazio. L'*emicampo visivo sinistro* proietta le proprie immagini sull'emiretina nasale dell'occhio sinistro e sull'emiretina temporale dell'occhio destro. L'*emicampo visivo destro* proietta invece le proprie immagini sull'emiretina nasale dell'occhio destro e su quella temporale dell'occhio sinistro. La luce che proviene dalla zona centrale del campo visivo colpisce entrambi gli occhi e va a formare la *zona binoculare*. Il disco ottico, che è la zona della retina da cui fuoriescono gli assoni delle cellule gangliari, non contiene fotorecettori, è del tutto insensibile alla luce e costituisce una *macchia cieca* nella retina. Poiché il disco ottico è posto medialmente ad entrambe le fovee, la luce proveniente da un punto qualunque della parte centrale del campo visivo non può mai andare a cadere simultaneamente su entrambe le macchie cieche e perciò noi, normalmente, non ci rendiamo conto della loro esistenza. Le informazioni visive che dalla retina vengono trasmesse alle stazioni centrali, sono il risultato di una corrispondenza esistente tra il campo visivo e le rispettive immagini visive. Tale corrispondenza segue due proprietà: anzitutto, il cristallino inverte le immagini visive. La metà superiore del campo visivo si proietta sulla metà inferiore della retina e viceversa. In secondo luogo, un punto luminoso presente nella parte binoculare dell'emicampo visivo di destra andrà a cadere sulla emiretina temporale dell'occhio sinistro e sulla emiretina nasale dell'occhio destro. Gli assoni di tutte le cellule gangliari retiniche vanno a confluire nel disco ottico e, a livello del chiasma ottico, le fibre delle metà *nasali* di ogni retina si incrociano e proiettano all'emisfero cerebrale contralaterale. Le fibre provenienti dalle emiretine temporali, al contrario, non si incrociano. In tal modo, le fibre del chiasma ottico provenienti dalle due retine entrano a far parte dei tratti ottici destro e sinistro. Per via di tale disposizione, gli assoni provenienti dalle *metà sinistre di ogni retina* (l'emiretina temporale dell'occhio sinistro e l'emiretina nasale di quello destro) proiettano nel tratto ottico sinistro, che, in tal modo contiene una rappresentazione completa *dell'emicampo visivo destro*. Le fibre che provengono dalla *metà destra di ogni retina* (l'emiretina nasale dell'occhio sinistro e l'emiretina temporale di quello destro) proiettano invece al tratto ottico destro e contengono una rappresentazione completa *dell'emicampo visivo sinistro*. Questa separazione dell'emicampo visivo di destra nel tratto ottico sinistro e dell'emicampo visivo sinistro nel tratto ottico destro viene mantenuta in tutte le proiezioni ai nuclei visivi sottocorticali (figura 2.4).

Gli assoni di tutte le cellule gangliari retiniche vanno a confluire nel disco ottico, dove divengono mielinici e vanno a formare, bilateralmente, i nervi ottici. I nervi ottici proiettano al chiasma ottico e da questo gli assoni proiettano a tre principali stazioni sottocorticali: il pretetto, il collicolo superiore e il copro genicolato laterale. L'area pretettale del mesencefalo controlla i riflessi pupillari e permette di regolare la quantità di luce che colpisce la retina, il collicolo superiore controlla invece i movimenti saccadici dell'occhio, mentre il corpo

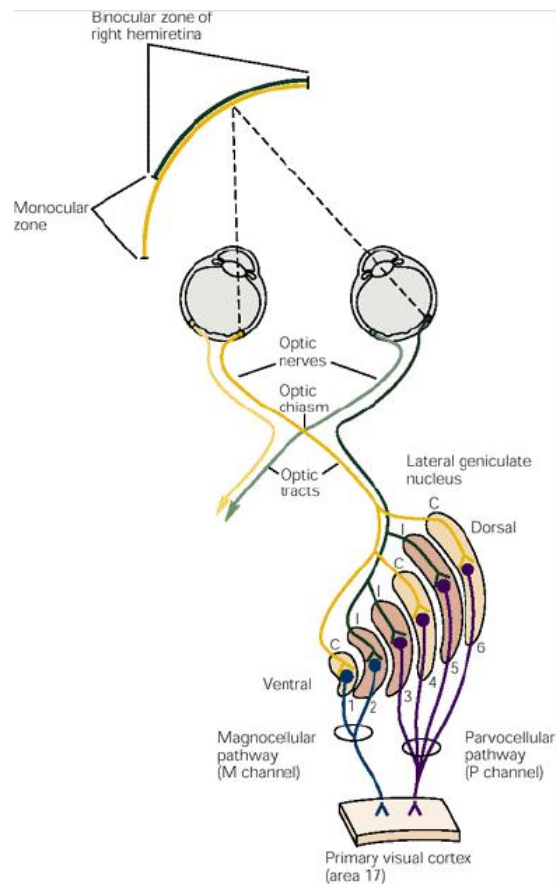


Figura 2.4: Il corpo genicolato laterale è il sito principale subcorticale per l'elaborazione delle informazioni visive. I nuclei dall'emiretina destra di ogni occhio proiettano in strati differenti del corpo genicolato laterale destro completando l'informazione dell'emisfero visivo sinistro. In modo analogo, le fibre dell'emiretina sinistra di ogni occhio proiettano nel corpo genicolato laterale sinistro.

genicolato laterale è la principale stazione di ritrasmissione delle informazioni afferenti destinate alla corteccia visiva. Il novanta per cento degli assoni retinici va a terminare nel corpo genicolato laterale, che è la più importante stazione sottocorticale che trasmette le informazioni visive alla corteccia cerebrale. Le cellule gangliari della retina vanno a terminare ordinatamente nei diversi punti del corpo genicolato laterale in modo che in quest'ultimo viene a costituirsi una rappresentazione retinotopica della metà contralaterale del campo visivo. La rappresentazione non è però spazialmente uniforme, la fovea, che è l'area retinica dove maggiore è la densità delle cellule gangliari, ha una rappresentazione molto maggiore di quella della periferia della retina. Circa la metà della massa del corpo genicolato laterale (e della corteccia visiva primaria) è deputata a rappresentare la fovea e le aree ad essa circostanti. In tale zona la densità delle cellule gangliari è molto elevata a compenso del fatto che l'area centrale della retina è molto minore di quella della sua parte periferica (per via della concavità della retina stessa). Poiché tale limitazione al di fuori della retina non esiste più, la distribuzione dei neuroni diviene del tutto uniforme e di conseguenza, le proiezioni che provengono dai neuroni della fovea, che sono molto numerosi, andranno ad occupare un'area molto estesa. Nei primati, compreso l'Uomo, il corpo genicolato laterale è composto da sei strati di corpi cellulari separati da strati intralaminari di fibre nervose e di dendriti. I diversi strati vengono indicati con i numeri dall'1 al 6 procedendo in senso ventro-dorsale. I due strati più ventrali del corpo genicolato contengono cellule relativamente grandi e vengono perciò detti *strati magnocellulari*; le loro principali afferenze provengono dalle cellule gangliari M della retina. I quattro strati più dorsali sono invece detti *strati parvocellulari* e ricevono afferenze dalle cellule gangliari P della retina. Ogni strato del nucleo riceve afferenze da un solo occhio: le fibre che provengono dall'emiretina nasale contalaterale prendono contatto con gli strati 1, 4 e 6: quelle che prendono origine dall'emiretina temporale ipsilaterale terminano negli strati 2, 3 e 5. L'organizzazione strutturale centro-on e centro-off delle cellule gangliari viene pure mantenuta (figura 2.4).

La separazione anatomica di cellule gangliari retiniche (di tipo M e di tipo P) viene mantenuta sia a livello del corpo genicolato laterale che a livello della corteccia visiva primaria e ciò induce a pensare che esistano due vie parallele, distinte anche da un punto di vista funzionale. Tali vie vengono anche indicate come via M e via P. Esistono differenze notevoli fra le proprietà delle cellule della via M e della via P. La differenza più evidente riguarda la loro sensibilità verso i contrasti di colore. Le cellule P rispondono bene alle variazioni di colore (rosso/verde e giallo/blu) indipendentemente dalla luminosità relativa dei colori. Altra differenza è il *contrasto di luminanza* che è una risposta della differenza fra la parte più illuminata e la parte più scura dello stimolo. Le cellule M rispondono anche quando tale contrasto si riduce al 2%. Le cellule M e P differiscono anche per le loro risposte alle frequenze spaziali e temporali. La *frequenza spaziale* di uno stimolo è il numero di volte con cui una certa struttura viene replicata in un certo spazio. La *frequenza temporale* è una misura della rapidità con cui una struttura visiva varia nel tempo; per esempio, se le linee di un reticolo vengono illuminate e oscurate 10 volte al secondo, la loro

frequenza temporale è di 10 Hz. Le cellule M presentano risoluzioni spaziali minori ma risoluzioni temporali maggiori delle cellule P. La differenza nel tipo di risposta delle cellule M e P implica che esse forniscono contributi diversi alla percezione visiva. Le cellule P sono fondamentali per la visione dei colori ed hanno massima importanza per le discriminazioni visive che necessitano alta frequenza spaziale e basso potere risolutivo temporale. Le cellule M, al contrario, danno il loro massimo contributo visivo alle discriminazioni nelle quali è richiesto un basso potere risolutivo spaziale e un alto potere risolutivo temporale. Dal corpo genicolato laterale, le informazioni visive passano direttamente alla corteccia visiva primaria o area visiva 1 (V1), detta anche *corteccia striata*. Al pari del corpo genicolato laterale e del collicolo superiore, la corteccia visiva primaria riceve informazioni soltanto dalla metà contralaterale del campo visivo. Essa è composta da sei strati di cellule e il principale strato d'ingresso delle fibre che arrivano dal corpo genicolato laterale è lo strato IV, che viene ulteriormente suddiviso in quattro sottostrati. Gli assoni delle cellule M e P del corpo genicolato laterale vanno a terminare in sottostrati diversi dello strato IV. La separazione delle vie parvicellulare e magnocellulare continua perciò a mantenersi anche a questo livello di elaborazione. Nella corteccia visiva primaria è possibile riscontrare un'organizzazione retinotopica della struttura cellulare nella quale, i campi recettivi continuano ad avere forma circolare. La corteccia elabora le informazioni in almeno tre modi: informazioni relative all'orientamento, necessario per la discriminazione delle forme; informazioni relative ai colori, che non posseggono selettività per l'orientamento; e informazioni provenienti dai due occhi che vengono riunite in specifiche colonne permettendo la percezione del senso di profondità. La via M e la via P si continuano due vie distinte extrastriate: una via dorsale ed una via ventrale. Studi dimostrano che la via dorsale, che si estende fino alla corteccia parietale posteriore, media le informazioni riguardanti la percezione del movimento e del senso di profondità, mentre la via ventrale, che raggiunge la corteccia inferotemporale è interessata al riconoscimento delle forme e dei colori (figura 2.5).

2.4 La percezione del movimento

Molti neuroni della corteccia striata (V1) sono sensibili all'*orientazione*: cioè, se una linea viene a cadere nel campo recettivo di una cellula e ruota sul proprio centro, la cellula risponderà soltanto quando la linea avrà assunto una particolare angolatura ([29] capitolo 28). Tali neuroni sono definiti *cellule semplici*. Tali cellule hanno una periferia di tipo inibitoria: una linea con particolare orientazione eccita una particolare cellula solo se si trova al centro del proprio campo recettivo, altrimenti produce un effetto inibitorio. Un altro tipo di cellula viene definita *cellula complessa* e risponde bene a una linea orientata in modo particolare, ma non sembra possedere una periferia inibitoria: in altre parole, questo neurone continua a rispondere quando la linea viene spostata all'interno del suo campo recettivo. Infatti, molte cellule complesse aumentano la loro frequenza

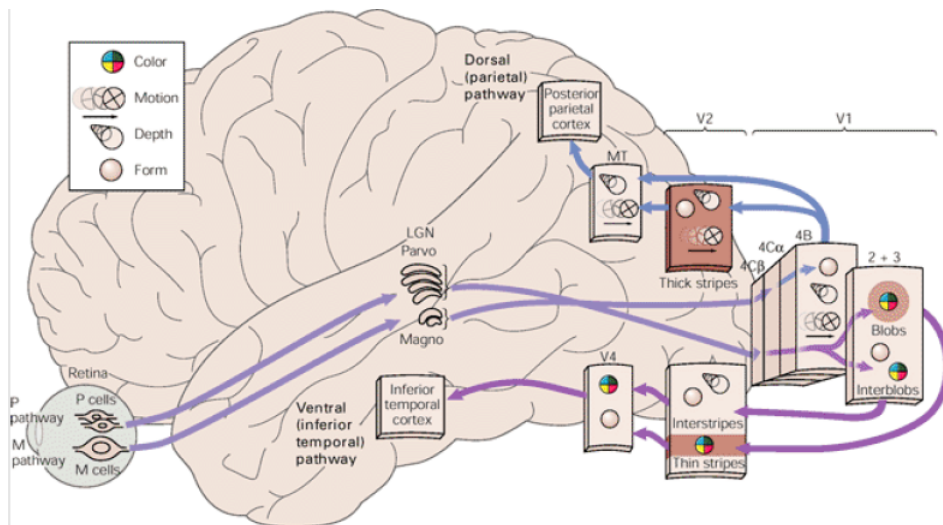


Figura 2.5: Possibili funzioni mediate dai due percorsi distinti del processamento nella corteccia cerebrale delle informazioni visive. Il percorso che inizia dalle cellule M si estende fino alla corteccia parietale posteriore processando informazioni di movimento, profondità e località spaziale. Il percorso che inizia dalle cellule P arriva fino alla corteccia temporale inferiore che è relazionata al riconoscimento di forme e colori.

di scarica, quando la linea viene spostata perpendicolarmente al suo angolo di orientazione; in tal modo, esse svolgono la funzione di rilevatrici del movimento.

Poiché la maggior parte delle cellule del sistema visivo sono estremamente sensibili alla posizione retinica delle immagini e possono distinguere eventi separati nel tempo da 10 o 20 millisecondi, in teoria la maggior parte dei neuroni del sistema visivo dovrebbero essere in grado di ottenere tutte le informazioni necessarie paragonando la localizzazione di un oggetto sulla retina istante per istante. Spesso, tuttavia, noi vediamo muoversi gli oggetti non perché essi si muovono sulla nostra retina, ma perché noi ne cogliamo l'immagine muovendo gli occhi; l'immagine resta ferma sulla retina ma noi percepiamo un movimento in quanto abbiamo mosso gli occhi. Occorre supporre, dunque l'esistenza di un sottosistema neurale specializzato per l'analisi del movimento. Nei primati, un'area situata al bordo della corteccia parietale, l'area *mediotemporale* (MT, negli esseri umani nota come V5), risulta dedicata all'analisi del movimento. Analogamente alla V1, anche MT possiede una mappa retinotopica del campo visivo contralaterale, ma i campi recettivi delle cellule di questa mappa sono circa 10 volte più estesi di quelli delle cellule della corteccia striata. Ogni parte del campo visivo è rappresentata da un gruppo di colonne nel quale le cellule rispondono ad ogni direzione possibile del movimento in quella parte del campo visivo. Le cellule di MT rispondono bene al movimento di macchie o sbarrette luminose rilevandone i contrasti di luminanza. L'area V1 risponde a movimenti di stimoli semplici come quelli di un margine o una linea che si muove nel campo visivo, tuttavia, nell'esperienza quotidiana ci imbattiamo spesso in sagome complesse, il cui movimento può dare origine ad ambiguità o percezioni illusorie. Un esempio tipico è detto il *problema dell'apertura* (figura 2.6) nel quale si osserva l'immagine di una struttura a righe che si muove in tre direzioni diverse. Quando l'immagine viene osservata attraverso un'apertura circolare molto piccola, la direzione in cui si muove appare sempre la stessa in tutti e tre i casi illustrati. L'osservatore riesce a rilevare soltanto la componente del movimento che è perpendicolare all'orientamento delle linee dell'immagine stessa. Per risolvere il problema dell'apertura, in uno stadio iniziale i neuroni che rispondono ad assi di orientamento specifici segnalano le componenti del movimento perpendicolari al loro asse di orientamento preferenziale; successivamente, nel secondo stadio, la percezione del movimento dell'intera struttura viene ricavato a partire dalle diverse componenti dello stadio iniziale di elaborazione. I neuroni di V1 e la maggior parte di MT si limitano a mettere in evidenza la direzione del movimento delle singole componenti. Al contrario circa il 20% dei neuroni di MT risponde soltanto al movimento globale dell'intera struttura. Questi neuroni ricevono afferenze dalle cellule selettive per le componenti della direzione del movimento. Anche un'altra area corticale adiacente ad MT, l'area *mediotemporale superiore* (MST, negli esseri umani nota come V5a), possiede neuroni che rispondono a stimoli visivi in movimento e sembra che tali neuroni siano deputati ad analizzare un tipo globale di movimento, detto *flusso ottico*, che risulta particolarmente importante in quanto permette ad ogni individuo di avere la percezione del proprio movimento attraverso l'ambiente circostante (figura 2.7). Il flusso ottico è fonte di due tipi di informazioni: informazioni relative all'am-

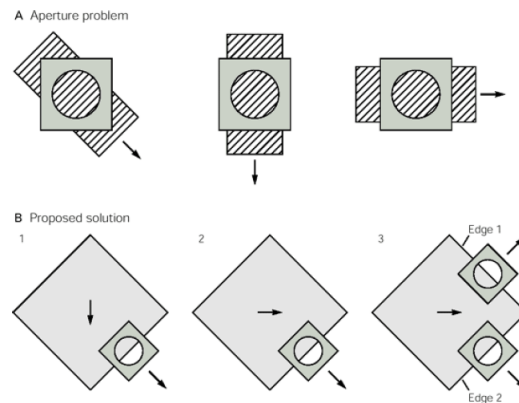


Figura 2.6: Problema dell'apertura. In A tra pattern si muovono in direzioni differenti, ma producono lo stesso stimolo fisico all'interno del foro, quindi vengono percepiti come se si stessero muovendo nella stessa direzione. In B la soluzione risulta quella di possedere più aperture. Questo permette l'identificazione della direzione del movimento. Occorre assumere che le aperture rappresentino il campo recettivo di un neurone.

biente circostante (gli oggetti più vicini sembrano muoversi più velocemente di quelli più lontani) e informazioni relative al controllo della postura (i movimenti in senso laterale del campo visivo inducono oscillazioni del corpo). Per esempio, quando un individuo si muove in avanti con gli occhi e la testa sempre diretti verso l'avanti, il flusso si espande verso l'esterno a partire dal punto centrale del campo visivo. I neuroni dell'area MST posseggono campi recettivi che coprono gran parte del campo visivo e rispondono in maniera preferenziale a movimenti che interessano parti molto estese del campo visivo stesso. Questi neuroni sono molto sensibili allo spostamento del punto dal quale prende origine il movimento globale del campo visivo, nonché alle differenze di velocità che si manifestano fra il centro e la periferia del campo visivo. I neuroni di quest'area ricevono anche afferenze correlate ai movimenti oculari e ciò appare particolarmente interessante in quanto il movimento verso l'avanti si accompagna, in generale, a movimenti degli occhi e del capo.

2.5 Apprendimento visivo

Riassumendo il paragrafo precedente, il secondo livello di elaborazione visiva è diviso in due vie. La via ventrale, che è coinvolta nel riconoscimento degli oggetti, comincia nella corteccia striata e continua ventralmente nella corteccia temporale inferiore ([29] capitolo 28). La via dorsale che elabora la percezione della localizzazione degli oggetti, comincia anch'essa nella corteccia striata del lobo occipitale, ma continua dorsalmente nella corteccia parietale posteriore. Si può dunque affermare che la via ventrale è coinvolta nel *cosa* della percezione

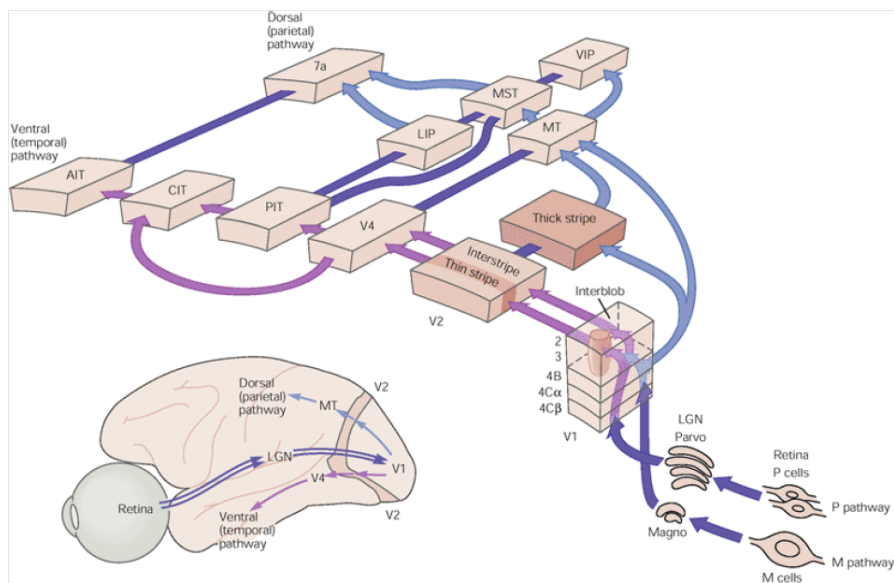


Figura 2.7: Percorsi magnocellulari (M) e povicellari (P). I due percorsi restano sempre separati e dalla corteccia visiva primaria (V1) procedono verso le aree dorsali (M) e temporali (P). In particolare dalla V1 le informazioni che riguardano il movimento attraversano l'area mediotemporale (MT) e l'area mediotemporale superiore (MST).

visiva; la via dorsale con il *dove* (figura 2.5). Per quanto riguarda il riconoscimento di un particolare pattern visivo, è stato osservato che l'apprendimento si realizza mediante cambiamenti delle connessioni sinaptiche nella corteccia temporale inferiore, che stabiliscono nuovi circuiti neurali. In seguito, quando viene rivisto lo stesso stimolo e lo stesso pattern di attività viene trasmesso alla corteccia temporale inferiore, questi circuiti si riattivano permettendo il riconoscimento dello stimolo (lettura della memoria visiva). L'apprendimento percettivo può avvenire molto rapidamente e il numero di elementi che può essere ricordato è enorme. Studi specifici di immagine funzionale hanno mostrato che il recupero delle memoria percettive riguardanti l'identità e la localizzazione degli oggetti comportano, rispettivamente, attività nella via ventrale e in quella dorsale del cervello umano. La capacità di memorizzazione delle informazioni visive si estende anche in compiti che richiedono un confronto tra lo stimolo ricordato e uno stimolo che stiamo percependo nell'istante attuale. In altre parole, in certe situazioni occorre confrontare una percezione con la memoria a breve termine di qualcosa che abbiamo appena percepito. Molti studi sulla memoria a breve termine hanno impiegato un compito di confronto differito con un modello che chiede ad un animale di ricordare uno stimolo particolare per un periodo di tempo. E' stato mostrato che i neuroni nel lobo parietale posteriore, nella via dorsale, conservano l'informazione relativa alla localizzazione spaziale di uno stimolo visivo che è stato appena percepito. La memoria a breve termine visiva coinvolge anche altre regioni cerebrali, specialmente la corteccia prefrontale. Entrambe le regioni principali della corteccia associativa visiva, la via ventrale della corteccia temporale inferiore e la via dorsale della corteccia parietale posteriore, hanno connessioni reciproche dirette con la corteccia prefrontale. Nel cervello di scimmia, la via dorsale ha connessioni con la regione posta all'interno e immediatamente dorsale rispetto al *solco principale* e la via ventrale ha connessioni con le regioni ventrali di questo solco [30]. La corteccia prefrontale è coinvolta nei compiti di memoria a breve termine per tutte le modalità sensoriali. Molti neuroni in questa regione rispondono selettivamente a stimoli visivi specifici e presentano un'attività sostenuta durante il periodo di attesa del compito di confronto con un modello. Da quanto detto, sembrerebbe dispendioso per il cervello conservare informazioni identiche sugli stimoli appena percepiti in due zone diverse del cervello (corteccia parietale posteriore e corteccia prefrontale). Miller, Erickson e Desimone mostrarono che durante la presentazione di uno stimolo visivo (test), che doveva essere confrontato con un modello appena appreso, si attivavano entrambe le aree [31]. L'attività nella corteccia associativa visiva che rappresenta lo stimolo scompariva quando veniva tolto il test, mentre l'attività nella corteccia prefrontale continuava. Tale attività continuava finché gli sperimentatori presentavano uno stimolo test che corrispondeva allo stimolo modello. Quindi, l'attività nella corteccia prefrontale è in grado, presumibilmente, di conservare il ricordo a breve termine dello stimolo modello, lasciando la corteccia associativa visiva libera di occuparsi della percezione visiva. Altri esperimenti sulla corteccia prefrontale codificano anche informazioni che riguardano la risposta che l'animale sta per emettere e non semplicemente lo stimolo che ha visto. L'associazione tra uno stimolo sensori-

ale e una risposta è detto *condizionamento operante* ed è il mezzo mediante il quale noi (e altri animali) traiamo profitto dall'esperienza. Se, in una particolare situazione, emettiamo una risposta che produce esiti favorevoli, tenderemo a produrre nuovamente quella risposta. Se quell'azione non è stata mai eseguita in precedenza, la prestazione sarà lenta ed impacciata, mentre continuando a produrre la risposta, il comportamento diventerà più veloce, fluido ed automatico. Questo viene definito apprendimento motorio. I circuiti responsabili del condizionamento operante hanno origine in diverse regioni della corteccia associativa sensoriale, dove ha luogo la percezione, e terminano nella corteccia associativa del lobo frontale, che controlla i movimenti. Ci sono due vie principali tra la corteccia associativa sensoriale e la corteccia associativa motoria: connessioni transcorticali dirette e connessioni che passano dai gangli della base e dal talamo. Si ritiene che le connessioni dirette siano coinvolte nei processi di memoria a breve termine. Insieme alla formazione ippocampale, esse sono anche implicate nell'acquisizione di memorie episodiche, memorie percettive complesse di sequenze di eventi. Alcune prove suggeriscono che, quando i comportamenti appresi diventano automatici e di routine, sono trasferiti ai gangli della base. Dopo diverse iterazioni dello stesso comportamento, essi si fanno carico della maggior parte dei dettagli del processo, lasciando i circuiti transcorticali liberi di fare altro. Studi su animali o su esseri umani indicano che in gangli basali giocano un ruolo importante nell'apprendimento automatico, non volontario. I ricercatori hanno studiato persone affette dal morbo di Parkinson, un disturbo neurologico che colpisce i gangli della base. Nel passato i sintomi del morbo di Parkinson sono stati descritti come deficit motori. Tuttavia, alcuni di essi possono essere visti come fallimenti della memoria automatica. Infatti, tali pazienti incontrano difficoltà nello svolgere molti compiti quotidiani, come alzarsi dalla sedia. Anche questa semplice azione richiede un apprendimento precedente su come spostare il baricentro sui piedi prima di iniziare a contrarre i muscoli estensori delle gambe. Diversi esperimenti hanno mostrato che le persone con il morbo di Parkinson presentano deficit attribuibili a difficoltà nell'apprendimento di risposte automatiche come compiti di discriminazione visiva.

Alla luce di quanto descritto è possibile presumere che le informazioni sensoriali vengano elaborate da zone specifiche della corteccia e continuino nella corteccia associativa del lobo frontale che permette le connessioni con l'area premotoria per il controllo del movimento. Movimenti appresi di natura automatica vengono anche elaborati dai gangli della base per permettere movimenti più fluidi in base alle risposte precedentemente ottenute. Nello specifico di questa tesi, le informazioni visive di movimento vengono elaborate nei vari strati della corteccia visiva primaria, attraversano le aree MT e MST deputate al riconoscimento del tipo di movimento presente nel campo visivo e terminano nella corteccia parietale posteriore che gestisce la memoria visiva a breve termine. Tale informazione viene successivamente trasmessa alla corteccia prefrontale che mantiene il ricordo dell'informazione visiva anche quando lo stimolo non è più presente per poi associare tale rappresentazione alla sedi coinvolte nel controllo del movimento.

2.6 Controllo dello sguardo

L'importanza del sistema che regola la direzione dello sguardo scaturisce dall'anatomia dell'occhio. Sebbene l'occhio sia in grado di rilevare la presenza di un oggetto davanti a noi dovunque si trovi, siamo tuttavia in grado di vedere gli oggetti in maniera ottimale soltanto con la *fovea*, che è l'area specializzata disposta al centro della retina con un diametro inferiore ad 1 mm e in grado di ricevere informazioni da una piccola parte del campo visivo ([29] capitolo 39). Quando vogliamo esaminare un oggetto presente nel mondo esterno, dobbiamo indirizzare la fovea verso di esso. Il *sistema dello sguardo* svolge questa funzione attraverso due componenti: il *sistema oculomotore*, che fa ruotare gli occhi all'interno delle orbite, e il *sistema del movimento del capo*, che fa muovere le orbite nello spazio. Il sistema del *movimento lento di inseguimento* (*smooth pursuit*) è uno dei sistemi neuronali di controllo che mantengono la fovea sul bersaglio visivo. Esso è in grado, dunque, di calcolare la velocità di movimento dell'oggetto e permette all'occhio di muoversi alla medesima velocità. Per poter determinare con esattezza la velocità del movimento oculare questo sistema richiede la presenza di uno stimolo in movimento. Di conseguenza, un comando verbale o l'immaginazione di uno stimolo non sono in grado di generare movimenti lenti di inseguimento. Essi possono raggiungere una velocità massima di $80^\circ/\text{s}$, molto più bassa dei movimenti oculari saccadici che permettono velocità fino a $900^\circ/\text{s}$. I movimenti oculari consistono nella rotazione del globo oculare all'interno della cavità dell'orbita. L'orientamento dell'occhio può essere definito facendo riferimento a tre assi di rotazione, orizzontale verticale e torsionale, che si intersecano nel centro del globo oculare e i movimenti oculari possono essere descritti come rotazioni attorno a questi assi. L'*abduzione* è un movimento di rotazione in direzione opposta al naso e l'*adduzione* un movimento di rotazione verso il naso. L'*elevazione* consiste nella rotazione verticale verso l'alto e l'*abbassamento* nella rotazione verticale verso il basso. I movimenti torsionali non modificano la linea dello sguardo, ma fanno ruotare l'occhio attorno ad essa. La muscolatura estrinseca dell'occhio è formata da tre coppie complementari di muscoli ed in tutti i movimenti oculari, ad eccezione di quelli di vergenza, i due occhi lavorano appaiati. Ciò comporta che ad ogni coppia di muscoli di un occhio corrisponda una coppia funzionale di muscoli dell'altro occhio che possa far ruotare un occhio nello stesso piano ma in direzione opposta rispetto al movimento compiuto dall'altro. I muscoli oculari estrinseci sono innervati da tre gruppi di motoneuroni i cui corpi cellulari sono situati in tre nuclei del tronco dell'encefalo. Esaminando l'attività di un motoneurone che innerva un muscolo estrinseco si osserva una frequenza di scarica direttamente proporzionale alla posizione dell'occhio ed alla sua velocità. Quando la velocità dell'occhio passa da $0^\circ/\text{s}$ a $900^\circ/\text{s}$, la frequenza di scarica del motoneurone aumenta rapidamente e quest'aumento di attività, che è di tipo impulsivo è detto *impulso* (figura 2.8). Una volta che l'occhio ha raggiunto la nuova posizione, i muscoli oculari estrinseci presentano uno stato di contrazione tonica. La differenza fra la frequenza di scarica iniziale e finale viene denominata *gradino*. Il segnale emesso dai neuroni oculomotori quando viene generato una saccadico è perciò impulso-gradino. L'altezza del gradino

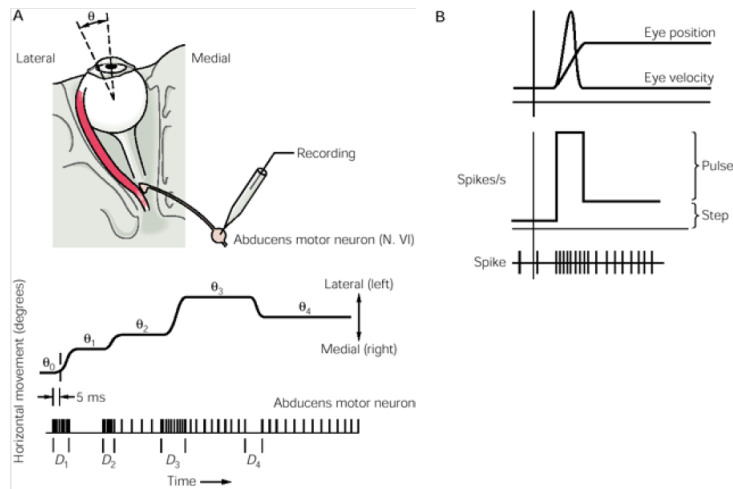


Figura 2.8: Segnali dei motoneuroni per la posizione e la velocità dell'occhio. In A e B si osserva la relazione tra la scarica del motoneurone abducente in una scimmia. Scariche di tipo burst (D) permettono il cambio della posizione dell'occhio e per mantenere tale angolo occorre una scarica di tipo treno di impulsi tonici di frequenza proporzionale alla posizione che si vuole mantenere.

determina l'ampiezza del saccadico, mentre l'altezza dell'impulso determina la velocità.

La funzione del sistema del movimento lento di inseguimento è diversa da quella del sistema saccadico. Invece di portare il più rapidamente possibile gli occhi verso un punto nello spazio, questo sistema deve adeguare la velocità del movimento dell'occhio a quella di un bersaglio presente nello spazio visivo. Neuron che segnalano la velocità dell'occhio sono stati identificati a livello dei nuclei vestibolare mediale e prosito dell'ipoglosso e ricevono afferenze dal foccolo del cervelletto (figura 2.9). Anche neuroni della formazione reticolare pontina parmediana ritrasmettono segnali per il movimento di inseguimento e ricevono informazioni dal verme del cervelletto. Sia i neuroni del verme che quelli del foccolo ritrasmettono segnali relativi alla velocità dell'occhio che sono correlati con i movimenti lenti di inseguimento. Queste regioni del cervelletto ricevono segnali dalla corteccia cerebrale ritrasmessi attraverso il nucleo pontino dorso-laterale. Le principali afferenze corticali per il sistema di inseguimento lento, sono due. Una è costituita dalle regioni del solco temporale superiore e dalle aree temporale mediale (MT) e temporale superiore mediale (MST). L'altra nasce dai campi oculari frontali. Sia a livello della MT che della MST sono stati identificati neuroni che calcolano la velocità dei bersagli visivi. Quando l'occhio si muove per adeguare la propria velocità a quella di un bersaglio in movimento, la velocità relativa dell'immagine del bersaglio sulla retina diminuisce. I neuroni di MT, che codificano il movimento dell'immagine sulla retina, cessano di scaricare quando la velocità di movimento dell'immagine retinica diminuisce,

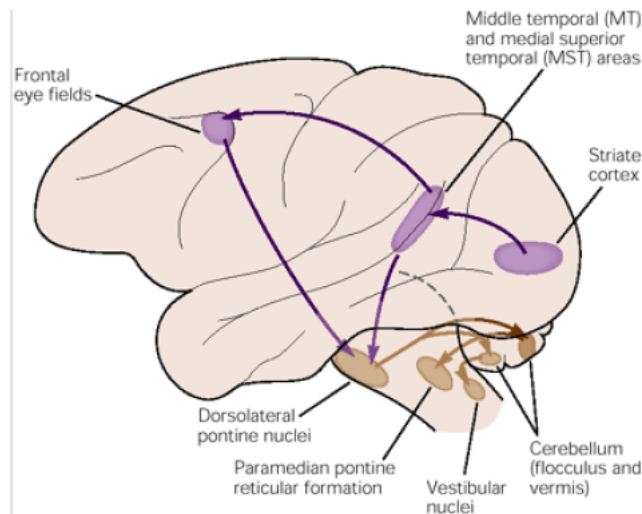


Figura 2.9: Percorso corticale del movimento di smooth pursuit nelle scimmie. La corteccia cerebrale processa le informazioni relative al movimento che vengono inviate ai motoneuroni attraverso il ponte dorsolaterale, il verme e il foccolo del cervelletto, i nuclei vestibolari e la formazione reticolare pontina paramediana. L'avvio del segnale di inseguimento parte dai campi oculari frontali.

anche se il bersaglio continua a muoversi nello spazio. Invece i neuroni di MST continuano a scaricare e rimangono attivi anche se il bersaglio scompare per un breve lasso di tempo. Questi neuroni mettono in azione un processo che somma le velocità del movimento dell'occhio e del bersaglio sulla retina per calcolare la velocità del bersaglio stesso nello spazio. Tali aree forniscono informazioni sensoriali necessarie per guidare i movimenti di inseguimento, ma non sono in grado di provvedere alla loro esecuzione. La stimolazione elettrica di MT e MST non determina l'inizio del movimento ma influisce sulla loro esecuzione. I campi oculari frontali hanno probabilmente maggiore importanza per l'inizio dei movimenti di inseguimento.

Capitolo 3

Modello proposto di movimento oculare di inseguimento

3.1 Introduzione

Un modello è una rappresentazione di un fenomeno, che ne riproduce alcune caratteristiche o comportamenti fondamentali, in modo tale che questi aspetti possano essere mostrati, studiati, conosciuti laddove l'oggetto modellato non sia direttamente accessibile. La costituzione di un modello scientifico o tecnico è sempre il risultato di un contesto della prova rigoroso, predisposto in modo tale da non essere minimamente influenzato dalle aspettative e dall'interpretazione soggettiva dell'osservatore. Il grado di verosimiglianza tra il fenomeno studiato e il modello elaborato dipende dalle caratteristiche che si vuole ottenere e studiare e, di solito, è direttamente proporzionale al grado di complessità del modello stesso. Per questo motivo i primi modelli che caratterizzano un fenomeno sono spesso semplici e poco complessi, ma per emulare la complessità funzionale degli apparati biologici, essi necessitano di continue aggiunte in modo tale che sia possibile ampliarne le caratteristiche o migliorarne l'integrabilità con altri modelli di apparati vicini. I modelli biologici necessitano, infatti, di una visione olistica, che è basata sull'idea che le proprietà di un sistema non possano essere spiegate esclusivamente tramite le sue componenti. La complessità di un organismo biologico deriva proprio dal fatto che le sue parti sono mutualmente dipendenti. Il modello del movimento oculare di inseguimento lento (*Smooth Pursuit*) è stato ampiamente trattato in letteratura, ma in questo lavoro si sono voluti evidenziare i vantaggi derivanti dalla modellizzazione del fenomeno nelle caratteristiche di predizione e apprendimento.

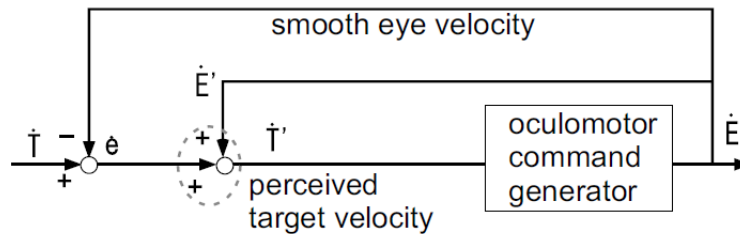


Figura 3.1: Modello di Yasui e Young in cui la velocità del target viene calcolata in base allo slip retinale e alla velocità dell'occhio.

3.2 Modello di inseguimento classico

La modellizzazione del movimento oculare di inseguimento lento (Smooth Pursuit) è stata oggetto di numerosi studi sia al fine della conoscenza biologica sia per le possibili applicazioni tecnologiche. Nel precedente capitolo è stato mostrato che tale movimento riguarda numerose aree cerebrali che interpretano e gestiscono le informazioni provenienti dalla retina per controllare in modo accurato l'apparato muscolare oculomotore. La comprensione del fenomeno da un punto di vista modellistico aiuta la localizzazione e l'identificazione funzionale delle aree cerebrali coinvolte. Occorre quindi comprendere la problematica tenendo conto del sistema sul quale il fenomeno si svolge, comprendendone i limiti e le possibilità. Il modello più semplice ed intuitivo di inseguimento di un target in movimento prevede l'estrazione delle coordinate dell'oggetto e lo spostamento del sistema di visione, cercando di annullare la differenza relativa tra l'oggetto e il centro dell'immagine. Tale modello risulta essere sempre in ritardo rispetto al target e può funzionare discretamente solo su sistemi con tempi di acquisizione molto rapidi rispetto alla dinamica del target. La frequenza di aggiornamento dell'immagine sulla retina è, invece, piuttosto bassa, intorno ai 10 Hz. L'illusione del movimento è opera del cervello il quale assembla le immagini che gli vengono trasmesse in modo unitario e, ciò che più conta, in movimento. Yasui e Young [32] hanno proposto un primo modello di smooth pursuit nel quale il diagramma di controllo è sviluppato nello spazio delle velocità (figura 3.1), nell'assunzione che esista nel cervello umano, un adeguato controllore che realizzi perfettamente la velocità desiderata. L'idea di Yasui e Young è che i primati possano generare una stima \dot{T}' della velocità del target \dot{T} sommando lo slip retinale \dot{e} (*slip retinale*), cioè lo spostamento dell'oggetto sulla retina, e il segnale di ritorno propriocettivo della velocità dell'occhio \dot{E}' :

$$\dot{e} + \dot{E}' = \dot{T}' \simeq \dot{T} \quad (3.1)$$

Si consideri l'errore retinale e come l'errore di posizione sull'immagine, mentre lo slip retinale \dot{e} , è l'errore di velocità sull'immagine. L'importanza di questo

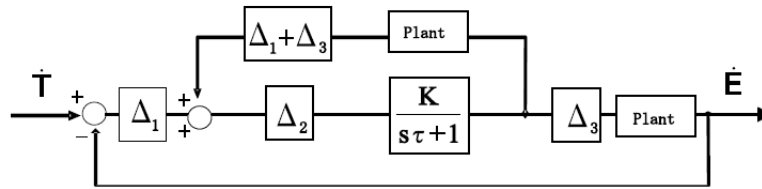


Figura 3.2: Il modello feed forward di Robinson in cui sono cancellati i ritardi dovuti al processamento delle informazioni Δ_1 e al movimento Δ_3 . Il modello però non considera il ritardo Δ_2 e per questo risulta sempre in ritardo.

modello risiede nel fatto che viene stimata la velocità del target da variabili sensoriali come lo slip retinale e la posizione dell'occhio. Ma tale approccio non tiene conto che \dot{E} al tempo t porta le informazioni generate da $\dot{T}(t - \Delta)$ e $\dot{E}(t - \Delta)$. Per prima cosa, dunque, occorre tener conto che è possibile stimare solo $\dot{T}(t - \Delta)$ e questa stima richiede $\dot{E}(t - \Delta)$ e non $\dot{E}(t)$. Inoltre in modo di seguire accuratamente un oggetto in movimento con una dinamica arbitraria, è necessario predire $\dot{T}(t)$ al tempo t da una storia degli stati del target per ovviare ai ritardi dovuti al processamento delle immagini.

Un miglioramento sostanziale è stato presentato da Robinson et al. [33] nel quale viene cancellato il segnale di feedback negativo in modo da permettere un alto guadagno di velocità (figura 3.2). Tale modello lavora come un controllore *feed-forward*: il segnale di feedback con un ritardo Δ_1 viene cancellato da un loop a feedback positivo con ritardo $\Delta_1 + \Delta_3$. Tale percorso contiene un ritardo significativo determinato da Δ_2 e τ in modo tale che questo modello non può ottenere un inseguimento senza perdita di un target sinusoidale. Un altro importante lavoro è stato eseguito da Bradshaw et al. [34], che implementa un filtro di Kalman per eseguire la predizione della velocità del target. Tale idea si basa, però, su conoscenze a priori della dinamica del target e quindi non fornisce indicazioni su come inseguire accuratamente una dinamica sconosciuta. Il filtro di Kalman è un efficiente filtro ricorsivo che valuta lo stato di un sistema dinamico a partire da una serie di misure soggette a rumore. E' utilizzato per il filtraggio dei dati sulla base di una media ragionata tra il prossimo valore predetto e il prossimo valore stimato. Questo filtro è spesso utilizzato per ottenere una migliore valutazione di un dato ottenuto dalla lettura di più sensori, ognuno caratterizzato da un rumore di misura avente caratteristiche differenti nel tempo. Il filtro ha bisogno però della conoscenza di un modello matematico lineare descrittivo del sistema in modo da stimarne i coefficienti a partire dalle diverse misure del processo. Nel caso in esame, il filtro adatta i coefficienti della dinamica del target nota in modo da predire correttamente lo stato del sistema anche in presenza di rumore. Generalmente quindi occorre stabilire a priori una serie di possibili dinamiche e, rispetto alla dinamica presentata, scegliere quella che meglio assicura la convergenza dell'algoritmo. E' lecito supporre che la predizione che il cervello compie sulla dinamica del target derivi esclusivamente

di dinamica inversa del sistema oculomotore presente nel cervelletto e nel tronco encefalico, permette agli occhi di seguire la velocità predetta. Il cervello non può osservare lo stato del target direttamente, quindi esso stima lo stato basandosi sull'informazione di slip retinale ritardata di un tempo Δ . Nell'equazione (3.2) viene rappresentata la predizione della velocità corrente del target $\hat{X}(t)$ da una stima dello stato del target passato $\bar{X}(t - \Delta)$ attraverso la funzione f :

$$\hat{X}(t) = f(\bar{X}(t - \Delta)) \quad (3.2)$$

E' dunque necessario applicare ricorsivamente la funzione f per predire continuamente la dinamica del target a partire dalla conoscenza dello stato del target ritardato. Si assume che il controllore inverso del sistema oculomotore sia in grado di allineare la velocità degli occhi alla velocità predetta del target:

$$\dot{E}(t) = \hat{X}(t) \quad (3.3)$$

Lo stimatore dello stato $\bar{X}(t - \Delta)$ nell'equazione (3.2) deriva da:

$$\bar{X}(t - \Delta) = \hat{X}(t - \Delta) + K(X(t - \Delta) - \hat{X}(t - \Delta)) \quad (3.4)$$

Il primo termine a destra in questa equazione è lo stato stimato al tempo $(t - \Delta)$, mentre il secondo termine è l'errore di predizione moltiplicato per una matrice di pesi K , che corregge lo stato predetto dalla media dei segnali passati. Lo stato $X(t - \Delta)$ non è direttamente osservabile nell'equazione (3.4), ma componendo la (3.3) con la (3.1) si ottiene:

$$\bar{X}(t - \Delta) = \hat{X}(t - \Delta) + K(X(t - \Delta) - E(t - \Delta)) = \hat{X}(t - \Delta) + K\dot{e}(t - \Delta) \quad (3.5)$$

E' così possibile stimare lo stato correttamente basandosi sul segnale di slip retinale ritardato.

Al tempo t , il predittore può solo vedere lo stato stimato del target ritardato $\bar{X}(t - \Delta)$. Attraverso il vettore dei parametri w nel predittore dello stato espresso nella (3.2), la predizione della velocità del target diventa:

$$\hat{X}(t) = f(\bar{X}(t - \Delta), w(t)) \equiv f(t) \quad (3.6)$$

Il vettore dei parametri w deve essere regolato dal sistema. In accordo con le conoscenze biologiche riguardo all'area MST, il predittore predice esclusivamente la velocità del target invece che il completo stato del target. Sia la funzione di perdita J un semplice errore quadratico di predizione:

$$J(t) = \frac{1}{2}\dot{e}(t)^2 = \frac{1}{2}(\dot{X}(t) - \dot{E}(t))^2 \quad (3.7)$$

Unendo le equazioni (3.3) e (3.6) è possibile scrivere la legge di apprendimento della discesa del gradiente per il vettore dei parametri w come:

$$\frac{dw_i(t)}{dt} = -\varepsilon \frac{\partial J(t)}{\partial w_i} = \varepsilon \frac{\partial f(t)}{\partial w_i} \dot{e}(t) \quad (3.8)$$

in cui ε è la velocità di apprendimento. Con lo scopo di permettere al modello di seguire una dinamica di tipo sinusoidale si noti che essa corrisponde ad un sistema del secondo ordine che garantisce che la velocità del target possa essere predetta dalla conoscenza dello stato passato del target, quindi posizione x e velocità \dot{x} . Occorre considerare che il predittore non può ottenere $\dot{e}(t)$ al tempo t , quindi occorre spostare tutta l'equazione (3.8) un Δ nel passato.

$$\frac{dw_i(t)}{dt} = \varepsilon \frac{\partial f(t - \Delta)}{\partial w_i} \dot{e}(t - \Delta) \quad (3.9)$$

L'apprendimento è eseguito implementando l'algoritmo ricorsivo ai minimi quadrati (RLS Recursive Least Squares) che garantisce una convergenza veloce. Originariamente l'RLS è formulato come in (3.10, 3.11) dove w è il vettore dei parametri di regressione che devono essere stimati, P è la matrice di covarianza inversa dei dati di input, x è il vettore di input, y è l'uscita, e \hat{y} è l'uscita predetta.

$$P(t) = \frac{1}{\lambda} \left[P(t - \Delta) - \frac{P(t - \Delta)x(t)x(t)^T P(t - \Delta)}{\lambda + x(t)^T P(t - \Delta)x(t)} \right] \quad (3.10)$$

$$w(t) = w(t - \Delta) + \frac{P(t)x(t)}{\lambda + x(t)^T P(t)x(t)} (y(t) - \hat{y}(t)) \quad (3.11)$$

$$\hat{y}(t) = w(t)^T x(t) \quad (3.12)$$

Nella (3.11) si mostra che il normale RLS richiede la presenza dell'uscita del target y per la legge di aggiornamento, ma come descritto precedentemente, il predittore non può vedere l'uscita in modo diretto, ma può utilizzare lo slip retinale come errore di predizione. In questo modo l'equazione deve essere modificata in:

$$w(t) = w(t - \Delta) + \frac{P(t)x(t)}{\lambda + x(t)^T P(t)x(t)} \dot{e}(t + 1) \quad (3.13)$$

Questa strategia corrisponde ad addestrare l'RLS su target falsi, cioè $y = \hat{y}(t) + \dot{e}$. Inizialmente, questi target falsi sono piuttosto distanti da quelli reali, così che l'RLS è addestrato con dati incorretti. Per questo motivo, occorre dimenticare i dati iniziali utilizzando un fattore di dimenticanza λ appartenente all'intervallo compreso tra $[0, 1]$. Per $\lambda = 1$, non avviene dimenticanza, mentre per piccoli valori, i precedenti valori della matrice P sono dimenticati con andamento esponenziale. Essenzialmente, il fattore di dimenticanza assicura che la predizione dell'RLS sia basata esclusivamente sui $1/(1 - \lambda)$ valori. Questa strategia, inoltre, permette al predittore di essere adattivo con i cambiamenti della dinamica del target. Un altro importante elemento dell'equazione (3.13) è che essa mostra la necessità dell'allineamento temporale tra l'uscita del predittore e l'errore, ma il predittore non può vedere $\dot{e}(t + 1)$ all'istante t . Così, come mostrato anche nella (3.9), tutte le variabili della (3.13) devono essere ritardate di un passo temporale e questo richiede la loro memorizzazione per un breve tempo.

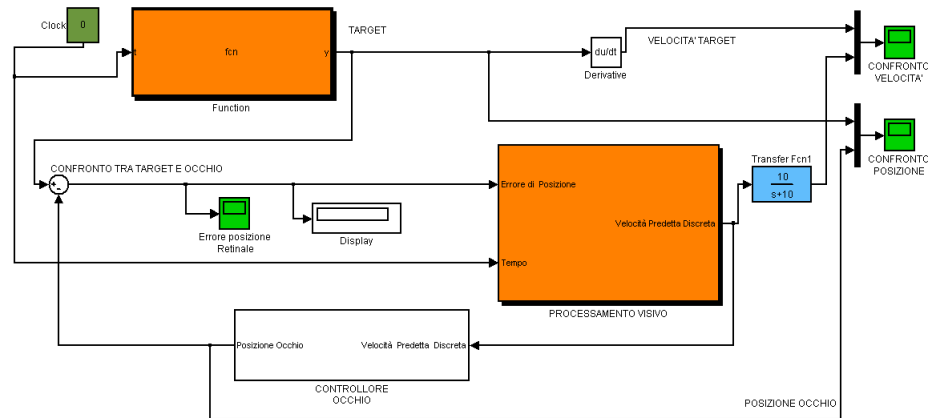


Figura 3.4: Modello di Schaal implementato in Matlab-Simulink. *Function* genera il segnale da seguire che corrisponde alla posizione del target, il *Confronto tra target e occhio* corrisponde alla retina che sottrae alla posizione del target quella dell'occhio, in uscita si ottiene l'errore di posizione; il *processamento visivo* restituisce la velocità del target predetta (in accordo con gli studi biologici sull'area MST) che viene inviata al *controllore dell'occhio* il quale modifica la posizione.

3.4 Simulazione in Matlab Simulink

Il modello descritto è stato simulato su Matlab Simulink con ingresso di tipo sinusoidale realizzato tramite il blocco *Function* (figura 3.4). Il ritardo di elaborazione visiva, (Δ mostrato nelle equazioni precedenti) è stato assunto dello stesso ordine di grandezza della persistenza dell'immagine sulla retina umana cioè 100 ms. Il modello è così rappresentato come un sistema tempo discreto che campiona il segnale di ingresso con frequenza pari a 10 Hz. Per il teorema del campionamento, questo implica che il sistema potrà seguire correttamente solo segnali sinusoidali con frequenza strettamente minore di 5 Hz, il che è in accordo con dati sperimentali su soggetti umani.

In figura 3.4 viene mostrato la prima parte del modello nel quale il nodo sommatore rappresenta essenzialmente la retina, in quanto confronta la posizione del target con quella dell'occhio. Questa informazione è inviata al sistema di *processamento visivo* che restituisce un segnale di velocità (in questo caso un segnale tempo discreto) che viene inviato al *controllore dell'occhio* per aggiornare la posizione dello sguardo. La velocità del target è poi confrontata con la velocità in uscita dal predittore per mostrare che l'anticipo di 200 ms in accordo con i dati presenti in letteratura. In figura 3.5 è mostrato il dettaglio del *processamento visivo*. Qui l'errore di posizione è campionato a frequenza di 10 Hz. Il campionatore è ottenuto con un blocco *Unit Delay* che aggiunge un ritardo di 100 ms come richiesto da modello. L'operazione di derivata discreta ottenuta attraverso il blocco *Difference* permette di ottenere lo slip retinale. A questo

CAPITOLO 3. MODELLO PROPOSTO DI MOVIMENTO OCULARE DI INSEGUIMENTO 42

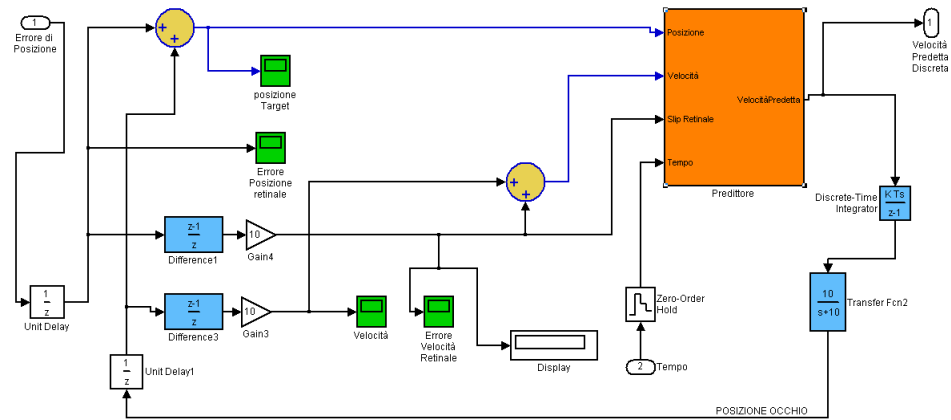


Figura 3.5: All'interno del *processamento visivo* si esegue il campionamento a frequenza 10 Hz dell'errore di posizione aggiungendo un ritardo di 100 ms. L'operazione di derivata discreta fornisce lo slip retinale che viene usato come errore nel *predittore*. La somma tra errore di posizione e la posizione occhio, nel nodo sommatore in alto, fornisce la posizione del target tempo discreto. La somma tra slip retinale e velocità dell'occhio, nel nodo sommatore in basso, fornisce la velocità del target tempo discreto. Queste due variabili rappresentano lo stato del target.

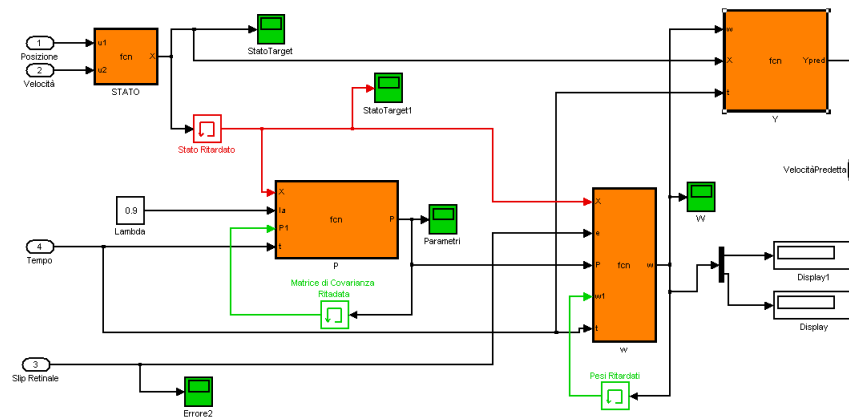


Figura 3.6: Per il *predittore* è stato implementato l'algoritmo RLS secondo le equazioni (3.10, 3.12, 3.13) descritte nel testo. Il fattore di dimenticanza Lambda è stato scelto pari a 0.9. Lo stato del target deve essere ritardato (in rosso) per permettere l'addestramento con lo slip retinale non ritardato. All'interno dei blocchi vengono calcolati i valori aggiornati della matrice P e del vettore dei parametri di regressione w . Il blocco Y calcola la velocità predetta moltiplicando lo stato X per il vettore w .

punto sommando l'errore di posizione e lo slip retinale rispettivamente alla posizione dell'occhio e alla velocità dell'occhio si ottiene lo stato del target (come mostrato dalla equazione 3.1) che raggiunge il *predittore*. Si noti che, come da modello, l'informazione sullo stato dell'occhio (posizione e velocità) vengono ottenute direttamente dall'uscita del predittore e ritardate della solita quantità di tempo (100 ms). All'interno del predittore è riprodotto l'algoritmo RLS ottenuto implementando in codice Matlab le equazioni 3.10, 3.12 e 3.13 per le quali è stato necessario utilizzare il blocco *Memory* per ricordare il valore della variabile ottenuto nella precedente iterazione. La matrice P è inizializzata con una matrice diagonale con valori $[1 \ 0; 0 \ 1]$, mentre tutti gli stati iniziali del sistema di apprendimento sono settati a zero. Nelle figure 3.7 e 3.8 sono mostrati i risultati per una sinusoide con pulsazione pari a 1 rad/sec e con fattore di dimenticanza $\lambda = 0.9$. La velocità dell'occhio segue perfettamente la dinamica del target sinusoidale (figura 3.8), lo slip retinale raggiunge lo zero entro 5 secondi. In accordo con gli studi biologici, il modello predice esclusivamente la velocità del target e di conseguenza non annulla l'errore di posizione (figura 3.7 in basso) che resta dunque costante. In figura 3.9 sono riportati i grafici del vettore dei parametri di regressione w . Mentre in figura 3.10 è possibile notare dal confronto della velocità predetta in uscita dall'algoritmo e la velocità del target che, dopo una determinato transitorio, la prima anticipa la seconda. Per mostrare il grado di anticipazione del modello è stata eseguita una prova in cui ad un determinato istante (es. 40 sec) la posizione del target viene portata a zero. In figura 3.11 in alto, è possibile vedere il confronto della posizione del target e dell'occhio, mentre nel dettaglio in figura 3.11 in basso, è possibile notare che la posizione dell'occhio risponde alla brusca variazione della dinamica del target solo dopo 200 ms in accordo con i risultati riportati in letteratura. Questo ritardo è dovuto al fatto che il modello predice la traiettoria del target al fine di annullare la differenza di velocità, superando il ritardo dovuto alla conduzione e al processamento dell'immagine (rappresentato dal blocco *Time Delay* che campiona l'errore di posizione) e il ritardo dovuto alla risposta motoria (questo è rappresentato dal *controllore dell'occhio* che prende in ingresso la velocità predetta discreta e restituisce la posizione dell'occhio tempo continuo). Nel modello in figura 3.3 si osserva che l'area MST necessita di confrontare i segnali provenienti dalla retina con la predizione precedente che dunque dovrà essere opportunamente ritardata (ritardo rappresentato dal blocco *Time Delay* che riporta il segnale in ingresso). Altra interessante prova (figura 3.12) è stata eseguita con una dinamica composta da una sinusoide smorzata:

$$y = A * \sin(k1 * t) * e^{-t*0.2} \quad (3.14)$$

Con y posizione del target, $A = 1$ ampiezza della sinusoide (in radianti), $k1 = 1$ pulsazione (rad/sec) e t tempo. Tale dinamica potrebbe corrispondere all'andamento di un pendolo smorzato per effetto della gravità. Il modello continua a seguire la velocità del target ed i valori dei parametri risultano essere pari a $w(1) = -0.3352$ e $w(2) = 0.831$.

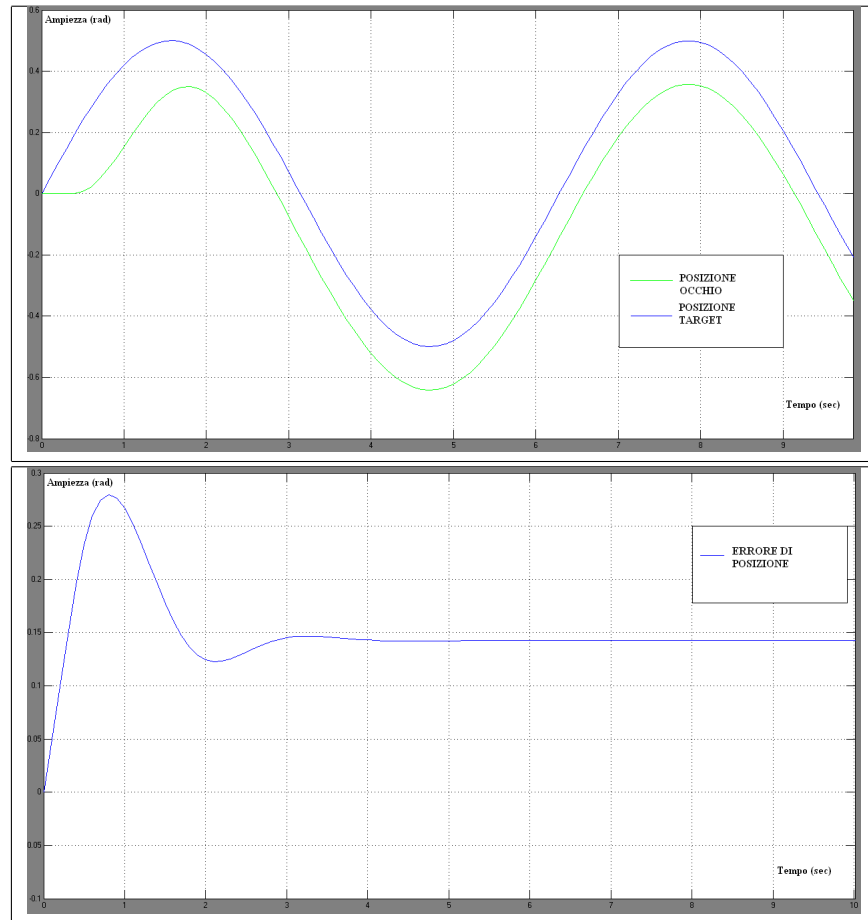


Figura 3.7: In figura sono mostrati i primi 10 secondi della posizione del target confrontata con la posizione dell'occhio (in alto) con dinamica sinusoidale con pulsazione pari a 1 rad/sec e $\lambda = 0.9$. L'errore di posizione (in basso) risulta costante e di valore 0.1425 entro i 5 secondi. Il modello non annulla l'errore di posizione.

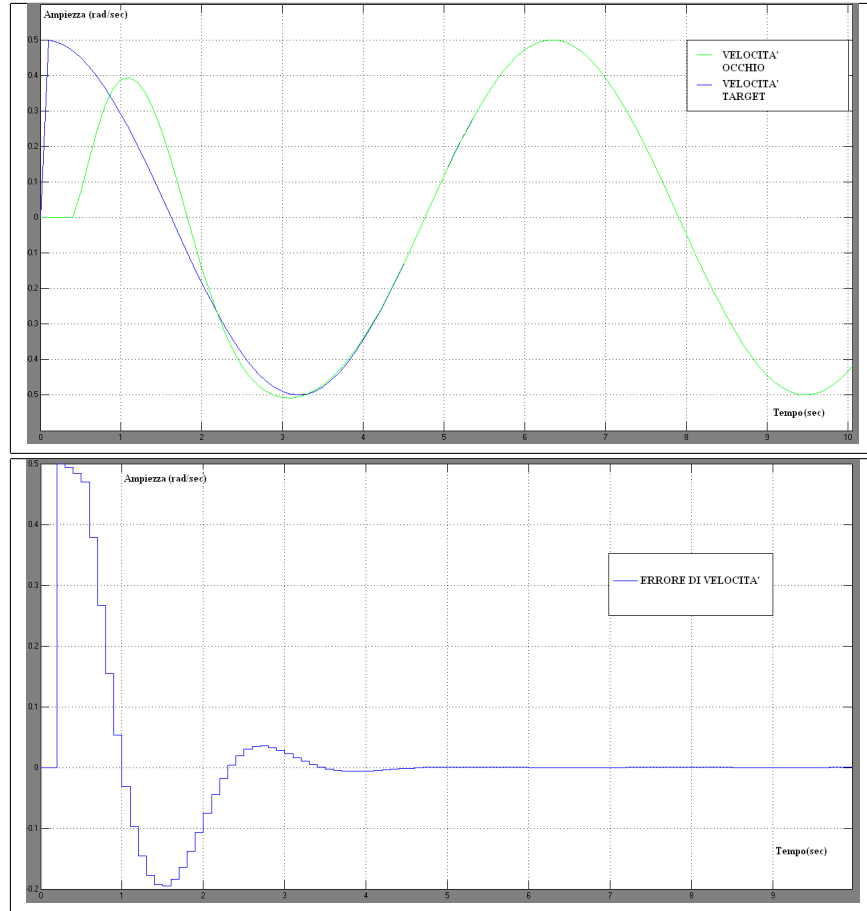


Figura 3.8: Confronto dei primi 10 secondi della velocità del target con la velocità dell'occhio (in alto) con dinamica sinusoidale con pulsazione di 1 rad/sec e $\lambda = 0.9$. Il modello segue perfettamente la dinamica del target. Lo slip retinale tempo discreto (in basso) si annulla entro 5 secondi.

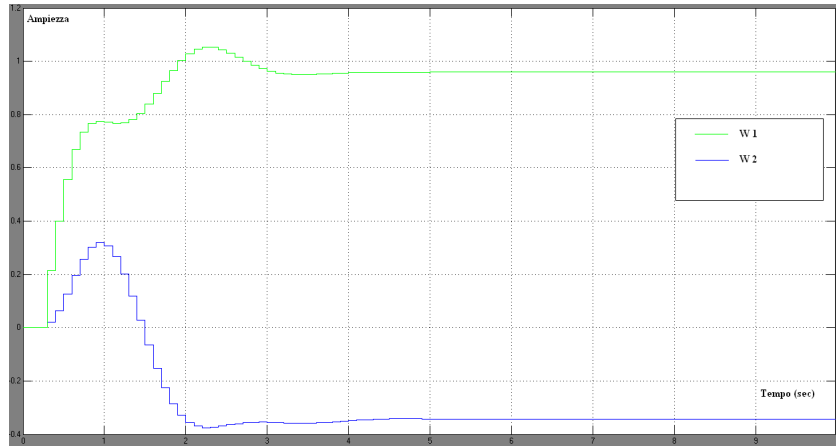


Figura 3.9: Vettore di regressione dei parametri. I due parametri raggiungono i valori di $w(1) = -0.3442$ e $w(2) = 0.9602$.



Figura 3.10: In figura si mostra il confronto tra la velocità del target e la velocità in uscita dal predittore che risulta essere in anticipo dopo un breve transitorio.

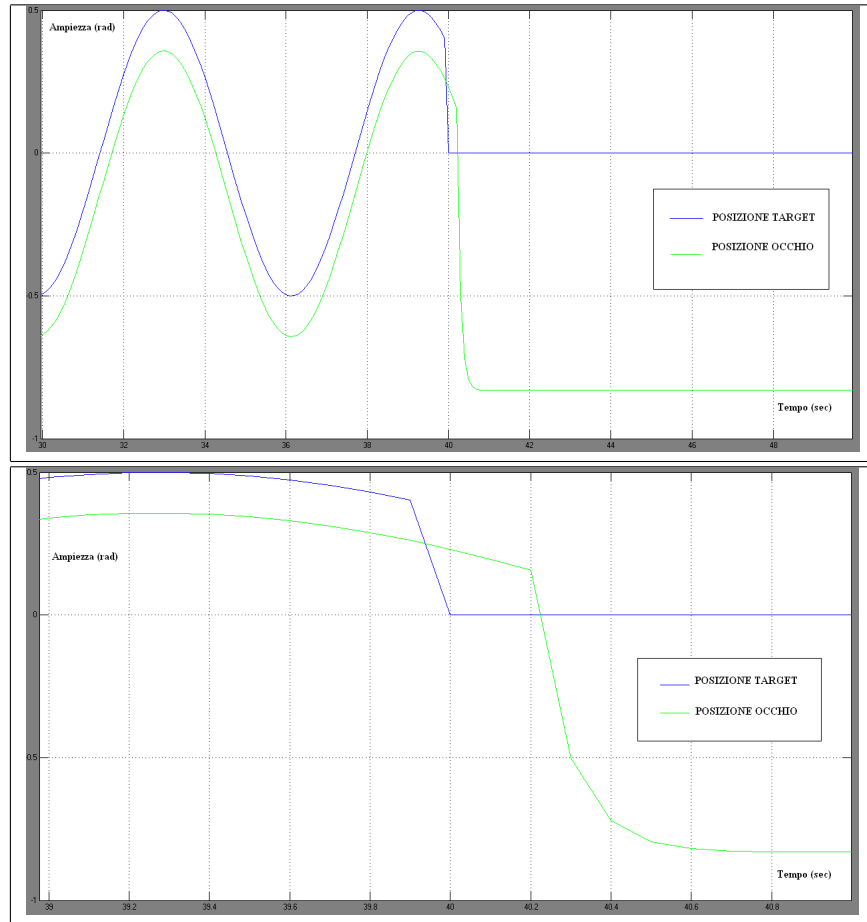


Figura 3.11: In alto viene mostrato il confronto tra posizione del target e quella dell'occhio nel caso in cui ad un predeterminato istante temporale (40 secondi) viene brutalmente arrestata la dinamica del target. In basso viene mostrato il dettaglio nel quale è possibile osservare che la posizione dell'occhio continua con la traiettoria predetta per altri 200 ms in accordo con i risultati biologici.

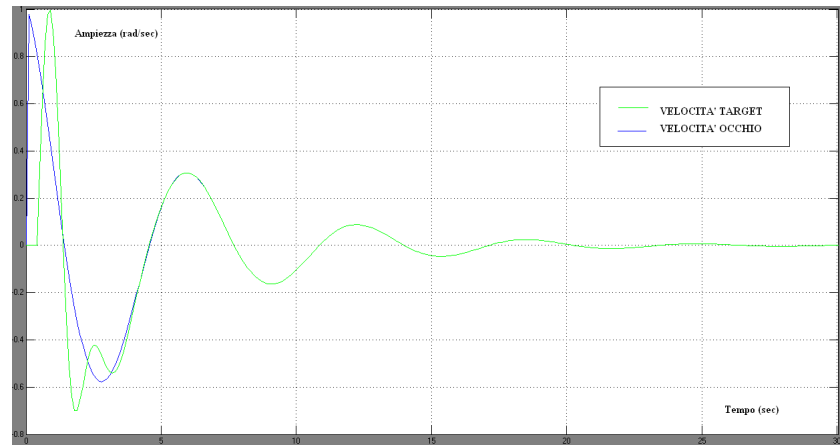


Figura 3.12: Confronto tra velocità del target e velocità dell'occhio nel caso di dinamica con sinusoidi smorzate. I tempi di convergenza risultano analoghi.

3.5 Miglioramento del modello di Schaal

Il modello presentato permette una predizione corretta della dinamica del target, consentendo al sistema oculomotore di restare in linea con il target stesso. In altre parole, il predittore fornisce il segnale che permette all'occhio di mantenere l'oggetto al centro dell'immagine. Come è possibile notare dal grafico (3.8 in basso), lo slip retinale, cioè la differenza tra la velocità del target e la velocità dell'occhio, si annulla entro un tempo di 5 secondi dalla presentazione del target. Questo tempo è notevolmente superiore rispetto ai tempi che si osservano in natura. Questo implica che il modello presentato da Schaal e Shibata non spiega come sia possibile seguire correttamente un oggetto in movimento in tempi rapidi. Il vettore dei parametri di regressione w di fatto genera la predizione della velocità del target a partire dallo stato attuale, come mostrato dalla (3.12), quindi l'inizializzazione dei parametri gioca un ruolo fondamentale nella velocità di predizione. Dalle prove eseguite è stato dimostrato (figura 3.13 e tabella 3.1) che il vettore w dipende esclusivamente dalla frequenza di oscillazione della dinamica del target, in quanto essa determina, a parità di ampiezza della sinusoidi, la velocità del movimento. Esso è infatti indipendente dall'ampiezza della sinusoidi e dal fattore di dimenticanza λ . Il limite principale di questo modello è che esso non conserva memoria delle traiettorie precedentemente acquisite, ma deve necessariamente adattarsi ogni volta che la dinamica del target cambia. In figura 3.14 è mostrato un esempio di tale situazione: inizialmente il sistema si adatta ad una dinamica del target con pulsazione di 0.5 rad/sec, successivamente la frequenza raddoppia e il modello modifica i parametri per adattarsi alla nuova dinamica; a questo punto, il target torna alla precedente velocità ed il sistema necessita dello stesso tempo impiegato nel primo apprendimento, per annullare lo slip retinale. In figura 3.15 è

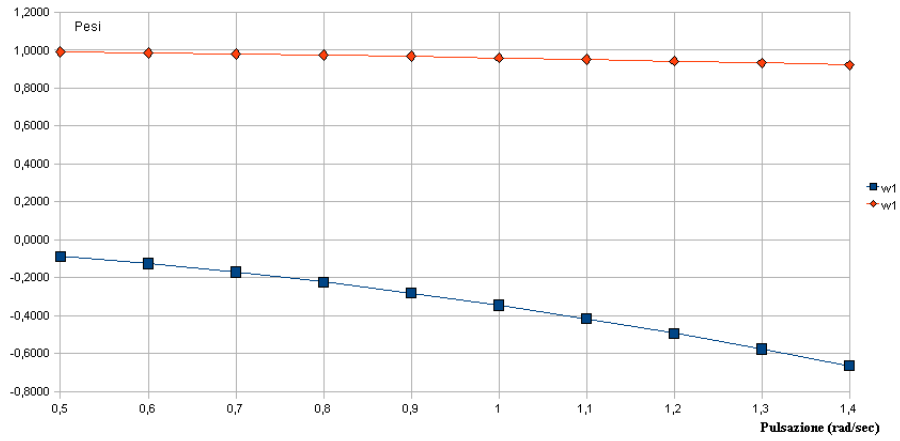


Figura 3.13: Grafico dell'andamento del vettore di regressione dei parametri in funzione della pulsazione della sinusoide in ingresso.

Pulsazione (rad/sec)	Ampiezza (rad)	λ	w_1	w_2
1	1	0.9	-0.3442	0.9602
1	2	0.9	-0.3442	0.9062
1	1	0.85	-0.3442	0.9062
0.5	1	0.9	-0.08714	0.99
0.5	2	0.9	-0.08714	0.99
0.5	1	0.85	-0.08714	0.99

Tabella 3.1: In tabella è mostrata l'invarianza del vettore di regressione dei parametri rispetto all'ampiezza e del fattore di dimenticanza λ .

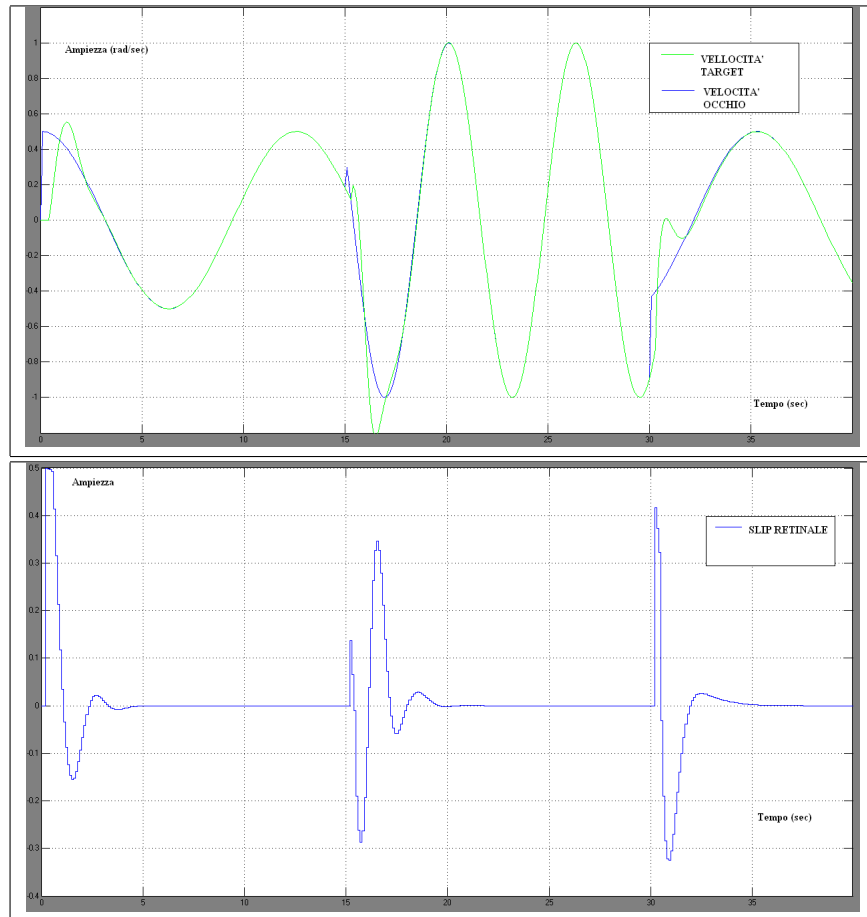


Figura 3.14: Variazione della dinamica del target. In alto sono mostrate la velocità del target e la velocità dell'occhio, nel caso in cui la pulsazione della sinusoide in ingresso passi da 0.5 rad/sec a 1 rad/sec e di nuovo da 1 rad/sec a 0.5 rad/sec. Il modello si adatta perfettamente alla variazione della dinamica, in entrambi i casi. In basso è mostrato lo slip retinale che si annulla con i medesimi tempi in ogni apprendimento.

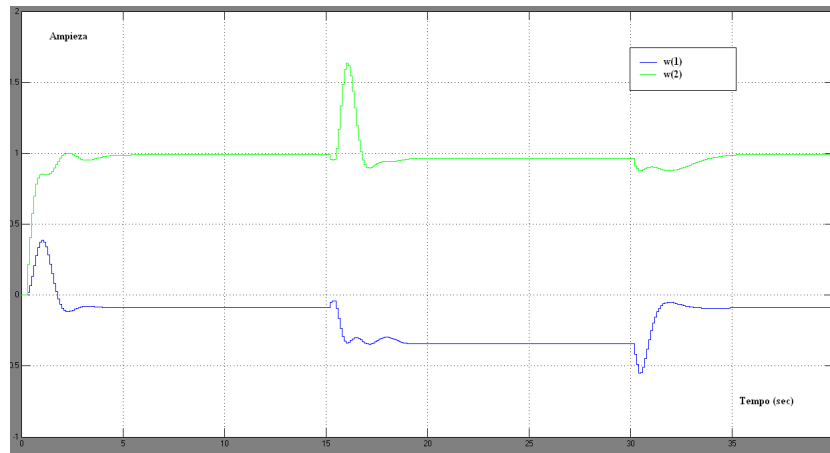


Figura 3.15: Variazione del vettore w nel caso di raddoppio della pulsazione della dinamica e successivo ritorno alla dinamica iniziale. I pesi iniziali e finali, coincidono.

possibile osservare che il vettore di regressione dei parametri torna agli stessi valori precedentemente raggiunti. Se al contrario viene sfruttata la conoscenza dei parametri di una traiettoria precedentemente acquisita, il sistema annulla istantaneamente l'errore di velocità (figura 3.16). Nell'esempio è rappresentata una sinusoide a pulsazione di 1 rad/sec in cui dopo 1 secondo (quindi all'istante di campionamento di 1.1 secondi) vengono forzati i parametri rispettivamente a $w(1) = -0.3442$ e $w(2) = 0.9602$ (figura 3.17). In questo sistema, il corretto annullamento dell'errore, necessita di una forzatura del vettore w per almeno 600 ms. In questo tempo, gli altri parametri del sistema devono continuare ad aggiornarsi in modo tale che a seguito della forzatura, il modello possa riprendere a guidare la traiettoria. E' possibile notare dal grafico che entro 1 secondo dalla forzatura dei parametri il sistema annulla lo slip retinale.

In questo lavoro si ipotizza dunque un'estensione al modello di Schaal che possa memorizzare i parametri necessari alla corretta predizione relativi ad una specifica traiettoria. In questo modo, i risultati di un adattamento, possono essere riutilizzati per migliorare la velocità di annullamento dello slip retinale. Il termine *apprendimento* che viene utilizzato da Schaal, deve essere inteso nella misura in cui il sistema impara a predire la dinamica del target, mentre nel modello proposto è la proprietà del sistema per la quale esso è realmente in grado di imparare dall'esperienza, come presumibilmente lo è il cervello biologico. Nel precedente capitolo, infatti, è stato evidenziato che le aree MT e MST non sono coinvolte nell'avvio dell'esecuzione dello smooth pursuit ma solo nell'elaborazione delle informazioni visive di movimento (si veda anche [37]). Tali informazioni vengono poi trasmesse alla corteccia parietale posteriore che gestisce la memoria visiva a breve termine e alla corteccia prefrontale dove viene eseguita l'associazione tra l'informazione sensoriale e il comando motorio. L'esecuzione

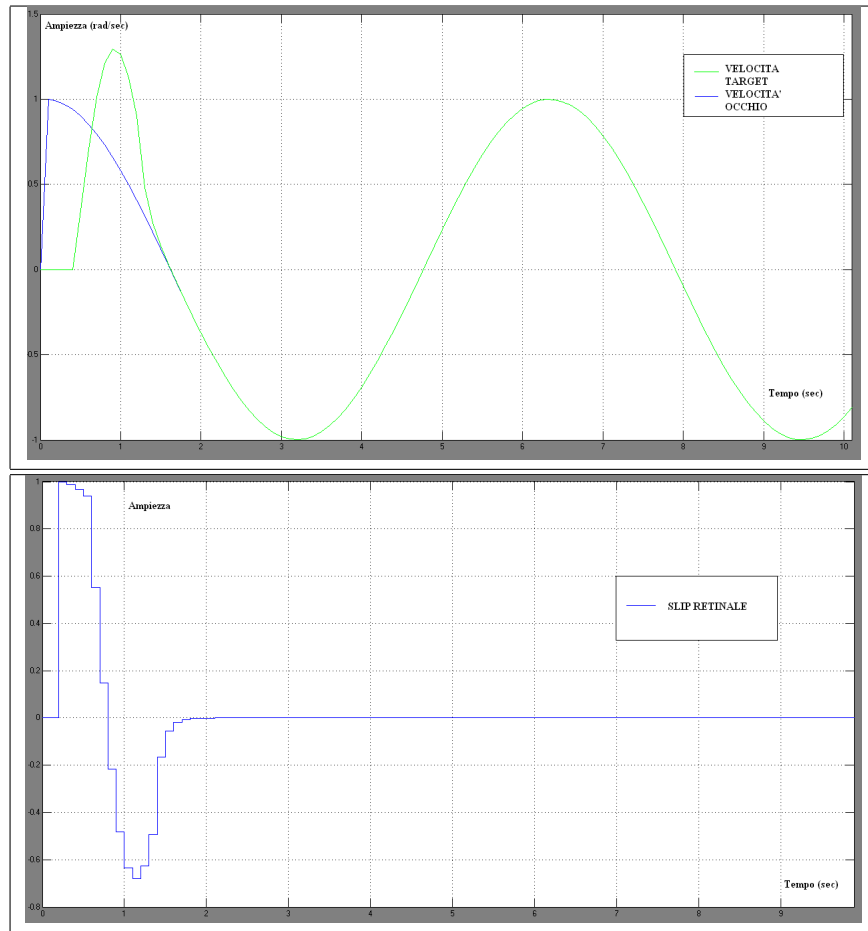


Figura 3.16: Forzatura del vettore di regressione dei parametri con valori noti da una precedente esecuzione. I parametri sono stati impostati dopo 1 secondo, lo slip retinale diminuisce rapidamente e raggiunge lo zero dopo circa un altro secondo.



Figura 3.17: Forzatura del vettore w ad un secondo alla comparsa del target. I valori degli elementi del vettore sono: -0.3442 e 0.9602 .

fluida dei comandi motori, inoltre, è una conseguenza dell'apprendimento di ripetute realizzazioni motorie, il quale ha sede nei gangli della base. Si deduce quindi che le informazioni del movimento di un oggetto e il conseguente movimento oculare vengano memorizzate per un migliore svolgimento del compito. Nel paragrafo 1.4 viene fatto riferimento a studi su bambini, i quali apprendono la corretta esecuzione dell'inseguimento di un oggetto in movimento durante il primo anno di vita. Il sistema, quindi, giunge a maturazione solo dopo una adeguata fase di apprendimento, durante la quale gli errori commessi vengono compensati da ripetuti movimenti saccadici. Sebbene la compensazione saccadica, che corregge l'errore di posizione, si manifesti anche in soggetti adulti, la percentuale di questi movimenti viene ridotta sensibilmente nel primo anno di vita, dimostrando l'apprendimento da parte del soggetto. L'inserimento di una componente mnemonica nel modello possiede dunque una evidente plausibilità biologica, nell'ottica di un approccio emergente (paragrafo 1.2) nella progettazione robotica. Il modello modificato è rappresentato in figura 3.18.

Per memorizzare il vettore w è stata scelta una semplice rete neurale supervisionata. In input alla rete vengono mandati 10 campioni di velocità del target calcolati attraverso la (3.1) che corrispondono ad 1 secondo di elaborazione visiva. La memorizzazione del profilo di velocità risulta essere un vantaggio in quanto permette di riconoscere la dinamica del target indipendentemente dalla posizione che il target stesso assume nello spazio. La rete è in grado di associare tale dinamica con i due corretti elementi del vettore w dopo un'adeguata fase di apprendimento. In figura (3.19) viene mostrata la rete neurale realizzata con Matlab Neural Network Toolbox. È stata scelta una rete *feed-forward* con *back-propagation* in quanto risulta robusta e di facile implementazione ed adeguata per questo scopo. La rete neurale, ispirandosi al sistema biologico, considera un numero elevato di processori che hanno una capacità computazionale elementare,

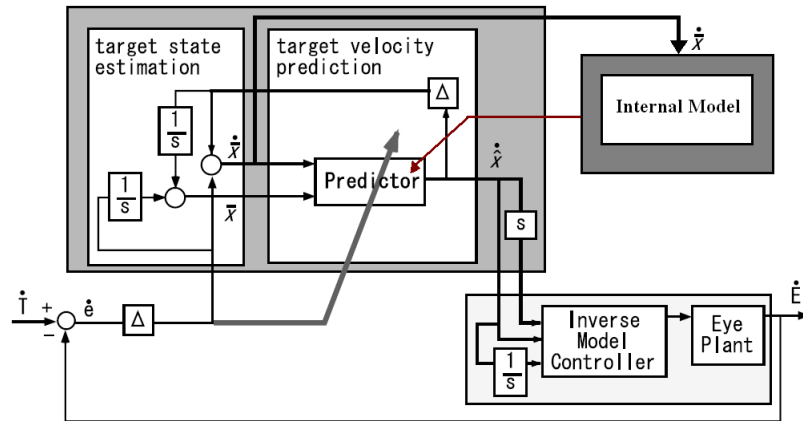


Figura 3.18: Rappresentazione del modello proposto. L'Internal Model costruisce una rappresentazione dei corretti valori del vettore w a partire dalla velocità del target stimata. Ad apprendimento completato, è possibile anticipare il risultato dell'algoritmo.

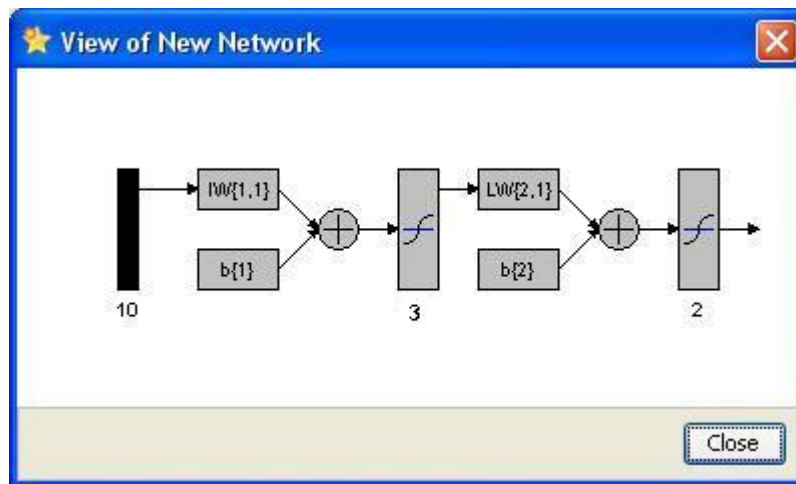


Figura 3.19: Rete neurale realizzata con Matlab Neural Network Toolbox. Sono stati utilizzati come ingressi 10 campioni della velocità del target. L'apprendimento supervisionato è stato eseguito con i risultati delle esecuzioni dell'algoritmo RLS con diverse dinamiche in ingresso.

i *neuroni artificiali* o nodi, connessi ad altre unità dello stesso tipo (di qui il termine connessionismo che designa questo approccio). Ad ogni connessione è assegnato un *peso*, cioè essa può essere più o meno debole in modo tale che un neurone possa influenzarne un altro in funzione della forza della connessione fra i due. Proprio come nei sistemi neurali biologici in cui un neurone (pre-sinaptico) influenza un altro neurone (post-sinaptico) in funzione del potenziale esercitato nella sinapsi fra l'assone del neurone pre-sinaptico e il dendrite del neurone post-sinaptico. La capacità sorprendente di apprendere, propria dei sistemi neurali biologici, risiede nella loro grande plasticità a livello microstrutturale (Hebb, 1949). L'apprendimento citato anche in [], potrebbe essere definito come la modifica del comportamento in conseguenza dell'interazione con l'ambiente e il risultato dell'esperienza che determina l'emergere di nuovi pattern di risposta agli stimoli esterni (McCord Nelson e Illingworth, 1991). Le sinapsi biologiche modificano le loro proprietà in relazione a particolari stimoli esterni per far sì che il sistema neurale nel suo complesso produca risposte adeguate a determinati stimoli ambientali. Questa è la caratteristica a cui si ispirano i sistemi neurali artificiali. In altre parole concetti come interazione con l'ambiente, stimoli e pattern di risposta trovano i loro corrispondenti all'interno della computazione neurale. Dire che una rete interagisce con l'ambiente significa sostanzialmente affermare che dall'ambiente riceve gli stimoli sotto forma di *pattern d'ingresso* e risponde attraverso la produzione di un *pattern di uscita*. La risposta della rete neurale deve essere modificata in modo da diventare la più appropriata possibile al pattern d'ingresso in quel momento. Questo viene ottenuto variando il valore dei pesi delle connessioni in modo che in presenza di un determinato pattern in ingresso alla rete si abbia una risposta corrispondente. La risposta della rete, il pattern di uscita, può essere decisa a priori. In questo caso si parla di *apprendimento supervisionato*: l'operatore umano decide che in corrispondenza di un determinato pattern vi debba essere una determinata risposta. Altrimenti, se la forma della risposta non è nota a priori, si parla di *apprendimento non supervisionato*. Un altro aspetto importante nelle reti neurali è la capacità di generalizzazione definito in come la capacità di interpolare una funzione dopo un addestramento basato su un numero limitato di campioni oppure di classificare correttamente dei nuovi pattern non presenti nei dati di addestramento. Si può parlare di *superimposizione* come la proprietà di una rete neurale di raggiungere, dopo l'apprendimento, una configurazione dei pesi stabile in modo tale che presentando pattern diversi in ingresso alla stessa rete essa fornisca risposte diverse per ognuno di questi pattern, coerenti con gli esempi forniti nella fase di apprendimento. I pattern utilizzati nella fase di apprendimento sono chiamati *pattern di addestramento*. Quando si presenta un pattern nuovo alla rete e la fase di apprendimento è finita di conseguenza tale pattern prende il nome di *pattern di richiamo* (recall). Questo viene processato in base alle stesse proprietà dei pattern di addestramento per cui la rete ha appreso ad associare tali pattern ad una specifica uscita. La rete è in grado quindi di generalizzare, proprietà molto importante nel campo delle reti neurali artificiali perché rende questo strumento versatile ed efficiente anche in situazioni inesplorate.

Durante la presentazione di una dinamica sconosciuta, l'algoritmo RLS resti-

tuisce i corretti parametri i quali possono essere successivamente impiegati per l'apprendimento supervisionato della rete neurale. Ad apprendimento completato, la rete neurale è anche in grado di restituire un valore corretto dei parametri anche in caso di presentazione di un movimento analogo a quelli precedentemente appresi, ma con differenti caratteristiche. Per esempio, se la rete ha imparato ad associare una serie di movimenti sinusoidali di diversa frequenza, essa può fornire un risultato corretto in presenza di una sinusoide con frequenza non precedentemente presentata. Nella visione completa del modello quindi, il sistema associa il movimento del target con gli elementi del vettore dei parametri di regressione w che consentono una corretta predizione della velocità del target. Il processo di riconoscimento necessita di 10 campioni (1 secondo con tempo di campionamento di 100 ms) per identificare il profilo di velocità, quindi risulta decisamente più rapido del precedente modello. Il sistema segue una dinamica nota in tempi più rapidi rispetto ai 5 secondi nel modello di Schaal e Shibata.

Capitolo 4

Implementazione sulla piattaforma robotica iCub

4.1 Introduzione

Il modello esposto nel precedente capitolo rappresenta la formalizzazione matematica di un fenomeno che può essere riprodotto su una qualsiasi piattaforma che ne consenta la corretta implementazione. L'ambiente di simulazione di Matlab Simulink rimane ad un alto livello d'astrazione e solo l'implementazione in un sistema robotico può validare l'efficacia della tecnica. Occorre, però, fare i conti con i limiti imposti dall'hardware utilizzato e necessariamente, sebbene sia proprio la volontà del superamento di tali limiti che induce la progettazione di tecnologie più avanzate, le prestazioni del sistema risulteranno inferiori a quelle mostrate dal puro ambiente di Matlab. In questa fase del lavoro il modello è stato implementato su un robot umanoide denominato *iCub* ([38, 39]). Questo, ha le stesse dimensioni di un bambino di due anni di età ed è stato progettato per essere capace di compiere i normali gesti di un bambino in fase di sviluppo: può camminare a quattro zampe e sedersi, le sue mani permettono la manipolazione abile, e ha testa ed occhi completamente articolati. Possiede capacità visive, vestibolari, e aptiche (figura 4.1). Essendo un sistema aperto, il progetto e la documentazione di tutto l'hardware e software possiede licenza *Free Software Foundation GNU* così che l'intero sistema possa essere riprodotto, utilizzato e modificato. Concordemente a questo principio è stato distribuito un software di simulazione che gestisce il robot in un ambiente simulato che riproduce le leggi fisiche ed è dunque possibile testare il modello presentato come se fosse implementato sul robot reale. In questo capitolo viene descritta la piattaforma robotica con particolare riferimento alle specifiche tecniche del sistema occhi-testa del robot. Viene inoltre mostrato l'ambiente di sviluppo utilizzato per le prove sperimentali del modello proposto.

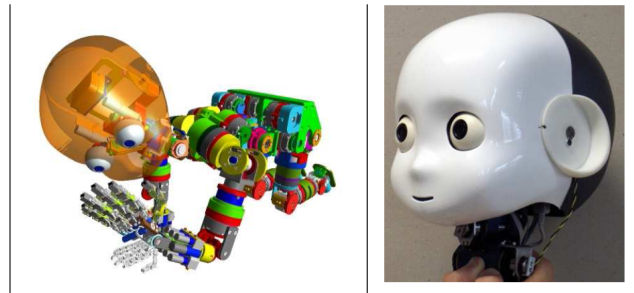


Figura 4.1: La piattaforma robotica iCub del progetto RobotCub ha le stesse dimensioni e capacità motorie di un bambino di due anni. A destra è mostrata la testa del robot che possiede 6 gradi di libertà.

4.2 Piattaforma iCub

Il progetto RobotCub consiste nella realizzazione di una piattaforma robotica umanoide, denominata iCub, per lo studio della cognizione umana. Gli obiettivi del progetto sono: 1) creare una piattaforma aperta e liberamente disponibile per la ricerca scientifica sulla cognizione e 2) l'avanzamento della comprensione dei sistemi cognitivi attraverso lo studio dello sviluppo di essi. Il sistema deve quindi essere capace di imparare come interagire con l'ambiente e come sviluppare le proprie capacità percettive, motorie e comunicative. La forma e la dimensione di iCub sono le stesse di un bambino di due anni [40] ed esso può ottenere le sue capacità cognitive attraverso lo sviluppo ontogenetico relazionato all'ambiente con cui interagisce (figura 4.1). Inoltre le ridotte dimensioni della piattaforma permettono sia una maggiore sicurezza in quanto si riducono i danni causati da una sua possibile caduta, sia di migliorare il controllo dato che sono più facili da manovrare. Questo robot possiede testa, torso, due braccia con mani e due gambe usate principalmente per camminare a quattro zampe. Nella versione attuale del progetto, l'iCub è alto circa 90 cm, pesa 23 Kg ed ha in totale 53 gradi di libertà distribuiti come segue: 7 per ogni braccio, 8 per ogni mano, 6 per la testa, 3 per il dorso e 7 per ogni gamba. Il sottosistema occhi-testa include primitive di processamento visivo di base, controllo di basso livello, sensori visivi, inerziali e propriocettivi. La maggior parte dei sistemi umanoidi esistenti hanno una struttura della testa semplificata con un ridotto numero di gradi di libertà. In questo caso, l'interazione con altri robot o individui è molto importante e questo giustifica la necessità di raggiungere una complessità maggiore, mantenendo stringenti le specifiche di progetto in termini di peso ed altezza. La progettazione meccanica finale è liberamente disponibile su web [41] in quanto sottoposta a licenza pubblica (General Public License GPL). Per questo lavoro risulta di particolare importanza citare le specifiche di progettazione che riguardano la testa di iCub [42]. Questa pesa 1.5 Kg e le sue dimensioni sono 13.6 x 17 x 17.3 cm, mentre il collo è 7 x 9 cm. I 6 gradi di libertà della testa sono mostrati in figura 4.2 e comprendono: pan (imbardata),

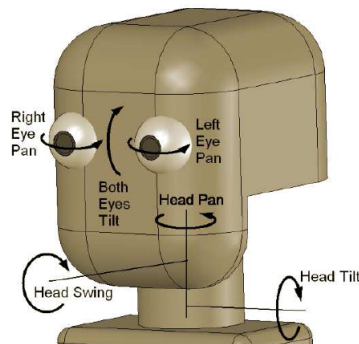


Figura 4.2: Illustrazione dei gradi di libertà della testa. In totale possiede sei gradi di libertà, tre per il collo e tre per il sistema degli occhi.

Adult values Head weight: 4.5 – 5Kg		Range [⁰]	Velocity [⁰ /s]		Acceleration [⁰ /s ²]	
			min	max	min	max
eyes	pan	90	166	850	16000	82000
	tilt	80				
neck	pan	110	23	352	330	3300
	tilt	90				
	swing	80				
Neck/eye (pan) ratio			14%	41 %	2%	4%

Tabella 4.1: Valori di range di movimento (in gradi), velocità (gradi per secondo) ed accelerazione (gradi per secondo quadro) per la testa adulta. Una specifica importante è il rapporto tra la velocità del collo e la velocità degli occhi.

tilt (beccheggio) e swing (rollio) della testa e pan (indipendente) e tilt (comune) degli occhi. La torsione degli occhi è stata ignorata perché non utile per il controllo. L'elevazione e l'abbassamento (tilt) degli occhi è sempre comune negli umani, nonostante l'esistenza di muscoli indipendenti e in modo analogo, un attuatore unico è usato per l'elevazione degli occhi di iCub. La vergenza degli occhi è invece assicurata da motori indipendenti. Nei movimenti di versione i due occhi si muovono nella stessa direzione sincronizzati da un controllore DSP. Per il movimento di vergenza gli occhi invece si muovono in verso opposto. Per le specifiche dei raggi d'azione dei gradi di libertà della testa sono stati usati come riferimenti i dati della testa umana adulta (tabella 4.1). I dati sono stati ridimensionati considerando che i bambini hanno velocità significativamente più basse e che le saccadi di piccola ampiezza sono molto simili al movimento di smooth pursuit. E' stato inoltre considerato un rapporto tra la velocità degli occhi e quella del collo compreso tra il 14% e 41% e per le accelerazioni, tra il 2% e il 4%. Usando queste informazioni e ipotizzando un profilo di movimento trapezoidale per ogni movimento degli occhi, sono stati definite le specifiche per ogni giunto (tabella 4.2). E' stata scelta la velocità massima di pan per gli occhi

CAPITOLO 4. IMPLEMENTAZIONE SULLA PIATTAFORMA ROBOTICA ICUB60

		Range [$^{\circ}$]	Max Vel. [$^{\circ}/s$]	Acceleration Full range		
				acc.[$^{\circ}/s^2$]	T _{tot} [s]	mean vel
eyes	pan	90	180	1440	0.625	144
	tilt	80	160	1280	0.625	128
neck	pan	110	90	295	1.528	72
	tilt	90	73.6	241	1.528	59
	swing	80	65.5	214	1.528	52

Tabella 4.2: Specifiche di velocità angolare ed accelerazione dei giunti, usate pre la testa di iCub.

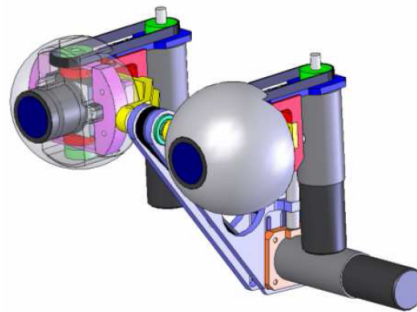


Figura 4.3: Meccanismo che gestisce i tre gradi di libertà degli occhi. La cinghia centrale permette il tilt, mentre il movimento di pan è interno alla palla dell'occhio.

pari a $180^{\circ}/s$ assumendo che il collo possa raggiungere la metà di questa velocità. Il 20% del tempo necessario per il movimento è usato per l'accelerazione mentre un altro 20% per la decelerazione. La rimanente parte della traiettoria è eseguita alla massima velocità. Il meccanismo degli occhi è mostrato in figura (4.3). Esso ha tre gradi di libertà, entrambi gli occhi possono eseguire il pan (indipendentemente) e tilt (simultaneamente). Il movimento di pan è guidato attraverso un sistema a cinghia con il motore all'interno della palla dell'occhio. Il tilt è attuato attraverso un'altra cinghia posizionata in mezzo ai due occhi. Ogni sistema possiede un meccanismo di aggiustamento della tensione. In modo da garantire un facile assemblaggio e semplici procedure di manutenzione, l'architettura meccanica è completamente modulare, in questo modo è possibile rimuovere e sostituire un modulo senza dover smontare l'intera struttura. Per la visione sono state utilizzate due telecamere di tipo DragonFly [43] con risoluzione VGA e 30 *fps* di velocità. Queste telecamere sono particolarmente facili da installare in quanto il sensore CCD è montato su una testina remota connessa con il resto dell'elettronica attraverso un cavo flessibile. In questo modo, il sensore è montato nel globo oculare, mentre l'elettronica è fissata su una scheda non mobile nella parte posteriore della testa. La testa contiene sensori inerziali MTi che possono fornire la posizione del corpo. Sono stati inoltre installati diversi microfoni intorno alla testa per poter localizzare la provenienza

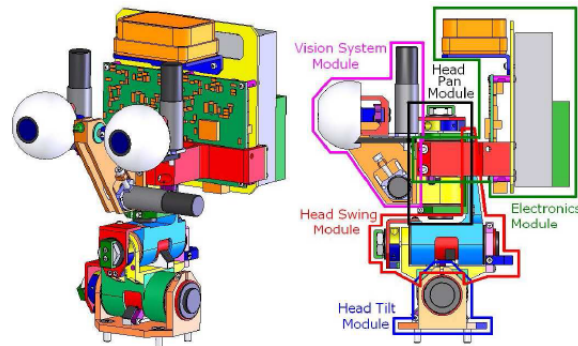


Figura 4.4: L'architettura meccanica è completamente modulare consentendo la sostituzione rapida ed indipendente di ogni modulo.

del suono in tutta l'area circostante. Tutte le schede di controllo sono integrate all'interno della testa e sono connesse ad un computer remoto attraverso un bus CAN. Per misurare la posizione dei giunti della testa, i motori sono forniti di encoders magnetici.

4.3 YARP e l'ambiente di simulazione

YARP è una struttura software scritta da e per i ricercatori in robotica, in particolare per la robotica umanoide. E' formata da un insieme di librerie, protocolli e strumenti che permettono di mantenere i moduli e i dispositivi completamente disaccoppiati. Nei progetti robotici di solito è necessario cambiare il tipo di software o di hardware spesso per effettuare prove ed identificare il dispositivo che meglio risponde alle specifiche richieste. Occorre però che non si abbia traccia del dispositivo sostituito in modo da non influenzare il funzionamento del sistema. Infatti cause comuni di malfunzionamenti sono le dipendenze di dispositivi o librerie non utilizzati o obsoleti. YARP è programmato per rendere i software di robot più stabili e durevoli permettendo all'utente la possibilità di poter cambiare sensori, attuatori, processori e reti, aiutando la comunicazione tra i dispositivi. Nel modello di comunicazione di YARP il flusso di dati è disaccoppiato dai dettagli delle reti e protocolli da cui derivano. Inoltre esso usa una metodologia per l'interfacciamento con i dispositivi (sensori ed attuatori) che migliora il disaccoppiamento e rende il cambio dei dispositivi più sicuro, riducendo i problemi di incompatibilità. Il software YARP è completamente open-source permettendo lo sviluppo del software più veloce.

La ricerca su piattaforme robotiche implica la necessità di controllare numerosi hardware utilizzando, in tempo reale, informazioni visive, uditive e tattili. Questo richiede molte risorse computazionali e, attualmente, la soluzione migliore adottabile è quella di dividere tale costo in una rete di computer. YARP

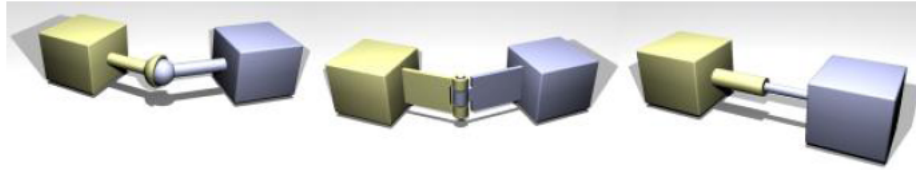


Figura 4.5: Tipologie di vincoli simulati dall' ODE. In questo modo è possibile calcolare le reazioni delle forze in seguito ad interazioni tra oggetti.

permette di gestire le informazioni provenienti dai dispositivi indipendentemente dal tipo di macchina dalla quale provengono.

I componenti di YARP possono essere divisi in:

- `libYARP_OS` = interfaccia con sistemi operativi per supportare il flusso di dati attraverso diversi *thread* e diverse macchine. YARP è scritto per essere compatibile con ogni sistema operativo, quindi può essere utilizzato su sistemi operativi Linux, Microsoft Windows, Mac OSX e Solaris. YARP utilizza le librerie open-source ACE (ADAPTIVE Communication Environment) che sono anch'esse compatibili con ogni sistema operativo. Esso è scritto quasi interamente in C++;
- `libYARP_sig` = gestione dei compiti di processamento di segnali (video, audio) in modo facilmente interfacciabile con altre librerie comunemente utilizzate, per esempio le OpenCV;
- `libYARP_dev` = interfaccia con i dispositivi comuni utilizzati in robotica: telecamere, schede di controllo per motori, etc.

Questi componenti sono mantenuti separati. Il nucleo del sistema è `libYARP_OS`, che deve essere disponibile prima che gli altri componenti possano essere usati. Per le operazioni in tempo reale YARP utilizza una rete isolata ed è consigliabile che resti separata dalla rete globale. I dispositivi hardware sono spesso dipendenti dal sistema operativo che supportano. Quindi `libYARP_dev` è strutturata per interfacciare facilmente quel particolare dispositivo e per proteggere il resto del sistema da quel particolare codice in modo da rendere agevole il cambio dell'hardware.

Per la fase di test sul robot è stato utilizzato un prototipo di simulatore di iCub che usa il motore grafico open-source Open Dynamics Engine (ODE)[44]. Esso calcola la leggi fisiche e la dinamica dei motori e permette al robot simulato di considerare il ruolo dei vincoli fisici all'interno di un ambiente riprodotto. Inoltre può calcolare e risolvere forze che emergono attraverso l'interazione di oggetti. In questo modo è possibile calcolare le forze relative ai vincoli nei giunti (figura 4.5). ODE include una interfaccia alle OpenGL che facilita la resa grafica degli oggetti nell'ambiente (sfera, scatola, etc.). Il modello simulato è una replica dell'attuale piattaforma iCub che utilizza dati e matrici di inerzia. La piattaforma e i dati stessi possono essere aggiornati e modificati per una

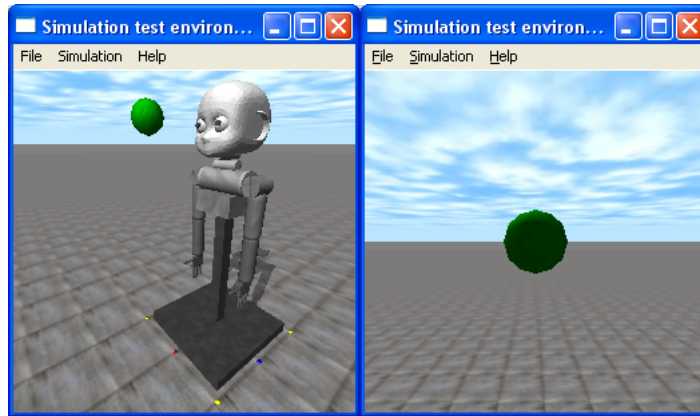


Figura 4.6: Immagini estratte dal simulatore. E' possibile osservare il robot dall'estero (a sinistra) o estrarre la visione da entrambi gli occhi. A destra è mostrata la posizione della palla che risulta essere centrata ad 1 metro di distanza dall'occhio sinistro.

migliore resa. Questo simulatore è stato sviluppato per gli utenti del progetto RoboCub come alternativa alla piattaforma fisica per poter sviluppare e testare rapidamente i prototipi studiati. L'obiettivo è rendere la simulazione più vicina possibile al reale robot iCub usando lo stesso tipo di interfaccia. Infatti YARP controlla la simulazione attraverso la stessa interfaccia che controlla il robot, quindi sfrutta gli stessi comandi e gli stessi parametri. Perciò gli sviluppatori possono sperimentare lo stesso ambiente software con l'addizionale sicurezza dell'esecuzione in un mondo simulato. Per gli obiettivi di questo progetto, risulta particolarmente utile il fatto che è stato implementato il sistema di visione. Esso appare in un'altra finestra grafica e mostra cosa il robot simulato vede nel suo ambiente. La visione può essere ricevuta da entrambi gli occhi. Nella versione del simulatore utilizzata, vengono riprodotte solo le braccia, il busto e la testa del robot (figura 4.6).

4.4 Scelte implementative

Il modello proposto è stato implementato in codice C++ sulla piattaforma robotica iCub gestendo le immagini visive provenienti dall'occhio sinistro e controllando il giunto 4 della testa che permette il movimento di versione degli occhi. Il programma dunque, si connette a YARP attraverso il metodo `Network::init()` e crea una porta di tipo buffer con `BufferedPort<ImageOf<PixelRgb>> portleft`. La porta si connette con la telecamera sinistra del robot con i metodi `portleft.open("/leftcam")` e `Network::connect("/icubSim/cam/left", "/leftcam")` e riceve le immagini che vengono inviate. L'immagine è prelevata ad ogni ciclo dal metodo `imgl=portleft.read()`. Da queste è possibile determinare l'oggetto presente nel campo visivo rintracciando il colore dell'oggetto. In questa sede è stato così semplificato il proces-

CAPITOLO 4. IMPLEMENTAZIONE SULLA PIATTAFORMA ROBOTICA ICUB64

samento delle immagini in quanto è stato preferito concentrare le risorse computazionali sul controllo del movimento. Il target da inseguire è una palla di colore verde che si muove con dinamica sinusoidale parallelamente al terreno. E' stato necessario, dunque, eliminare dalla simulazione la forza di gravità. La palla inizialmente si trova nel punto centrale dell'immagine a distanza di un metro dall'occhio. In accordo con gli studi neuroscientifici il processamento del pursuit orizzontale è separato da quello verticale, di conseguenza è stato implementato un solo modello che gestisce l'inseguimento della traiettoria longitudinale. Dall'immagine ricevuta è necessario estrarre il centroide della palla come punto unico da inseguire. In questa operazione l'immagine viene filtrata in base al colore verde, ottenendo così una seconda immagine in bianco e nero nella quale appunto, il verde assume il colore nero e dunque la palla risulta ben visibile. Da questa è stato calcolato il centroide cercando per ogni riga della palla il valore centrale, il centroide lungo l'asse x risulta essere la media dei valori ottenuti. La distanza dal centroide dal centro dell'immagine rappresenta l'errore di posizione, mentre la variazione del punto nel tempo rappresenta lo slip retinale, come definito dal paragrafo 3.2. Tali valori sono espressi in pixel e pixel al secondo rispettivamente ed occorre convertirli in gradi e gradi al secondo per essere confrontati con i valori provenienti dall'encoder. Si assume per semplicità che la telecamera non commetta errori dovuti alla distorsione della lente e quindi è possibile determinare empiricamente una costante di trasformazione che in questo caso vale $k = 0.34$. Questo significa che un pixel di distanza dal centro dell'immagine corrisponde circa ad un angolo di 0.34° che l'occhio deve compiere per centrare il target. Il controllo del robot può essere effettuato direttamente in velocità grazie ai controllori di basso livello che gestiscono gli occhi di iCub. L'uscita dall'algoritmo viene direttamente inviata al giunto di controllo della versione del robot attraverso il metodo `vel->velocityMove(4,(output))` nel quale 4 rappresenta il giunto da inseguire e output rappresenta l'uscita dal predittore. Valori positivi corrispondono ad un movimento verso destra degli occhi. I motori ed i sensori del robot sono calibrati automaticamente dal simulatore. Per settare il controllo in velocità è necessario abilitare la funzione attraverso il metodo `vel->setVelocityMode()`. L'algoritmo RLS è implementato nei file `rls.h` e `rls.cpp`. Nel main viene inizializzato il predittore attraverso il metodo `RLS *filter=new RLS (1,0.9,coeffW,initX)`. In cui occorre fornire il valore del fattore di dimenticanza che è stato scelto pari a $\lambda = 0.9$ e i vettori di `coeffW` e `initX` che rappresentano rispettivamente il vettore dei parametri di regressione e lo stato iniziale. La matrice P dell'algoritmo RLS è inizializzata con una matrice diagonale con valori [1 0; 0 1], mentre tutti gli stati iniziali del sistema di apprendimento sono settati a zero. Sono state eseguite prove a diversa frequenza di oscillazione della palla per ottenere i valori dei parametri di regressione w .

Capitolo 5

Risultati e conclusioni

5.1 Risultati

Il modello è stato implementato sul simulatore iCub ed è stato testato con dinamiche sinusoidali di pulsazione compresa tra i 0.5 rad/sec e 2.5 rad/sec. Il simulatore calcola ad ogni passo il nuovo stato di tutti i componenti all'interno del mondo simulato. Il passo temporale può essere scelto dall'utente modificando il TIMESTEP all'interno del file *iCub_Sim.h*. Bassi valori assicurano un'alta precisione nel calcolo delle funzioni, ma rallentano il tempo della simulazione. E' stato scelto un TIMESTEP di 0.01 secondi come buon compromesso tra velocità di simulazione e precisione. In questo modo il simulatore impiega 100 iterazioni per effettuare un secondo di attività del robot. Per simulare la dinamica del target, la palla viene, ad ogni iterazione, spostata secondo la seguente legge:

$$xs = A * \sin(count * \omega * TIMESTEP) + C \quad (5.1)$$

Dove xs rappresenta la posizione della palla nelle coordinate del mondo simulato, $A = 0.364$ rappresenta l'ampiezza della sinusoide corrispondente ad un angolo di 20° per ogni lato (infatti $\tan(20) = 0.364$), per un totale quindi di 40° di spostamento rispetto al centro dell'occhio, $count$ è il contatore dei passi di iterazione, ω è la pulsazione espressa in radianti al secondo e $C = 0.175$ è una costante che permette di centrare la palla rispetto all'occhio sinistro. Inizialmente la palla resta ferma per 400 passi (4 secondi nel tempo del robot), centrata alla posizione [1.36; 0.175; 3.5]. Le misure sono espresse in metri. Anche l'immagine sulla videocamera viene aggiornata con lo stesso TIMESTEP quindi, per ottenere un frame rate della telecamera simulata pari a 10 Hz è stato necessario prelevare 1 immagine ogni 10 iterazioni del simulatore. Occorre considerare che il simulatore è un software che richiede molte risorse computazionali e quindi esso risulta particolarmente dipendente dal tipo di macchina in cui viene utilizzato. Questo genera, su computer non eccessivamente veloci, diversi errori e influenza anche la corretta esecuzione del modello. Altro problema riscontrato è che le telecamere simulate risultano avere una risoluzione piuttosto bassa (256x256

pixel) quindi l'errore che si commette nel calcolo del centroide è piuttosto altro dato che 1 pixel che corrisponde a 0.34° . Inoltre occorre considerare che il calcolo della velocità di spostamento sulla retina viene eseguito attraverso la differenza della posizione attuale meno quella rilevata nell'immagine precedente diviso il tempo che è trascorso (0.1 secondi). Questo calcolo genera grandi errori che vengono rilevati dall'algoritmo e processati per la generazione del comando motorio. Tale problema è stato parzialmente risolto al livello del predittore in quanto quest'ultimo utilizza lo stato del target (inteso come posizione e velocità) ritardato di un passo. Quindi nel momento della processamento dello stato è già disponibile l'informazione riguardante la velocità del target di un istante in avanti. Il calcolo della velocità risulta particolarmente migliorato se mediato su tre istanti temporali secondo l'equazione:

$$v = \frac{1}{2} * \left[\frac{x(t+1) - x(t)}{\Delta t} + \frac{x(t) - x(t-1)}{\Delta t} \right] \quad (5.2)$$

Dove v è la velocità del target, x è la posizione del target data dalla somma della posizione retinale e la posizione dell'occhio fornita dall'encoder e $\Delta t = 0.1$ secondi. Alla fine dell'utilizzo dello stato memorizzato X per la predizione, *rls.cpp* aggiorna il vettore con lo stato attuale x . Occorre dunque mediare la velocità memorizzata con quella attuale attraverso la funzione $x[1] = (X[1] + x[1]) / 2$ presente nel file *rls.cpp*. Il metodo di rls denominato *sample* restituisce, infatti, il valore predetto della velocità del target e riceve in ingresso lo stato attuale x (che memorizza per la successiva iterazione) e lo slip retinale attuale e .

In figura 5.1 è rappresentato il risultato della simulazione in presenza di una dinamica del target con pulsazione pari a 0.5 rad/sec. Dal grafico del confronto delle velocità è possibile notare come la velocità dell'occhio segua quella del target. Il valore di slip retinale risulta essere minore di 1 grado al secondo 5.2. In figura 5.3 è invece mostrato l'andamento del vettore dei parametri di regressione che dopo un transitorio si assestano a valori circa costanti di $w1 = -0.025$ e $w2 = 0.9932$. Il grafico in figura 5.4 mostra l'andamento dei parametri in funzione della pulsazione della sinusoide. Sono state provate sinusoidi con pulsazione compresa tra 0.5 e 2.5 rad/sec. Sopra ai 1.2 rad/sec si consiglia di spostare la posizione iniziale della palla lungo la stessa sinusoide, come spiegato in seguito. L'andamento della caratteristica è paragonabile a quello ottenuto in Matlab-Simulink (vedi figura 3.13). Al crescere della frequenza, il sistema commette più errore e si può osservare in figura l'andamento per una sinusoide con pulsazione di 2.2 rad/sec. Sopra i 2.2 rad/sec il sistema non è in grado di seguire la dinamica del target. Questo è confermato dal fatto che lo smooth pursuit non supera su soggetti umani la velocità di 80° /sec. Per tale modello, la pulsazione di 2.2 rad/sec produce velocità dell'occhio di oltre 50° /sec.

Per alte frequenze è utile spostare la palla nella posizione iniziale. Infatti il valore centrale della sinusoide espressa dalla 5.1 risulta anche il momento in cui il target ha velocità più alta. Questo implica che la palla comincia a spostarsi subito con la velocità massima e il robot non riesce ad allinearsi alla dinamica abbastanza velocemente. Partendo invece da un quarto di periodo, quando cioè

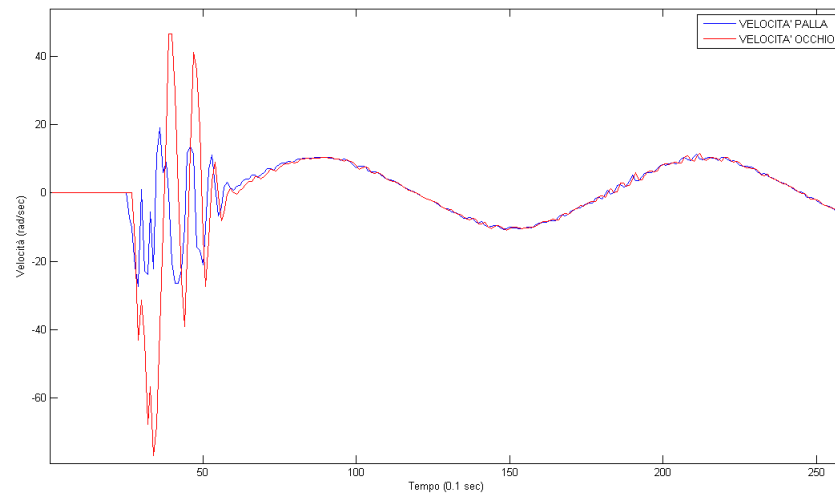


Figura 5.1: Confronto tra la velocità del target e la velocità dell'occhio misurata dall'encoder. Il target si muove con traiettoria sinusoidale con pulsazione pari a 0.5 Hz. Gli errori dovuti alla misura della velocità sono ben visibili. Dopo più di 5 secondi dal movimento del target è possibile considerare le dinamiche allineate.

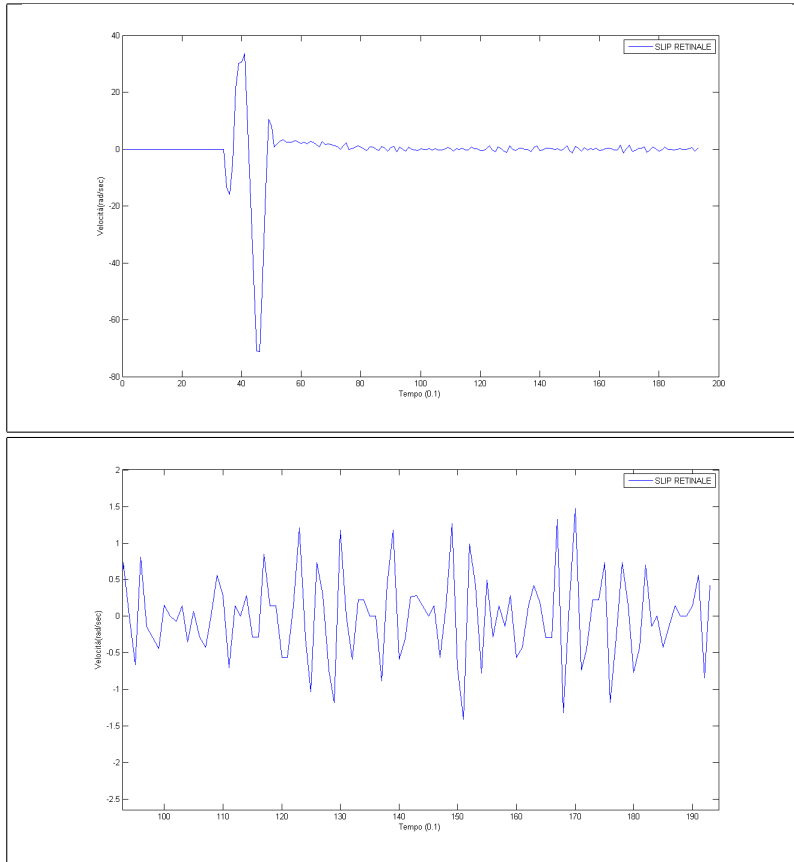


Figura 5.2: In alto è mostrato lo slip retinale in cui si osserva che dopo circa 5 secondi (ogni step nel grafico sono 0.1 secondi nel tempo del robot) dall'avvio della dinamica del target (8 secondi nel grafico), l'errore raggiunge valori minori di 1 grado (nel dettaglio in basso). Tale errore risulta accettabile ed è quindi considerato che l'occhio segue il target.

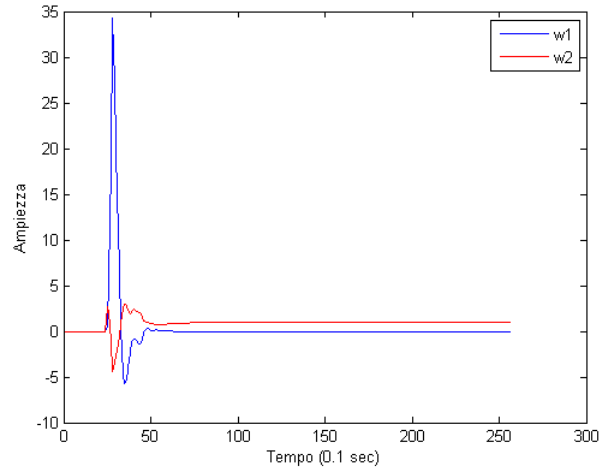


Figura 5.3: Grafico del vettore dei parametri di regressione per una senoide con pulsazione di 0.5 rad/sec. I parametri raggiungono una condizione stabile dopo circa 5 secondi dall'avvio della dinamica del target (8 nel grafico).

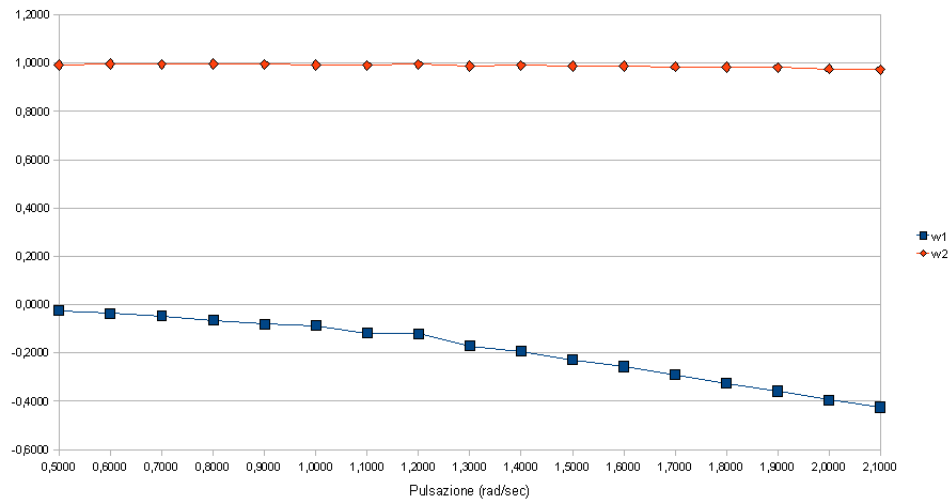


Figura 5.4: Il grafico mostra l'andamento dei parametri in funzione della pulsazione della senoide. Sono state provate sinusoidi con pulsazioni comprese tra 0.5 e 2.5 rad/sec ad intervalli di 0.1 rad/sec. I parametri sono ottenuti mediando i valori ottenuti nelle ultime 200 iterazioni considerando quindi la zona più stabile della caratteristica.

la palla è praticamente ferma, l'accelerazione di questa e di conseguenza dell'occhio risulta essere meno repentina. Occorre ricordare che il sistema non annulla l'errore di posizione, ma ha come unico scopo quello di allineare le due velocità. A differenza delle simulazioni in Matlab-Simulink, la finestra che rappresenta la visione del robot è fortemente limitata. Quindi con eccessive escursioni della posizione dell'occhio si rischia di perdere la palla dall'immagine. In questi casi è stato imposto al robot di ritornare alla posizione iniziale permettendo una nuova iterazione. A causa dei diversi errori nel calcolo della velocità l'algoritmo impiega più tempo ad annullare lo slip retinale, rispetto al modello in Matlab. L'RLS però converge e predice correttamente permettendo l'inseguimento del target.

Anche nell'implementazione robotica è possibile forzare i parametri precedentemente acquisiti per ottenere una velocità di convergenza migliore. E' dunque possibile implementare un dispositivo che memorizzi i valori dei parametri ottenuti per migliorare la velocità di convergenza in seguito a dinamiche già osservate. Sfortunatamente le problematiche riscontrate nella rilevazione dello slip retinale impediscono al modello di allineare correttamente i parametri di regressione, a causa dell'imprecisione intrinseca che si riscontra nel calcolo di una velocità attraverso la linearizzazione di due punti a due intervalli temporali differenti. Nelle prove mostrate da Schaal e Shibata [45], la piattaforma robotica utilizzata è dotata di 7 gradi di libertà in totale, tre per il collo e due per ciascuna videocamera. Per fornire al robot un sistema di visione ad alta risoluzione e contemporaneamente un'ampia periferia, ogni occhio è dotato di due videocamere. La videocamera ad alta definizione corrisponde alla visione foveale biologica mentre l'altra videocamera analizza il resto dell'ambiente con una risoluzione più bassa. Il calcolo dell'errore di posizione, ovvero della posizione del target sulla retina, e dello slip retinale avviene ad una frequenza di campionamento di 30Hz. L'aumento del numero di immagini per secondo e l'ausilio di videocamere ad alta risoluzione almeno nella parte centrale del campo visivo del robot, permettono una stima dello stato della palla molto più precisa rispetto a quella che è stata ottenuta nelle prove su iCub.

5.2 Conclusioni

Le prove sperimentali mostrate nel paragrafo 5.1 dimostrano che il modello è in grado di seguire perfettamente la dinamica del target. L'andamento dei parametri, inoltre, è paragonabile con i risultati di Matlab-Simulink e quindi è possibile impostare i parametri della traiettoria se questa viene riconosciuta in tempi rapidi. Le considerazioni biologiche affrontate mostrano che la retina umana è uno strumento ancora lontano dall'essere emulato. L'elaborazione visiva di tale organo permette una precisa stima di posizione e velocità dei target presenti nel campo visivo anche con frequenze di aggiornamento non elevate. Al fine di rendere il modello compatibile con i dati biologici occorre dunque considerare una visione olistica di esso all'interno di un sistema di elaborazioni delle immagini complesso e di un sistema di controllo del movimento in grado di

migliorare la propria risposta. Gli sviluppi futuri che emergono da questo lavoro quindi risultano essere separati in due categorie. Dalla parte del sistema di acquisizione video, il modello retinico fornisce utili indicazioni su come progettare sistemi di riconoscimento del movimento che abbiano contemporaneamente ampio spettro di angolo visivo e alta risoluzione centrale. Dal lato del controllo motorio, i movimenti oculari non sono mai del tutto separati, ma si influenzano ed interagiscono continuamente. Risulta dunque importante realizzare l'integrazione del movimento di inseguimento, almeno con il movimento saccadico. Le motivazioni di quanto detto si possono riscontrare in vari aspetti. In primo luogo, le aree motorie che gestiscono le saccadi risultano essere analoghe a quelle per lo smooth pursuit. Altro aspetto è che il movimento di inseguimento ha lo scopo di annullare lo slip retinale e quindi di allineare le velocità del target e dell'occhio, ma non prevede la compensazione dell'errore di posizione. Questo è visibile anche nel modello stesso. Infine l'interazione tra saccadi e smooth pursuit è stata ben studiata sia su soggetti adulti che durante la fase dello sviluppo. Le saccadi hanno un compito di compensazione che permette di riagganciare l'oggetto se c'è un errore nella velocità predetta. Sempre nella fase di controllo l'associazione mnemonica tra riconoscimento di movimento ed esecuzione motoria può essere sviluppata con strumenti più efficienti.

In conclusione questo lavoro si colloca nei settori tipici della bioingegneria con carattere altamente multidisciplinare. Lo studio anatomico e fisiologico del sistema visivo umano permette l'estrazione di modelli matematici che tentano di spiegare la complessità del sistema biologico di riferimento. La validazione dei modelli può essere ottenuta sia per confronto con i risultati sperimentali su soggetti umani e sia nella possibilità di riprodurre il sistema in applicazioni robotiche. La piattaforma robotica utilizzata deve essere considerata come uno strumento sul quale è possibile riprodurre i modelli studiati, tenendo ben presente le profonde differenze strutturali, ma anche le analogie, tra i sistemi biologici e quelli robotici. Capire il funzionamento di alcune delle funzionalità del cervello umano, fino ai compiti più complessi, consiste quindi nella possibilità di riprodurre tale soluzione, risolvendo spesso lo stesso tipo di problematiche, come, in questo caso, i ritardi dovuti al processamento delle informazioni. Nel sistema biologico, come in quello robotico, il ciclo di elaborazione dei segnali fino all'esecuzione dei comandi è troppo lento per essere eseguito da un normale sistema a feedback. La predizione risulta lo strumento fondamentale che consente di reagire in tempo reale alle variazioni dell'ambiente, ed anche di tenere presente la storia di tali variazioni. Da qui nasce una visione della memoria come strumento dinamico, che si aggiorna continuamente, sul quale basarsi per prendere decisioni e non come sistema di pura archiviazione. Il modello proposto in questo lavoro suppone due processi di apprendimento che consentono di predire le variazioni ambientali. Nel primo processo la storia attuale delle informazioni visive consente di adattare il sistema predittivo alla corretta dinamica del target. Nel secondo processo, ciò che è stato appreso in passato non viene perduto, ma utilizzato come base per le esperienze future. Il modello proposto permette infatti di migliorare l'esecuzione dell'inseguimento rispetto ai tempi previsti dai modelli presenti in letteratura. Tale modello ha evidenti riscontri

nelle proprietà neurofisiologiche del cervello dei primati ed i risultati ottenuti permettono un confronto con gli studi del mondo neuroscientifico.

Bibliografia

- [1] P. Dario, S. Martinoia, G. Rizzolatti, G. Sandini. Neuro-robotica. Neuroscienze e robotica per lo sviluppo di macchine intelligenti. Pàtron Editore.
- [2] G. Metta, G. Sandini, V. Tagliasco. *Mente e corpo nei robot*. Bioingegneria della mente. Pàtron Editore.
- [3] R.R. Murphy. *Introduction to AI Robotics*. The MIT Press, Cambridge, MA, 2000.
- [4] L. Fogassi, V. Gallese, L. Fadiga, G. Luppino, M. Matelli, G. Rizzolatti. Coding of perpersonal space in inferior premotor cortex (area F4), *Journal of Neurophysiology*. 76, pp. 141-157, 1996.
- [5] H. Gardner. *Multiple Intelligences: The Theory in Practice*. New York: Basic Books. 1993.
- [6] J. L. Krichmar, G. M. Edelman. Principles underlying the construction of brain-based devices. Paper presented at the AISB 06 Adaptation in Artificial and Biological Systems, Bristol. 2006.
- [7] D. Vernon. The space of cognitive vision. In H. I. Christensen & H.-H. Nagel (Eds.), *Cognitive Vision Systems: Sampling the Spectrum of Approaches*, In press, 7-26. Heidelberg: Springer-Verlag. 2006b.
- [8] R.C. Miall, D.J. Weir, D.M. Wolpert, J.F. Stein. Is the cerebellum a Smith predictor?. *Journal of Motor Behavior*, vol.25, no. 3, pp. 203-216, 1993.
- [9] M. Kawato. Internal models for motor control and trajectory planning. *Current Opinion in Neurobiology*, 9, 718-727, 1999. Elsevier Science Ltd.
- [10] A. Berthoz. *Le sens du mouvement*. Odile Jacob, Paris, 1997.
- [11] R. S. Johansson. Sensory input and control of grip. In *Sensory Guidance of Movements*, John Wiley, Chichester, UK, pp. 45-59, 1998.
- [12] R. C. Miall and D. M. Wolpert. Forward Models for Physiological Motor Control. *Neural Networks*, vol. 9, no. 8, pp. 1265-1279, 1996.

- [13] E. Datteri, G. Teti, C. Laschi, G. Tamburrini, P. Dario, E. Guglielmelli. Expected perception: an anticipation-based perception-action scheme in robots. IROS 2003, 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems, Las Vegas, Nevada, October 27 - 31, 2003.
- [14] E. Datteri, G. Teti, C. Laschi, G. Tamburrini, P. Dario, E. Guglielmelli. Expected perception in robots: a biologically driven perception-action scheme, ICAR 2003, 11th International Conference on Advanced Robotics, Coimbra, Portugal, June 30 - July 3, 2003.
- [15] E. Datteri, G. Teti, G. Tamburrini, C. Laschi, P. Dario, E. Guglielmelli. Expected perception: an experimental application in robotic sensory-motor coordination, Workshop "Embodied Artificial Intelligence - special issues on the design principles", Zurich, Switzerland, October 7th, 2002.
- [16] D. Zambarbieri. Movimenti oculari. Collana di Ingegneria Biomedica n.12. Parton Editore. 2006.
- [17] E. Kowler, R. M. Steinman. The effect of expectations on slow oculomotor control. I. Periodic target steps. *Vision Research* 19, 619-632, 1979.
- [18] E. Kowler, A.J. Martins, M, Pavel. The effect of expectations on slow oculomotor control. IV. Anticipatory smooth eye movements depend on prior target motions. *Vision Research* 24:197-210, 1984.
- [19] G. R. Barnes, P. T. Asselman. The mechanism of prediction in human smooth pursuit eye movements. *J Physiol (Lond)* 439:439±461, 1991.
- [20] C. von Hofsten, K. Rosander. The development of gaze control and predictive tracking in young infants. *Vision Research*, 36, 81-96. 1996.
- [21] C. Von Hofsten, K. Rosander. The development of smooth pursuit tracking in young infants. *Vision Research* 37, 1799-1810, 1997.
- [22] G. Gredeback, C. Von Hofsten, J. Karlsson, K. Aus. The development of two-dimensional tracking: a longitudinal study of circular pursuit. *Exp Brain Res* 163(2):204-213, 2005.
- [23] H. Gronqvist, G. Gredeback, C. von Hofsten. Developmental asymmetries between horizontal and vertical tracking. *Vision Research*, 46, 1754-1761. 2006.
- [24] J. E. Richards, F. B. Holley. Infant attention and the development of smooth pursuit tracking. *Dev. Psychol.* 35:856-67. 1999.
- [25] J. O. Phillips, D. V. Finocchio, L. Ong, A. F. Fuchs. Smooth pursuit in 1- to 4-month-old human infants. *Vision Research*, 37, 3009- 3020. 1997.
- [26] C. von Hofsten. Predictive reaching for moving objects by human infants. *Journal of Experimental Child Psychology*, 30, 369-382. 1980.

- [27] C. von Hofsten, P. Vishton, E. Spelke, Q. Feng, K. Rosander. Predictive action in infancy: tracking and reaching for moving objects. *Cognition*, 67, 255–285. 1998.
- [28] B. Jonsson, C. von Hofsten. Infants' ability to track and reach for temporarily occluded objects. *Developmental Science* 6:1, pp 86–99. 2003.
- [29] E. R. Kandel, J. H. Schwartz, T. M. Jessell. *Principi di Neuroscienze*. Terza edizione. Casa Editrice Ambrosiana.
- [30] C. Constantinidis, M. A. Steinmetz. Neuronal activity in posterior parietal area 7a during the delay periods of a spatial memory task. *Journal of Neurophysiology*, Vol 76, Issue 2 1352-1355, 1996.
- [31] E. K. Miller, C. A. Erickson, R. Desimone. Neural Mechanisms of Visual Working Memory in Prefrontal Cortex of the Macaque. *The Journal of Neuroscience*, August 15, 16(16):5154–5167, 1996.
- [32] S. Yasui and L.R. Young. Control of eye movements, chapter Pursuit eye tracking movements, pages 509– 513. Pergamon Press, Oxford, 1975.
- [33] Robinson, D., Gordon, J., & Gordon, S. (1986). A model of the smooth pursuit eye movement system. *Biological Cybernetics*, 55, 43–57.
- [34] K.J. Bradshaw, I.D. Reid, and D.W. Murray. The Active Recovery of 3D Motion Trajectories and Their Use in Prediction. *IEEE Trans on Pattern Analysis and Machine Intel.*, 19(3), 1997.
- [35] T. Shibata, S. Schaal. Biomimetic smooth pursuit based on fast learning of the target dynamics, *IEEE International Conference on Intelligent Robots and Systems*, 2001.
- [36] T. Shibata, H. Tabata, S. Schaal, M. Kawato. A model of smooth pursuit based on learning of the target dynamics using only retinal signals, *Neural Networks*, 18, pp.213-225, 2005.
- [37] D. Kawawaki, T. Shibata, N. Goda, K. Doya, M. Kawato. Anterior and superior lateral occipito-temporal cortex responsible for target motion prediction during overt and covert visual pursuit. *Neuroscience Research*, 54, 112–123, 2006.
- [38] Giulio Sandini, Giorgio Metta, and David Vernon. Robotcub: An open framework for research in embodied cognition. *International Journal of Humanoid Robotics*, 8(2), November 2004.
- [39] G. Sandini, G. Metta, D. Vernon. The iCub Cognitive Humanoid Robot: An Open-System Research Platform for Enactive Cognition. in *In 50 Years of AI*, M. Lungarella et al. (Eds.), *Festschrift, LNAI 4850*, pp. 359–370, 2007, Springer-Verlag, Heidelberg, in press, 2007.

- [40] A. R. Tilley. *The Measure of Man and Woman: Human Factors in Design*. Henry Dreyfuss, 2001.
- [41] RobotCub web page: <http://www.robotcub.org>.
- [42] R. Beira, M. Lopes, M. Praa, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, R. Saltarn. Design of the Robot-Cub (iCub) head. In volume = A, editor, *International journal on Advanced Robotics Systems*, 3(1):43-48, 2006.
- [43] Ptgrey. Firewire cameras. www.ptgrey.com. 2005.
- [44] ODE webpage: <http://www.ode.org/>
- [45] T. Shibata, S. Vijayakumar, J. Conradt, S. Schaal. Humanoid Oculomotor Control Based on Concepts of Computational Neuroscience. In: *Humanoids 2001*, second IEEE-RAS International conference on Humanoid Robots. Waseda University, Japan, pp 278–285