



UNIVERSITÀ DEGLI STUDI DI MESSINA

Dipartimento di Scienze Cognitive, Psicologiche, Pedagogiche e degli Studi Culturali

DOTTORATO DI RICERCA IN SCIENZE COGNITIVE
XXX CICLO

**Nuovi sistemi cognitivi sintetici di memoria
per l'interazione umano-robot a lungo termine**

Dottorando:

Mario Busà

Coordinatore del dottorato:

Prof. Antonino Pennisi

Tutor e Supervisore di Tesi:

Prof. Giovanni Pioggia

S.S.D: Ing-inf 06

Anno Accademico 2016 - 2017

I – La memoria naturale e artificiale	4
Review sulla memoria biologica ed artificiale.....	4
La memoria nei sistemi biologici.....	7
Memoria nei sistemi sintetici.....	10
Memoria a breve termine.....	12
Memoria Dichiarativa.....	13
Memoria Procedurale.....	14
II – I problemi aperti per l’implementazione dell’interazione uomo-macchina a lungo termine.....	19
Pervasive Memory.....	19
Percezione.....	23
Object segmentation.....	24
Embodiment.....	29
Il problema della Working Memory e le architetture.....	31
III – Implementazione di un’architettura per l’integrazione uomo – macchina tramite social robots	36
L'architettura oggetto della tesi.....	36
Sviluppo dei singoli componenti.....	42
Architettura.....	46
IV – Protocolli utilizzati per l’acquisizione dei dati di apprendimento del robot....	60
Perchè i robot?.....	60
Imitazione.....	67
Contatto oculare.....	75
CAA.....	79
T-Nao.....	81
Server.....	81
Client.....	82
Conclusioni	86
Appendice Tecnica.....	88
PyServer.py.....	88
Client.swift.....	90
NaoServiceBrowser.swift.....	93
Struttura delle classi.....	95
Bibliografia	96

I – La memoria naturale e artificiale

In questo capitolo, sarà affrontato lo studio di sistemi di memoria naturale ed artificiale, aggiornati allo stato dell'arte, che integrandosi nelle più moderne architetture di robot sociali, possa consentire un'interazione umano-robot naturale e costruttiva per un lungo periodo. La modellistica attuale difficilmente integra nei robot sistemi di memoria biologicamente ispirati, che consentano ai robot di intraprendere interazioni autosufficienti credibili. Il problema è complesso: si richiede, infatti, che i robot incorporino la capacità di memorizzare e richiamare esperienze, per imparare da queste e adattare i loro comportamenti sociali sulla base delle esperienze precedenti. Tutto ciò considerando l'ambiente esterno e gli effetti delle proprie azioni.

Sono state analizzate e riportate approfonditamente le ricerche già effettuate in questo settore. Possiamo sicuramente rendere evidente come i sistemi di memoria sintetica costituiscano il substrato attraverso il quale le modalità cognitive interagiscono, ottenendo schemi di comportamento socialmente coerenti a lungo termine. L'esperienza a lungo termine può essere acquisita consentendo al robot di adattare il proprio comportamento sulla storia delle precedenti interazioni.

Review sulla memoria biologica ed artificiale

Nella prima parte sarà svolto un riassunto degli aspetti neuroanatomici della memoria, che porterà a concludere che la memoria negli animali, inclusi gli umani non è affatto monolitica. Un breve excursus della letteratura neuroetologica mostrerà che la funzionalità della memoria non riguarda solo la specie umana. Successivamente verranno passati in rassegna gli approcci sintetici alla memoria, dalle implementazioni delle architetture cognitive alla memoria nelle implementazioni connessioniste per concludere

che in linea di massima la memoria è generalmente implementata come un sistema separato, che agisce come un dispositivo di archiviazione per la cognizione. Per questo motivo la maggior parte delle implementazioni della memoria non escono dal dominio per la quale sono state progettate. Sulla base delle prove neurologiche ed etologiche, sarà illustrato un nuovo modello sintetico, presente nella letteratura attuale, basato su quattro principi:

La memoria è fondamentalmente associativa.

La memoria non richiede la memorizzazione esplicita dell'informazione semantica.

La memoria è una funzionalità distribuita

La memoria supporta l'innescò dei ricordi e dei comportamenti.

In ambito scientifico è pacifica l'idea che la memoria a lungo termine sia un importante aspetto della cognizione. Tuttavia, le questioni relative a ciò che la memoria attualmente è, che cosa fa, e come lo fa, sono state poco indagate, in particolare nel campo dei sistemi cognitivi sintetici.

In termini d'ispirazione biologica per la progettazione del sistema cognitivo, la domanda è se i principi funzionali biologicamente-ispirati (o derivati) correntemente usati sono adeguati a permettere che auspicabili proprietà comportamentali di agenti biologici, come ad esempio l'adattività e la robustezza, siano replicate da sistemi artificiali. In caso contrario è necessario proporre altri principi funzionali (etologicamente validi) sui quali costruire delle architetture cognitive.

La memoria è generalmente discussa in termini di funzionalità, essa può quindi essere descritta in termini più generali come un sistema che permette l'archiviazione d'informazioni, acquisite da interazioni precedenti tra un agente e il suo ambiente, a servizio di comportamenti futuri.

Pur essendo molto generale, questa descrizione implica una serie di ipotesi che si riflettono sia negli approcci di memoria biologica, sia nell'implementazione delle sue funzionalità nei sistemi sintetici. Essa evidenzia la differenza tra il sistema mnemonico e le altre componenti cognitive in virtù delle loro differenti funzioni, dove 'l'archiviazione' è la funzione principale della memoria.

La funzione del sistema di memoria può essere ulteriormente suddivisa in tre distinti processi: codifica (delle informazioni ambientali), raccolta (dell'informazione), e recupero (al servizio del comportamento in atto) (Neath & Surprenant, 2003).

Questo paradigma delle funzioni della memoria è prevalente sia nelle scienze cognitive in generale, sia negli orientamenti inerenti alle architetture cognitive sintetiche, e può essere visto come un'estensione prevalente (ma spesso implicita) della mente come metafora del computer (Bickhard, 2009).

Tuttavia, esiste una crescente corrente di evidenze empiriche che non è compatibile con questo metodo e che sfida le ipotesi dell'approccio '*computational-mind*', rendendolo sempre più inadatto a capire ciò che la memoria è e come funziona (Glenberg 1997; Bar, 2007), nonché inadeguato all'applicazione delle implementazioni nella robotica cognitiva (Baxter, 2010). Per gli agenti artificiali, la funzionalità della memoria è un requisito teoretico e artificiale. Senza (per il momento) impegnarsi con particolari implementazioni computazionali, è necessario sviluppare un sistema di memorizzazione delle informazioni in modo che queste possono essere usate in futuro. Dov'è esplicitamente invocato un sistema di memoria per gestire tali funzionalità, è tipicamente usata una struttura di memorizzazione passiva.

Qualora l'implementazione di un meccanismo di apprendimento sia lo scopo centrale di un'architettura computazionale (usando per esempio alcune forme di strutture *network* sotto simboliche), la memoria è implicita nel fatto che l'addestramento e l'apprendimento precedenti migliorano i comportamenti attuali e futuri. È interessante notare che comunque questo tipo di memoria non è generalmente indicata come tale. Tuttavia, in entrambi i casi, la memoria può essere considerata in un certo senso come presente. Nel dominio d'interazione umano-robot, i requisiti per un sistema di memoria sono più specifici. Oltre all'esigenza di un requisito generale per l'adattamento dei comportamenti ad un'agente interagente (come la direzione dello sguardo, la distanza interpersonale, etc.), è anche chiara la necessità di incorporare informazioni semantiche specifiche o informazioni episodiche (come ricordare un nome, e la circostanza dell'ultima interazione).

Nelle attuali implementazioni computazionali, questo tipo d'informazione è tipicamente codificata in una forma simbolica e tenuta in un modulo di memoria orientato all'archiviazione, spesso invocando la funzionalità di un modulo di 'memoria episodica'. Tuttavia questo approccio all'implementazione della memoria è probabilmente inadeguato in termini di flessibilità e robustezza, poiché il sistema generalmente usa uno schema di codifica specifico per task.

La memoria nei sistemi biologici

La struttura della memoria umana, così come presentata nei libri di psicologia, può essere rappresentata come uno schema a blocchi che evidenzia i vari modi differenti con il quale noi richiamiamo il passato e le informazioni dagli input attuali alle informazioni trovate nel lontano passato. Questa presentazione pulita e ordinata della memoria smentisce la profusione (a volte contraddittoria) di evidenze empiriche dalle neuroscienze e dalla psicologia, e dimostra che la struttura e le funzioni della memoria sono fondate su complesse interazioni tra meccanismi distribuiti attraverso regioni multiple del cervello, alcune delle quali non sono state ancora propriamente comprese. Le funzioni della memoria possono essere scomposte in un'ampia varietà di modi, la cui validità è ancora oggetto di un notevole dibattito. Gli incrementi nella conoscenza ci hanno guidato dalle nozioni di memoria come struttura unitaria verso approcci basati sull'interazione (e differenziazione) di sottosistemi multipli. La memoria umana è generalmente affrontata come se fosse composta di funzioni di codifica, di archiviazione e recupero con l'archiviazione strutturata in sottosistemi scindibili, organizzati in base al contenuto.

Il problema di come identificare le sotto-componenti della memoria e la loro organizzazione rimane un'area di ricerca molto attiva. Questa sezione introdurrà e rivedrà i sottosistemi funzionali della memoria e presenterà brevemente alcuni dei principali modelli psicologici. In seguito la discussione si concentrerà sulla memoria dichiarativa a

lungo termine con un focus particolare sulla funzione episodica, che consente di accedere e richiamare in modo coscivo esperienze passate.

La tassonomia della memoria è sia funzionale sia basata su contenuti con sistemi differenziati, secondo una scala temporale nella misura in cui il contenuto è consapevolmente accessibile e con caratteristiche fenomenologiche ben precise. La divisione principale imposta alla memoria umana separa i processi per il richiamo delle informazioni a breve e lungo termine. A Hebb è accreditata la divisione originale della memoria in sistemi a lungo e breve termine (Hebb, 1949).

Hebb basa la memoria a breve termine (*STM*) su un'attività elettrica transitoria nel cervello, mentre definisce la memoria a lungo termine (*LTM*) consistente di adattamenti neurochimici più duraturi. Questa divisione temporale dei sistemi di memoria è stata anche riflessa nei meccanismi proposti per il '*forgetting*' – il dimenticare, ad es. il *trace-decay* nel *STM* o l'interferenza tra i ricordi memorizzati nel *LTM* (A.D. Baddeley, 2001).

La divisione dei sistemi di memoria a lungo e breve termine ha trovato supporto empirico nelle procedure di dissociazione doppia.

Un'altra distinzione fondamentale nella funzione della memoria è incentrata sulla questione dell'accesso cosciente ai contenuti; il '*knowing how*' di Ryle contro il '*knowing that*' (Squire, 2004). Così, nel *LTM*, l'accesso consapevole della memoria, a fatti e informazioni, è differente dalle abilità sottostanti e dalle capacità non coscienti della memoria. Questi sono stati etichettati rispettivamente come sistemi di memoria dichiarativa e procedurale, da Anderson nel 1976, anche chiamati memoria esplicita e implicita (Graf & Schacter, 1984). La differenziazione della memoria dichiarativa e procedurale ha trovato supporto dagli studi del *priming* e dell'abilità motoria nei pazienti amnesici e anche dalle sperimentazioni con primati non umani. Un successivo movimento lontano dalla semplice dicotomia della memoria a lungo termine è stato poi motivato da un accumulo di elementi probatori che puntano a preservare una collezione apparentemente disparata di funzioni di memoria procedurale nei pazienti con amnesia. Così la divisione procedurale-dichiarativa fu ulteriormente elaborata per fornire un quadro composto da più sottosistemi di memoria.

Il termine ‘non-dichiarativa’ è stato coniato per coprire una serie di ulteriori funzioni inconse di memoria, che comprendono adattamenti a sistemi di comportamento precedentemente appresi. La memoria procedurale, non dichiarativa, viene così suddivisa in quattro tipi: abilità e abitudini, innesco, condizionamento classico e apprendimento associativo. Altre distinzioni nella memoria dichiarativa sono disegnate tra i sistemi di supporto al mantenimento, il recupero di episodi personalmente vissuti e d’informazioni sulla conoscenza del mondo (ad es. il contenuto semantico).

La memoria dichiarativa è stata caratterizzata come la codifica di caratteristiche uniche di un singolo set di input mentre la memoria procedurale generalizza su più set di input basati su elementi comuni (Squire, 2004). Squire suggerisce che l’evoluzione di sistemi multipli di memoria riflette differenti ed incompatibili scopi, ad esempio il graduale apprendimento delle abilità motorie ha dei requisiti differenti dall’apprendimento di una sequenza temporale tramite sensori di input. Infine, gli approcci psicologici e neuropsicologici alla memoria evidenziano anche le divisioni funzionali tra i processi, per rilevare gli stimoli in precedenza incontrati (familiarità) e per il richiamo del contesto in cui questi stimoli sono stati riscontrati (ricordi) Atkinson e Shiffrin proposero un modello a tre stadi della memoria umana (1968). Secondo questo studio, gli ingressi passano dalla memoria sensoriale in una memorizzazione a lungo termine, filtrata dall’attenzione. Gli oggetti nella memoria a breve termine sono mantenuti tramite delle prove o dimenticati e la lunghezza del periodo di tempo in cui gli oggetti rimangono nel STM determinano la probabilità che questi siano trasferiti nel LTM. Il modello è influente ma suggerisce che i pazienti che hanno la STM compromessa dovrebbero anche avere deficit concomitanti nel LTM e nell’elaborazione d’informazioni complesse; una previsione che però non è confermata da evidenze.

La ritenzione nel LTM è stata dimostrata essere dipendente più dalla profondità del processo che dalla lunghezza dei ricordi nel STM (Craik & Lockhart, 1972). Baddeley e Hitch in seguito hanno proposto un modello con strutture simili a quello di Atkinson e Shiffrin ma con l’aggiunta di un modulo di memoria (Working Memory, WM) che ha rimpiazzato il concetto di memoria a breve termine con dei sottosistemi specifici di funzionalità (1974). Nella loro formulazione, questo modulo comprende tre componenti:

il *loop* fonologico di informazioni uditive, il blocco che memorizza le informazioni spaziali e visive e una centrale esecutiva che coordina e controlla l'attenzione. Questo modello è stato probabilmente il primo a rendere esplicita la divisione della memoria (come archivio) e della cognizione (come controllo).

Più tardi Baddeley ha riformulato questo modello per includere un'unità extra, che si occupa di organizzare ingressi multimodali dal *WM*, con contenuti attivati dalla memoria a lungo termine dentro episodi, in modo da mediare i trasferimenti da e per *LTM* (2000). Cowan ha rifiutato la nozione di *WM* come un sistema discreto in favore di un modello in cui il *WM* comprende una sezione attivata del *LTM*. In questo sistema unificato i livelli di attivazione del modello *WM* possono essere equiparati con il contenuto momento per momento della consapevolezza, mediata dall'attenzione (1988). Mentre ciascuno di questi modelli stimola un grande sforzo di ricerca nel tentativo di confermare o diminuire la decomposizione dei sottosistemi proposti, la validità di ognuno di questi schemi non è stata ancora dimostrata.

Memoria nei sistemi sintetici

Che la memoria (la capacità di utilizzare uno stato sperimentato in precedenza, o delle informazioni precedentemente acquisite, al servizio di comportamenti in corso e futuri) sia necessaria per la cognizione è ormai chiaro. Questa funzionalità costituisce pertanto una parte importante delle implementazioni sintetiche dei sistemi cognitivi. Inoltre, da questa caratterizzazione della memoria, diventa esplicito che molti sistemi adattativi sintetici mostrano le funzionalità della memoria - anche quelli che non implementano esplicitamente un sistema di memoria per sé. È quindi intenzione di questa sezione fornire una revisione ed una rivalutazione dell'uso di sistemi di memoria in strutture computazionali sintetiche. Questo dimostrerà che la memoria è diffusa in queste implementazioni secondo le caratteristiche di memoria descritte, anche se non

esplicitamente riconosciute come tali. Lo scopo di questa sezione non è necessariamente di fornire un quadro completo di tutte le implementazioni di sistemi sintetici, ma piuttosto di individuarne le principali categorie, e di riesaminare gli esempi tipici di ciascuna di esse. L'enfasi è quindi sui principi delle funzionalità del sistema di memoria piuttosto che su quelli di progettazione architeturale (Goertzel, Lian, Arel, Garis e Chen, 2010; Largley, Laird e Rogers, 2009; Sun, 2004). Il disegno di una distinzione funzionale e strutturale tra la memoria e la cognizione è derivato principalmente dalla psicologia e parallelamente vi è la tendenza più generale ad approcci modulari alla funzione cognitiva. Questa prospettiva è stata particolarmente influente nel campo della modellazione cognitiva e dell'implementazione nei sistemi cognitivi sintetici, poiché è stato da questo dominio che l'ispirazione per la dissociazione tra l'elaborazione cognitiva e la memoria ebbe origine. Conseguentemente, una divisione intrinseca tra memoria e cognizione rimane in queste implementazioni computazionali, infatti, è stato necessario che fosse dimostrato (Sun, 2004). Inoltre prove comportamentali hanno indicato una divisione della memoria biologica in un numero di componenti funzionalmente distinte, in particolare una separazione tra memoria a breve e lungo termine, e quindi tra diversi tipi di memoria a lungo termine.

Come con la dicotomia memoria / cognizione questa prospettiva si è dimostrata particolarmente influente nei sistemi cognitivi sintetici. Così le implementazioni possono incorporare una serie di funzioni discrete di memoria. Possono essere effettuate due serie principali di distinzioni: (a) la distinzione tra memoria a breve e lungo termine, e (b) la divisione di memoria a lungo termine in memoria procedurale e memoria dichiarativa. Va notato, che mentre una serie di sovrapposizioni può essere descritta tra questi sistemi di memoria, in particolare tra memoria a breve termine e memoria dichiarativa, le divisioni sono giustificate da motivi funzionali (la memoria a breve termine svolge una funzione specifica diversa da quella servita dalla memoria dichiarativa), come descritto sopra. Questo ci porta a tre grandi categorie di sistemi di memoria che coprono lo spazio dei tipi funzionali:

1. Breve termine
2. Dichiarativa / memoria esplicita

3. Procedurale / memoria implicita

Mentre possono essere eseguite altre suddivisioni, questa caratterizzazione grossolana dei sistemi di memoria costituisce una base adeguata per esaminare l'uso (esplicito o implicito) della memoria nelle implementazioni di calcolo, dati i diversi gradi di ispirazione biologica su cui si basano. Nel condurre questo esame, sono state analizzate una serie di caratteristiche d'implementazione che riflettono le ipotesi implicite per quanto riguarda il funzionamento dei sistemi di memoria di agenti biologici, e che possono quindi costituire la base di un'ulteriore rivalutazione.

Memoria a breve termine

La distinzione tra memoria a breve e lungo termine è effettuata per motivi puramente temporali: la memoria a breve termine contiene le informazioni per un periodo limitato di tempo (di solito commisurata a quella del contesto in cui questa informazione è legata), mentre la memoria a lungo termine detiene le informazioni a tempo indeterminato (nonostante il meccanismo ancora mal compreso del *forgetting*). Questa caratterizzazione molto generale può quindi essere applicata a una vasta gamma di sistemi sintetici, se esplicitamente considerati come memoria a breve termine o meno. Funzionalmente, l'idea di appartenenza temporanea è dunque evidente: le informazioni possono essere dentro o fuori dalla memoria a breve termine.

Una prospettiva più orientata alle funzioni sul sistema di memoria a breve termine deriva dalla proposta e dal conseguente studio del *WM*, che ha esteso il concetto di memoria a breve termine nella letteratura psicologica. Il *WM* può essere più generalmente descritto come una funzionale, e spesso strutturata, parte distinta, del sistema di memoria umano che memorizza temporaneamente le informazioni in uno stato accessibile, perché queste possano essere al servizio di un compito da eseguire (Cowan, 1999). E' stato quindi empiricamente esaminato come un sotto- sistema di funzionalità separabili del sistema

cognitivo (AD Baddeley e Hitch, 1974; Repovs & Baddeley, 2006), pur essendo in sé oggetto di ulteriori divisioni funzionali (AD Baddeley & Logie, 1999).

Questa prospettiva sulla memoria a breve termine è molto influente nello studio empirico della cognizione umana, e come tale ha fornito i principi fondamentali di funzionamento, o di giustificazione retrospettiva, per la progettazione di architetture cognitive sintetiche. Le implementazioni di sistemi di memoria a breve termine possono così essere divise in due categorie. Nella prima, i costrutti computazionali sono realizzati con le caratteristiche funzionali esplicite del *WM* psicologico, usualmente come parti di architetture cognitive generali (Langley et al., 2009). In questo tipo di applicazione, la proposta di un dispositivo di memorizzazione temporanea per le informazioni sulle attività rilevanti diventa una struttura simile a un *workspace* (ad esempio, Hayes-Roth, 1985).

Nella seconda categoria, è la proprietà dell'attività transitoria sostenuta che è prevalente, in cui la memoria a breve termine può essere descritta come memoria attiva a lungo termine (Oberauer, 2002). Va notato che le implementazioni che si trovano in questa categoria non sono necessariamente esplicitamente collegate al concetto di memoria a breve termine dai progettisti, ma piuttosto mostrano questa struttura generale.

Memoria Dichiarativa

La memoria dichiarativa può essere definita nei termini del tipo d'informazioni che codifica: nei sistemi sintetici dichiarativi il contenuto delle informazioni che sono direttamente accessibili dal sistema è più spesso simbolico (o basato sui simboli). La distinzione tra la memoria dichiarativa e la memoria a breve termine come definito in questo capitolo riguarda il ruolo funzionale delle informazioni nei due sistemi, in cui il tipo di informazione è del medesimo tipo. Così, mentre la memoria a breve termine è un deposito temporaneo di informazioni dichiarative, il sistema di memoria dichiarativa è il sistema di memorizzazione permanente per queste informazioni. Dove sistemi di memoria

sono esplicitamente inclusi come parte di architetture cognitive computazionali più ampie, questi sono in genere del tipo di memoria dichiarativa. Da questo punto di vista, la struttura del sistema di memoria dichiarativa è quindi di fondamentale importanza, per i fini della funzione, l'efficienza del recupero o l'efficienza del deposito stesso.

La memoria dichiarativa può essere ulteriormente suddivisa in componenti episodiche e semantiche, una distinzione sostenuta da notevoli evidenze empiriche. Questa divisione è evidente anche nelle implementazioni di sintesi. Nei sistemi sintetici di memoria semantica ci si riferisce generalmente alle informazioni simboliche per quanto riguarda gli oggetti e i concetti fisici o astratti, che sono indipendenti dagli incontri specifici, o dall'esperienza, con tali informazioni. Al contrario, la memoria episodica si riferisce a un evento con cui si ha avuto esperienza, e quindi ha un contesto esplicito, spaziale e temporale (cioè che cosa, dove, quando) con le informazioni codificate in esso (Tulving, 2002). La Memoria episodica nell'uomo ha il richiamo cosciente come una proprietà associata necessaria, anche se questo requisito è stato aggirato nella definizione di memoria simil-episodica per gli animali.

Memoria Procedurale

La memoria procedurale, in termini più generali, può essere definita in opposizione alla memoria dichiarativa: essa si preoccupa della conservazione delle informazioni concernenti le informazioni accessibili inconsciamente. Ad esempio, le capacità motorie (come andare in bicicletta, o a piedi), o le competenze di un dominio specifico (come ad esempio il riconoscimento di un oggetto, o di un suono), possono essere considerate come risiedenti all'interno del sistema di memoria procedurale. In termini d'implementazioni sintetiche questa definizione comprende quindi una vasta gamma di lavori, tra cui gli approcci generali di apprendimento delle reti neurali artificiali. È necessaria, come tale, un po' di reinterpretazione di esempi di tali lavori in questo contesto. Tuttavia, lo scopo

di tale reinterpretazione non è alterare lo scopo originario o l'intento dell'implementazione discusse, ma è unicamente favorire un'analisi applicabile al dominio dei sistemi di memoria. Un certo numero di tipi di memoria procedurale può essere pertanto distinto.

Data la definizione generale della memoria sopra proposta (la memoria prevede l'utilizzo d'informazioni acquisite attraverso l'esperienza passata per il comportamento presente e futuro), le funzionalità di apprendimento di reti neurali artificiali di piccole dimensioni, per l'acquisizione di conoscenze basate sull'esperienza, nel campo delle architetture ispirate alle neuroscienze, possono essere incluse nella categoria dei sistemi di memoria procedurale / implicita, oltre ai sistemi in cui i costrutti codificano esplicitamente un processo basato su regole. Inoltre, prendendo come requisito primario la storia di una precedente interazione, le implementazioni nel dominio dello sviluppo robotico (in particolare quelli concernenti l'acquisizione della coordinazione sensori-motore) diventano rilevanti.

In alcuni sistemi basati su regole, i costrutti computazionali espliciti (regole) sono usati per codificare le coppie condizione-azione, come ad esempio per l'apprendimento dei sistemi classificatori, (per esempio Bull, 2004; Bull & Kovacs, 2005), o dei sistemi esperti classici (ad esempio Duda & Shortlie, 1983). Queste regole possono essere generate al *run-time* del sistema, e codificano anche le statistiche riguardanti le prestazioni, e, quindi, possono essere descritte come un tipo di memoria procedurale. Esistono strutture simili in architetture cognitive generali, come ad esempio i moduli di memoria procedurale di *Soar* (es Derbinsky & Laird, 2010), *ATC-R*. (ad esempio, JR Anderson et al., 2004), o *CLARION* (Sun, 2007), dove le regole condizione - azione sono utilizzate per rappresentare certi tipi d'informazioni. Le Reti neurali artificiali sono tipicamente addestrate per imparare le mappature tra un *set* di determinati input per un diverso insieme di uscite. Prendendo ad esempio un *perceptron* multistrato, lo scopo del processo di formazione sarebbe modificare i pesi dei collegamenti nella rete in modo tale che l'uscita desiderata venga prodotta per ogni dato di ingresso (ad esempio la classificazione di un ingresso in una serie di pre-classi definite). In termini più generali, un segnale di errore è utilizzato per fornire la base per la procedura di aggiornamento del peso su ogni iterazione

addestrata. Il *perceptron* multistrato codifica così quello che può essere descritto come memoria implicita: l'effetto cumulato delle iterazioni di addestramento porta a un sistema in grado di applicare le informazioni apprese ai comportamenti futuri. Allo stesso modo, nelle mappe auto-organizzanti (Kohonen, 2001), come esempio di apprendimento non supervisionato, la classificazione è ottenuta attraverso l'uso dei dati di addestramento del campione. Mentre questi esempi sono alquanto semplicistici, l'analogia si estende sia ad altri tipi di reti neurali artificiali, sia a sistemi in cui il concetto di apprendimento è più facilmente applicato alla memoria.

Più in conformità con il concetto che la memoria sia acquisita attraverso l'esperienza (che corrisponde all'esperienza di un agente attraverso l'interazione con l'ambiente e / o altri agenti), alcune implementazioni di reti neurali artificiali consentono l'adattamento senza una fase separata di addestramento esplicito. In queste implementazioni, il ruolo della memoria implicita è più chiaramente visibile. In (Wyss, König, e Verschure, 2006), una rete neurale gerarchica, in cui le unità neurali computazionali sono identiche in ciascuno dei cinque strati, l'input è formato dalle informazioni sulle immagini che provengono da una telecamera montata su un robot mobile che compie movimenti casuali in un ambiente chiuso. Utilizzando solo due principi di calcolo (stabilità temporale e memoria locale), i livelli della rete auto-organizzante dentro ruoli funzionali distinti, con lo strato superiore (cioè quello più lontano dall'input dei dati sensoriali) che mostra un comportamento simile al posizionamento delle cellule, (ad esempio Poucet, Lenck-Santini, Paz-Villagran, & Save, 2003), dove l'esperienza guida questa emergenza. Il posizionamento delle cellule è stato implicato centralmente nei compiti di navigazione spaziale (es Burgess, Donnett, Jeery, e O'Keefe, 1997; Poucet et al., 2003), questo pone la memoria procedurale in un ruolo importante. In modo simile, varie implementazioni di mappe auto organizzanti sono state create che per adattarsi alle informazioni d'ingresso (Lang, 2007), o la creazione e l'utilizzo di associazioni tra più mappe (Johnsson & Balkenius, 2009; AF Morse, Gree, Belpaeme, e Cangelosi, 2010; Luciw e Weng, 2010). Inoltre, i meccanismi implicati nella memoria procedurale, come il *priming* o il condizionamento sono anche visti nelle implementazioni strutturali. Il *Priming* in questo contesto si riferisce all'attivazione avviata come risultato delle associazioni ai correnti *input* di ingresso, basati sulla

precedente esperienza di una relazione tra i due, e il condizionamento è il processo in cui una condizione a volte sperimentata porta ad un comportamento stereotipato. In (A. Morse & Aktus, 2009), sia il Priming che il condizionamento sono dimostrati in un sistema robotico controllato utilizzando una rete associativa psicologicamente plausibile. Analogamente, l'apprendimento di un sistema complementare è stato utilizzato per modellare la memoria basata sul riconoscimento e per la rapida codifica delle relazioni arbitrarie tra rappresentazioni, come l'ippocampo (O'Reilly, Braver, & Cohen, 1999; O'Reilly & Norman, 2002).

I modelli di sistemi nervosi biologici forniscono esempi di sistemi integrati che incorporano una vasta gamma di elementi di prova neurofisiologica e neuropsicologica. Questi modelli incorporano tipicamente alcune mappature, grossolane a livello strutturale, a un certo numero di regioni funzionalmente distinte del sistema nervoso animale. Ad esempio, una serie di dispositivi basati sul cervello, (Krichmar & Edelman, 2002; Krichmar, Nitz, Gally, e Edelman, 2005; McKinstry et al, 2008) implementa simulazioni su larga scala di più regioni del sistema nervoso dei mammiferi a dimostrare non solo come la funzionalità implicita della memoria procedurale sopra descritta è istanziata, ma anche come questo possa costituire la base di comportamenti più complessi. Altre implementazioni computazionali di tipo simile (integrando più strati di prove empiriche) forniscono tentativi analoghi non solo al modello di tali sistemi biologici, ma anche per consentire la generazione di comportamenti adattivi in agenti sintetici: es l'architettura *GNOSYS* (Taylor et al., 2009), è un modello basato sui gangli applicato ad un robot mobile (Prescott, Montes Gonzalez, Gurney, Humphries, e Redgrave, 2006). In ciascuno di questi esempi, l'apprendimento è l'elemento centrale, ma un sistema di memoria è implicito nel fatto che le associazioni apprese precedentemente sono la base per un comportamento adattativo. Tornando alla caratterizzazione della memoria procedurale come richiedente intrinsecamente un resoconto della storia d'interazione di un agente con l'ambiente, il campo dello sviluppo robotico diventa rilevante. Quest'approccio ad agenti cognitivi sintetici impiega una vasta gamma di metodologie in cui il ruolo della memoria non è reso esplicito (Weng, 2007), con alcune eccezioni, come ad esempio la memorizzazione di storie di interazione in (Mirza, Nehaniv, Dautenhahn, e Boekhorst, 2007). Nel contesto

specifico dello sviluppo della coordinazione senso-motoria, un certo numero di implementazioni si sono allontanate dall'approccio ispirato al sistema neurale facente uso di rappresentazioni esplicite di associazioni, (ad esempio, Baxter e Browne, 2009; Lee, Meng, e Chao 2007; Quinton & Buisson, 2008).

Questa natura esplicita della rappresentazione delle associazioni, consente che l'interpretazione basata sulla memoria diventi chiara: l'interazione (e quindi l'esperienza da parte dell'agente) consente alle associazioni di essere formate (come la memoria), e questa struttura è successivamente responsabile del comportamento dell'agente.

II – I problemi aperti per l'implementazione dell'interazione uomo-macchina a lungo termine

In questo capitolo è stato effettuato lo studio, aggiornato allo stato dell'arte, degli ostacoli che si presentano nell'utilizzare l'interazione umano-robot in ambienti *open* e come la memoria a lungo termine possa essere di supporto nell'affrontare tali problematiche. Nello specifico, la prima parte di questo capitolo esaminerà il ruolo pervasivo della memoria, in particolare verrà introdotto un nuovo concetto, secondo cui la memoria è coinvolta in tutti gli aspetti dei comportamenti sociali, andando oltre la pura memorizzazione di informazioni semantiche.

Nella seconda parte analizzerò le difficoltà che si hanno nel portare l'interazione nel mondo reale e in ambienti *open* e allo stesso modo le opportunità che si possono sfruttare. Infine lo studio sarà focalizzato sullo sviluppo percettivo nei neonati, come il loro giudizio percettivo cambi con il passare del tempo e come la conoscenza basata sulle esperienze cambi questo giudizio, verrà quindi fatta una correlazione tra bambini e robot in modo da poter comprendere come essi possano imparare a riconoscere l'ambiente senza averne una conoscenza a priori.

Pervasive Memory

La memoria come definita in precedenza, può essere vista come un meccanismo (e non come un *repository* d'informazioni passive) che codifica l'esperienza in modo da facilitarne il susseguente coinvolgimento in tutti i processi cognitivi, incluse quelle competenze richieste per i comportamenti sociali.

Il primo obiettivo di queste funzionalità è l'interazione sociale a lungo termine. Per raggiungere questo scopo sono richiesti vari gradi di adattività (Goodrich and Schultz, 2007). Tuttavia, nei recenti lavori che coinvolgono l'interazione sociale tra bambini e robot è emerso che questo processo di adattamento lavora in entrambe le direzioni (anche i bambini adattano il loro comportamento in funzione del robot), e che persino un adattamento superficiale, come la conoscenza del nome dell'agente interagente o la conoscenza di un'interazione precedente, ha profondi effetti.

Per l'*HRI* (*human robot interaction*) la memoria è quindi pervasiva. Mantenendo la generalità dell'esempio, l'adattività si basa sulla capacità di permettere alle esperienze precedenti di influenzare comportamenti presenti e futuri (Belpaeme et al, 2012).

La memoria può essere quindi ridefinita come un supporto necessario per l'adattamento del *social robot* ai suoi partner d'interazione. Nel caso semplice in cui si deve riconoscere che un'interazione precedente abbia avuto luogo, un *flag* booleano sarebbe sufficiente, ma si tratterebbe di una codifica molto specifica per caso. Un passo avanti è prendere la nozione dell'adattività su periodi di tempo estesi, considerando il fenomeno più specifico di allineamento del comportamento, dove un individuo modifica il suo comportamento contingente su un comportamento del suo partner d'interazione e viceversa (De Jaegher et al. 2010). Questo effetto è stato osservato su molteplici modalità d'interazione, incluse il linguaggio (Branigan et al, 2010), il dialogo (Pickering and Garrod, 2004), l'espressività non verbale (Yamazaki, 2012). Esso sembra essere una feature fondamentale dell'interazione sociale, ed è quindi fondamentale che un robot sociale la faccia sua. Consideriamo il ruolo di un robot di prossimità (Torta et al., 2013): trovare un'appropriata distanza per un utente necessita una differenziazione, poiché non necessariamente questa distanza andrà bene per altri utenti, e richiede che la memoria conservi quest'informazione per incontri futuri. Come per il primo esempio, la soluzione è codificare le esperienze precedenti, ma anche restringerle al singolo contesto. In aggiunta all'interazione stessa, vi è un numero di competenze cognitive di supporto che è necessario considerare. Per la *long-term HRI*, questi aspetti devono essere presi in considerazione poiché sono alla base dei comportamenti che vanno oltre la mera reattività. Per esempio, ci sono delle informazioni necessarie per garantire il funzionamento della memoria episodica, come la

memorizzazione di una specifica sequenza di eventi che sono occorsi in un dato momento (Clayton and Russell, 2009). Inoltre l'abilità di imparare, processare e applicare concetti (sia fisici che astratti) è un'importante abilità cognitiva. La necessità della memoria per ciascuna di queste funzioni è chiara. Esse possono essere implementati tramite un database: tipicamente, ad esempio, la memoria episodica è codificata utilizzando *timestamps* su una chiave per ogni istanza, queste fragili rappresentazioni (in termini di generalità) però sono in contesti più generali sconvenienti.

Gli esempi sopra descritti illustrano i requisiti per la memoria nelle specifiche sfaccettature del *HRI*, anche se potremmo estenderli a tutti gli aspetti della cognizione (Wood et al., 2012). Possibili mezzi per archiviare le informazioni precedenti da poter usare per questi esempi sono stati suggeriti (stati booleani, *lookup tables* e *databases*), ma queste potenziali soluzioni sono specifiche per *task*, adempiono alle funzioni desiderate ma non facilitano la generalizzazione per nuovi tipi di *behaviour*. Possiamo distinguere tre domini distinti:

- Percezione:
 - Es: la *face recognition*
- *Decision making*
 - Es: l'uso di conoscenze precedenti per decidere un argomento di conversazione
- Azioni
 - Es: approcci appropriati

Tuttavia, le implementazioni del *HRI* tipicamente considerano solo uno o due di questi domini. Dato che la memoria è intrinsecamente coinvolta in ognuna di queste funzioni, è auspicabile avere una quantità di dati in memoria che può oltrepassare questo problema e per raggiungere questo obiettivo, occorre avvalersi della conoscenza di varie discipline. Partendo dalle neuroscienze (Bar et al., 2007), e dai fondamenti filosofici (Riegler, 2005), è stata formulata un'ipotesi basata su un *framework* centrato sulla memoria di processi cognitivi nel contesto della robotica cognitiva. Il principio di questo *framework* è che la memoria stia alla base di un singolo tipo di strato computazionale, e che questo rimane

consistente attraverso i vari contesti applicativi. Questo sottostrato è essenzialmente associativo e distribuito, riflettendo così due *features* centrali del sistema cognitivo umano (Fuster, 1997). La memoria può tuttavia essere considerata indipendente da un *task* specifico di un contesto, con un *set* di meccanismi che la costituisce come il punto di partenza per la progettazione di un sistema (Baxter et al. 2011). L'implementazione computazionale studiata è basata su un modello di calcolo parallelo distribuito, che utilizza uno schema di rappresentazione della conoscenza locale.

Quest'implementazione codifica l'esperienza dell'agente in una serie di *link* associativi che sono incrementalmente acquisiti e aggiornati. Usando il risultante network associativo come substrato per le dinamiche di attivazione, le informazioni codificate possono essere 'richiamate' o 'processate' attraverso la riattivazione delle associazioni. Essendo uno schema di rappresentazioni sub-simboliche distribuite, comunemente agli altri approcci connessionisti, un naturale effetto è la presenza di un grado di generalizzazione sia nel processo di codifica sia in quello di recupero. L'ipotesi che si vuole dimostrare è che se un robot allinea il suo comportamento al suo *partner* d'interazione, l'essere umano tenderà a rimanere coinvolto nell'interazione più a lungo rispetto a quando il robot non adatterà il suo comportamento. Questo punto di vista offre una nuova prospettiva, nella quale l'interazione cognitiva è prodotta da due agenti, dove non è necessario che uno di essi sia pienamente cognitivo. In altri termini la cognizione sociale è il prodotto di un'interazione ad alto livello tra due o più agenti e quando un agente impoverisce l'interazione, il *gap* è spesso riempito dalla cognizione dell'altro agente.

Si può prendere come esempio l'interazione verbale. La conversazione vocale non segue le stesse regole della conversazione scritta, eppure nonostante le interruzioni, gli errori e le esitazioni queste non influenzano negativamente l'interazione. La ragione per cui questo avviene è che le conversazioni sono il prodotto di due agenti, il quale combinando le loro cognizioni portano l'interazione avanti.

Questo ha un tremendo potenziale per l'*HRI* sociale. Le capacità di un robot sono ben lontane dalle capacità cognitive umane, nonostante questo, poiché la *social cognition* è il prodotto di due agenti, i fallimenti del robot possono essere coperti dall'essere umano. Questo ovviamente può accadere anche nell'interazione umano-robot. Ad esempio il

deficit che un robot possa non avere una percezione visiva, può essere sottovalutato, basta che l'essere umano creda che il robot possa vedere per far sì che questa lacuna nell'interazione possa passare inosservata.

Percezione

Per la maggior parte dei sistemi intelligenti la percezione resta un collo di bottiglia, è nel *HRI* è certamente così; Gli utenti spesso si aspettano che il robot abbia le stesse capacità percettive che loro stessi hanno, è queste funzionalità nella storia si sono mostrate molto difficili da raggiungere. Ad esempio nella percezione visiva ci sono voluti cinquanta anni di ricerca per raggiungere dei risultati che rappresentano una minima parte delle capacità umane. Funzionalità come la *Face detection*, la *Face recognition*, *l'object recognition*, la separazione delle figure e la comprensione dei comportamenti umani, hanno percorso un lungo cammino e tuttavia possono essere usate solo in scenari ristretti. Un'interazione aperta in un ambiente reale non è ancora possibile. Lo stesso vale per la *speech recognition*, ma mentre questa negli ultimi anni con l'acquisizione di grandi moli di dati ha raggiunto elevati livelli di efficienza, lo stesso non si può dire per la *child speech recognition* che rappresenta ancora una sfida e che forza l'interazione umano-robot ad usare il linguaggio naturale in situazioni limitate.

Un altro ostacolo all'interazione autonoma è la selezione delle azioni: il problema di cosa fare in risposta ai passati e ai correnti stimoli provenienti dai sensori di input. Gli utenti hanno aspettative ben definite di quale la risposta dovrebbe essere, ma è tecnicamente molto difficile selezionare una risposta corretta in un ambiente aperto e non vincolato. Il meccanismo di selezione delle azioni e le architetture cognitive non sono ancora all'altezza quando vengono usati nel dominio delle interazioni sociali. Vi è quindi bisogno di nuovi e radicali approcci. Comunque come già scritto in precedenza, le lacune negli agenti intelligenti possono essere colmate dai giovani utenti. Ad esempio la mancanza di

percezione visiva può essere colmata fornendo ai robot *behaviours* che fanno sembrare come se questo stesse vedendo e l'aggiunta di due occhi può essere un trucco che rafforza l'utente a credere che il robot veda. Tecnologie come lo *scanner laser* o i *tags RFID* possono rivelare gli oggetti vicini e generare l'illusione che il robot possa vedere. Diventa quindi fondamentale fornire ai robots la capacità di riconoscere l'ambiente in cui si muovono e gli oggetti che lo circondano.

Object segmentation

Gli sviluppi tecnologici avuti in questi anni nella sensoristica e nella motoristica rendono possibili nuove competenze richieste per l'interazione uomo – robot, tuttavia bisogna comprendere non solo come queste competenze debbano essere coordinate, ma anche come un sistema robotico deve utilizzarle per facilitare l'interazione in domini multipli, ad esempio come relazionarsi con gli umani a livello sociale e come mantenere questa relazione nel tempo, in ambienti e *task* differenti. Le ricerche in questo campo non hanno solo lo scopo di migliorare le abilità dei sistemi sintetici (Belpaeme et al., 2012), ma anche di inserire quest'ultime nei fenomeni naturali che come ricercatori stiamo provando a replicare, peraltro spesso questi fenomeni, come nel caso della memoria a lungo termine, non sono stati pienamente compresi.

Questa enfasi sulla generalità delle applicazioni nei domini multipli ci porta a fare delle considerazioni sui principi generali dell'interazione sociale e dei suoi meccanismi sottostanti che possono essere implementati dal sistema di controllo del robot (Sun, 2004; Langley et al., 2009), piuttosto che dalle competenze in un singolo dominio. Per l'*HRI* l'applicazione di architetture cognitive è stata proposta come un mezzo in grado di fornire ai robot modelli più accurati di interazione, in modo da essere in grado di migliorare il loro comportamento non sempre ottimale (Trafton et. al, 2013), una competenza necessaria per avere comportamenti cognitivi simili a quelli umani (Holland et al., 2013).

L'obiettivo è applicare architetture cognitive nel *HRI*, e per farlo bisogna migliorare le funzioni cognitive dei robot al servizio dei comportamenti *human-oriented*. La rivalutazione della memoria con le architetture cognitive sintetiche porta a un'integrazione molto vicina tra memoria e cognizione, dove la cognizione dovrebbe essere considerata come il risultato delle dinamiche di attivazione sul substrato associativo che è la memoria (Baxter & Browne, 2011; Wood et al., 2012). Nonostante il *focus* su cosa può essere considerato meccanismo a basso livello, questa prospettiva può essere applicata su questioni considerate di alto livello come l'interazione sociale (Baxter et al., 2011). Per incorporare questi principi si può sviluppare un sistema sperimentale di acquisizione di concetti, dove l'applicazione possa fornirne una spiegazione coerente con i dati umani, senza l'introduzione di meccanismi specializzati.

Per far ciò bisogna considerare due aspetti: una valutazione centrata sull'essere umano e la capacità del robot di incorporare dati. Questo implica che il robot deve essere controllato attraverso un'architettura reattiva che interagisce con gli esseri umani e registrarne i risultati attraverso le varie modalità con cui i sensori percepiscono il mondo.

I robot per essere utili e in grado di stabilire una fiducia con gli esseri umani devono avere una forte capacità di riconoscere gli oggetti e la rappresentazione del mondo circostante. È impossibile pensare che un robot possa conoscere in anticipo tutti gli oggetti con cui dovrà interagire, quindi avrà bisogno di essere in grado di imparare ogni volta nuovi oggetti; idealmente dovrebbe essere in grado di ottenere tutte le informazioni senza complessi meccanismi di training, ma semplicemente attraverso l'interazione con gli esseri umani o con gli oggetti stessi.

La maggioranza degli approcci esistenti di *computer vision* per il riconoscimento degli oggetti è basata su un database iniziale e varie forme di conoscenze pregresse o dalla presenza di target come rilevatori di pelle, faccia, colori. Questi approcci però non permettono ai robot di essere autonomi in un ambiente *open*. Per questo uno degli obiettivi di questa ricerca è sviluppare un sistema basato sull'apprendimento incrementale dove il robot avrà delle capacità di base e potrà acquisire nuove capacità di difficoltà sempre crescente. Il sistema deve essere in grado in ogni scenario di imparare a riconoscere gli oggetti, indipendentemente da tutte le differenti condizioni che si possono verificare in un

ambiente. L'idea di un robot umanoide che acquisisce la sua conoscenza del mondo da un'esplorazione graduale e dall'interazione con l'ambiente è ispirata dal modo in cui i bambini imparano gli oggetti. Tuttavia, la maggior parte dei *challenges* che esistono nel campo della *computer vision* sono risolti dai bambini durante i primi anni della loro vita. Iniziando con una conoscenza basilare, o addirittura senza nessuna conoscenza, i bambini recuperano tutte le informazioni dall'intensità della luce o dai suoni. Loro imparano dall'accumulo iterativo di dati, di associazioni e dai risultati delle loro azioni. L'auto sperimentazione inizia con l'esplorazione del loro corpo, come l'associazione dei loro arti con il loro aspetto e il movimento del loro corpo. Questo processo è chiamato '*body babbling*' ed ha lo scopo di ottimizzare la prevedibilità dei loro movimenti. Quando un certo progresso è raggiunto, la fase di auto apprendimento si sovrappone all'esplorazione del mondo esterno, agli oggetti e le persone circostanti. L'interazione con gli oggetti è importante poiché rafforza l'esplorazione e fornisce agli oggetti delle proprietà fisiche addizionali, oltre a fornirne la visione completa da differenti prospettive. Per realizzare un sistema di memoria a lungo termine occorre poter avere una rappresentazione visiva del mondo e quindi avere un approccio generale in grado di poter riprodurre gli *step* di sviluppo sopra descritti. Questo sistema deve essere in grado di riconoscere e imparare gli oggetti senza supervisione. Gli sviluppi fatti fino ad oggi nel campo della *computer vision* forniscono una grande quantità di *object detection* e di algoritmi di riconoscimento che sono però basati su una conoscenza e priori. Nel campo *dell'imaging processing*, la conoscenza a priori prende la forma degli algoritmi scelti o del *database* delle immagini, dove ogni oggetto è associato a molte immagini o a proprietà visuali codificate dai descrittori. In questo caso, il riconoscimento degli oggetti si basa sulla similarità con le entità esistenti nel database; Questi sistemi sono veloci e realizzabili, ma non sono facilmente applicabili per l'apprendimento autonomo, poiché questo richiede al robot l'abilità di costruire *online* una rappresentazione degli oggetti adattabile all'ambiente. Esempi esistenti sono la *face detection* che lavora con delle *features Hair-like*, o lo *skin-detector* che processa il colore dei *pixels*.

Ci sono molti metodi per rappresentare gli oggetti. Spesso, piuttosto che analizzare l'intensità dei *pixel*, sono usati dei descrittori locali in delle posizioni salienti, in tal modo

le immagini sono processate più velocemente e più efficacemente. Il contenuto visuale viene così codificato con una diminuzione importante del numero di dati processati. Il mondo attorno a noi ha una struttura complessa, e per un adulto questo è formato da oggetti più o meno ben definiti. Percepire il mondo in questo modo sembra banale, ma da una prospettiva ingegneristica è un problema molto complesso. Nel 1990 Elizabeth Spelk scriveva:

the ability to organize unexpected, cluttered, and changing arrays into objects is mysterious: so mysterious that no existing mechanical vision system can accomplish this task in any general manner.

Nonostante siano passati venticinque anni questa affermazione è ancora d'attualità. L'abilità di riconoscere i confini degli oggetti presenti nella scena (*'object segregation'* in psicologia o *'object segmentation'* in ingegneria) non può essere ancora automatizzata con successo per oggetti presenti in ambienti *open*.

Per quanto concerne la ricerca ingegneristica ci sono stati dei progressi algoritmici; Per esempio date delle misure locali di similarità tra elementi vicini di una scena, un appropriato *set* di confini può essere dedotto in modo efficiente (Shi & Malik, 2000). C'è anche una crescente consapevolezza nell'importanza di raccogliere e sfruttare la conoscenza empirica circa la combinazione statistica di materiali, ombre, luci e punti di vista che sono presenti nel nostro mondo (Martin et al., 2004). Certamente tale conoscenza può essere catturata ed efficacemente usata grazie ai progressi fatti in *machine learning*, ma la conoscenza di per sé non è specificata da nessun algoritmo.

Empiricamente, la conoscenza non algoritmica gioca un ruolo chiave nella percezione di un robot. Per esempio la *face recognition*, la *speech recognition* hanno avuto un grande miglioramento dovuto all'aumento esponenziale della loro conoscenza (dovuta alla possibilità di fornire training set sempre crescenti). Gli algoritmi di *machine-learning* sono in grado di approssimare il *mapping* dai sensori di *input* all'interpretazione *dell'output* desiderato, basandosi sui *training set* e applicando queste approssimazioni alle nuove situazioni che vengono presentate.

Quest'approssimazione è influenzata dalle proprietà che il programmatore ritiene rilevanti. Studi iniziali hanno dimostrato che i neonati usano un insieme di *features* per analizzare ciò che vedono. Studi successivi suggeriscono che la forma degli oggetti è la *feature* principale che i neonati utilizzano per identificare i confini tra oggetti adiacenti. Questo metodo può portare a delle analisi errate ma può anche fornire delle ragionevoli interpretazioni di ambienti complessi. I neonati, infatti, non riescono a segregare gli oggetti dalla nascita ma imparano con il tempo a utilizzare le caratteristiche degli oggetti per riuscire a distinguerli (Needham & Baillargeon, 1997). Successivamente, degli esperimenti hanno dimostrato che i bambini sono in grado di analizzare le differenze percettive tra superfici adiacenti e segregare le superfici con caratteristiche differenti (Needham, 1999). Intorno agli otto mesi i bambini hanno dimostrato di essere in grado di utilizzare le informazioni specifiche di un oggetto o d'interi classi per guidare il loro giudizio (Needham et al., 2006). Una grande quantità di dati è utilizzata per fare da *training* su questo sistema. Tuttavia, è stato investigato che l'esposizione a degli eventi chiave potrebbe fornire ai neonati dei parametri di apprendimento per la segregazione degli oggetti.

Anche se quest'idea può sembrare intuitiva ed evidente per gli esseri umani, non è ancora studiato come possa essere usato nei sistemi meccanici.

Un esperimento del 2005 usò un semplice evento chiave per far sì che il robot potesse imparare un oggetto. Il robot portando davanti alla propria telecamera l'oggetto si diede un ottimo punto di vista per apprenderlo e in seguito segregarlo dall'ambiente (Natale et al., 2005).

In una situazione reale è molto difficile per un agente robotico isolare un oggetto e poterlo muovere può portare dei grandi vantaggi. Le classi degli algoritmi per il riconoscimento degli eventi possono essere così suddivise:

- Riconoscimento di eventi chiave
 - algoritmi che riconoscono l'evento interessato, come quando il robot colpisce l'oggetto.
- Machine Learning:

- algoritmi che usano i *training set* per identificare le *features* che possano riconoscere gli oggetti e che possano essere usate in altre situazioni.
- Applicazioni di apprendimento:
 - Algoritmi che utilizzano queste *features* per riconoscere gli oggetti.

Embodiment

Lo studio della percezione nei sistemi biologici non può negare il ruolo del corpo e della morfologia nella generazione delle informazioni sensoriali che raggiungono il cervello. Uno dei più grandi passi avanti fatti in neurofisiologia durante l'ultimo ventennio è che l'unità di base del controllo non è l'attivazione di un muscolo specifico ma piuttosto un'azione che include un obiettivo, un motivo per agire, specifiche modalità di percezione su misura per quest'obiettivo e la ricombinazione di moduli funzionali e di sinergie per ottenerlo.

Una prospettiva moderna del *motor control* biologico considera multipli *controllers* come *goal specific*. Questo particolare tipo di generalizzazione è cristallino nell'area premotoria che correla l'atto dell'afferrare. Quest'area chiamata *F5* (area 5 frontale) contiene neuroni che sono usati per afferrare con la mano sinistra, con la mano destra o persino con la bocca (Gallese et al., 1996). Il successivo *step* concettuale nel cambiare la nostra visione del movimento fu fatto grazie alla scoperta dei neuroni sensoriali (es. visivi) nella stessa corteccia premotoria, area *F5*. Le due rappresentazioni, motorie e visuali, non solo coesistono nella stessa area del cervello ma anche coesistono nella stessa popolazione di neuroni. Simili responsi sono stati trovati nella corteccia parietale. Questo forma una forte connessione bi-direzionale con la corteccia premotoria che è utile per parlare del sistema fronto-parietale. Neuroni parietali sono stati trovati per rispondere alla *features* di oggetti geometrici (come il loro orientamento 3D). I neuroni multi sensoriali sono responsabili sia dell'azione che della percezione, corpo e cervello sono strettamente intrecciati per

plasmarsi l'un l'altro durante lo sviluppo e attraverso l'età adulta. Attraverso il corpo, il cervello esegue azioni per esplorare l'ambiente e raccogliere informazioni circa le sue proprietà e regole. Nello sviluppo iniziale l'esplorazione del mondo avviene attraverso gli occhi, le mani e la bocca. Le primissime competenze vengono acquisite attraverso gli occhi (Haith, 1980). Con l'esperienza la visualizzazione diventa più comprensiva e si focalizza sulle *features* più significative. I bambini spesso focalizzano, per lunghi periodi, la loro attenzione sulle loro mani, infatti, loro utilizzano le capacità del loro corpo per riconoscere gli oggetti sia con gli occhi, che con bocca e mani.

Negli anni sono stati condotti vari esperimenti a tal proposito ed è chiaro da questi che lo sviluppo motorio e l'uso del proprio corpo giocano un ruolo importantissimo nello sviluppo percettivo dei bambini. Nei robot è possibile studiare i collegamenti tra azione e percezione, e le implicazioni sulla realizzazione dei sistemi artificiali. I robot come i bambini, possono esplorare fisicamente gli ambienti per arricchire e controllare le loro esperienze sensoriali, ma come i bambini, il robot deve prima identificare e controllare il proprio corpo, tuttavia il controllo dei propri motori per un robot rappresenta una sfida da superare, soprattutto quando questa comprende l'interazione fisica tra il robot e l'ambiente.

La robotica ha adottato l'idea del reclutamento attivo del sistema motoristico per la costruzione di abilità percettive persino prima della scoperta dei neuroni specchio. Per esempio, il paradigma dell'*Active Vision* (Blake & Yuille, 1992) in *computer vision* ha proposto che i movimenti dei sensori possono aiutare il sistema percettivo a estrarre le informazioni direttamente in relazione all'obiettivo dell'osservatore.

Più specificatamente, molti autori hanno esplicitato il modello dei neuroni specchio o approssimato l'idea generale di moduli di processo comuni condivisi dalle azioni di produzione e comprensione. È possibile mostrare che le performance di un classificatore visivo sono migliorate mappando le informazioni visuali in una rappresentazione motoria come una fase di pre-elaborazione. In particolare, la complessità del classificatore è più bassa e la generalizzazione a nuove viste è migliorata. Questi risultati da un lato supportano il ruolo *dell'embodiment*, dall'altro evidenziano i benefici che ha un robot nello sviluppo di alcune *skills* presenti negli esseri umani.

Il problema della Working Memory e le architetture

Nel corso della storia sono stati molti i tentativi di costruire sistemi artificiali in grado di simulare le abilità cognitive. Questo ha fomentato il dibattito tra due differenti approcci teorici: Il connessionismo e il computazionalismo. La teoria computazionale della mente sostiene che il cervello è un sistema che processa informazioni e il pensiero può essere descritto come un processo computazionale che opera sugli stati mentali. Questa prospettiva ha guidato l'implementazione di una categoria di architetture cognitive chiamate simboliche. In queste architetture l'informazione è rappresentata da simboli ad alto livello, la cognizione diventa un processo di computazione che lavora sulle strutture simboliche e produce simboli in *output*. Il principale problema di questo approccio è che tutte le informazioni devono essere rappresentate come simboli appartenenti ad un dominio predefinito, il che rende difficile il suo utilizzo in grandi *datasets*.

L'idea centrale nell'approccio connessionista è che i processi mentali possono essere modellati come processi emergenti in reti ad unità funzionali altamente connesse. L'informazione in questo approccio è rappresentata dai segnali di attivazione che passano attraverso la rete.

Il modello connessionista più utilizzato è quello delle reti neurali artificiali (*ANN*), la quale sono state utilizzate per una vastissima varietà di aspetti della cognizione umana come memoria, percezione, attenzione, riconoscimento visivo e linguaggio. In molti casi architetture connessioniste sono state molto utili per spiegare alcune caratteristiche del comportamento umano che sono state descritte da osservazioni psicologiche. Tuttavia, fino ad oggi non sono state mai implementate simulazioni su larga scala per attività che richiedono ragionamenti complessi. Recentemente Eliasmith et. al (2012) hanno proposto un modello del cervello contenente 2,5 milioni e mezzo di neuroni in grado di elaborare sequenze visive e rispondere attraverso i movimenti del modello di un braccio fisico. Sono stati proposti anche altri modelli di simulazioni del cervello ma tutti questi modelli si sono concentrati sul realismo biologico del sistema senza affrontare apertamente il problema dell'elaborazione del linguaggio.

L'approccio simbolico è stato dominante per l'elaborazione del linguaggio naturale per molte decadi, poiché il linguaggio naturale è di per sé stesso un'attività fortemente simbolica, possiamo infatti considerare le parole come simboli che rappresentano oggetti, concetti, eventi ed azioni. La teoria del linguaggio formale usava l'algebra per definire i linguaggi formali con una sequenza di simboli.

Oggi il campo dell'elaborazione del linguaggio naturale è dominato dagli approcci di *machine learning*, che includono gli approcci basati sulle reti neurali, macchine a vettori di supporto, approcci Bayesiani e molti altri. I modelli delle reti neurali basati sul linguaggio hanno dimostrato migliori performance nel predire le parole successive rispetto agli approcci convenzionali, come i modelli *n-grams*. Recentemente tecniche di *deep learning* sono state usate con successo per molti problemi legati all'elaborazione del linguaggio come il riconoscimento delle parole o l'analisi sintattica. Alcuni di questi modelli sono biologicamente ispirati e sono stati disegnati come soluzioni ingegneristiche a specifici problemi *NLP*.

L'approccio connessionista ha dimostrato di essere eleggibile per modellare i fondamenti cognitivi del processamento del linguaggio. I suoi modelli sono stati utili per spiegare l'emergere delle abilità del linguaggio attraverso delle semplici regole di apprendimento, anziché richiedere una dettagliata conoscenza innata. Questo approccio enfatizza il ruolo dell'apprendimento attraverso l'interazione con l'ambiente, (fondamento di questa ricerca). Sebbene i modelli connessionisti sono stati ampiamente usati, non sono molti i modelli presenti che integrano i modelli neurali verbali dentro modelli comprensivi compatibili con la conoscenza attuale di come le informazioni sono memorizzate e processate nel cervello, come i modelli del *working memory*. Come descritto nei precedenti capitoli, Baddeley e Hitch proposero un modello di *working memory* composta da 4 componenti principali. Una centrale esecutiva, due sistemi *slave* (*loop fonologico* e il blocco di informazioni spaziali e visive) e un'unità *extra* per gli ingressi multimodali, il *buffer episodico*.

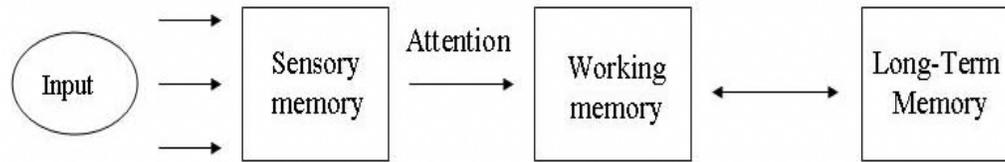


Figura 1 - working memory

Il modello di *working memory* proposto da Baddeley e Hitch nello specifico è così composto.

Logie e Baddeley (1989) hanno riconosciuto la centrale esecutiva come un meccanismo di controllo che alloca le risorse ai specifici sottosistemi ed è responsabile per il trasferimento delle informazioni nella memoria a lungo termine. Il *loop fonologico* e il blocco di informazioni spaziali e visive sono responsabili rispettivamente per l'analisi e la memorizzazione delle informazioni verbali e visive.

Il *buffer episodico* ha una limitata capacità di memorizzazione e tiene contemporaneamente attivate le informazioni da entrambi i sottosistemi, nonché

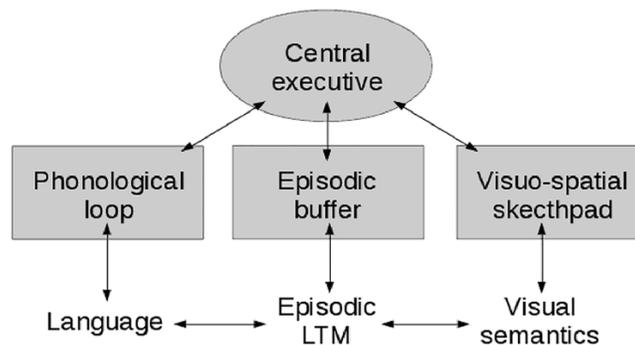


Figura 2

l'informazione dalla memoria a lungo termine. Baddeley ha notato che queste due features del *buffer episodico*, hanno la capacità di creare la rappresentazione episodica e permettere una breve memorizzazione delle informazioni che eccedono la capacità dei due sottosistemi.

Nella sua formulazione Cowan (1999) sostenne che il sistema umano di memoria opera grazie ad una ricca interazione tra i meccanismi di attenzione e quelli di memorizzazione. L'informazione nel sistema di memoria può essere tenuta sia in uno stato d'attivazione sia in uno stato di non attivazione, e in quest'ultimo stato gli elementi dell'informazione rappresentano l'*LTM*.

Le risorse di attenzione sono usate per stimolare l'informazione dalla memoria a lungo termine con l'obiettivo di raggiungere lo scopo richiesto. In questo modello le unità di attivazione possono essere reperite da sensori multi-modali. Golosio et al (2015) hanno proposto un modello di architettura cognitiva che utilizza una rete neurale come centrale esecutiva che prende in input gli stati di attivazione neurale derivanti dalla memoria a breve termine e fornisce in *output* le azioni mentali, la quale controlla il flusso di informazioni tra le componenti di *working memory* attraverso meccanismi neurali di *gating*. L'obiettivo del progetto di Golosio è quello di proporre un sistema in grado di comunicare attraverso il linguaggio naturale partendo da una situazione di tabula rasa, ovvero senza alcuna conoscenza a priori degli elementi del linguaggio, come struttura della frase, significato delle parole etc. etc., solo interagendo con un umano attraverso un'interfaccia basata sul testo. Nello specifico è ipotizzato che un tale sistema deve avere almeno i seguenti componenti: uno *store* fonologico, un *focus* d'attenzione, una struttura di recupero che utilizza il *focus* dell'attenzione per recuperare le informazioni dall'*LTM*, un *goal stack* e un sistema di supervisione che controlla il flusso delle informazioni tra i vari componenti (ad es. una centrale esecutiva). Il modello su descritto tuttavia si concentra sul ruolo delle funzioni esecutive nel processamento del linguaggio. La domanda che ci si potrebbe porre è: perché utilizzare una rete neurale per modellare questo processo e non utilizzare un modello simbolico. Innanzitutto possiamo partire dalla considerazione che il cervello è di per sé un'architettura neurale, oltre ciò l'architettura proposta non è un'architettura basata su regole ma prende decisioni statistiche, e se la centrale esecutiva non fosse basata su scelte statistiche non sarebbe in grado di generalizzare.

Il modello di architettura in corso di sviluppo in questa tesi si basa invece su un sistema modulare in grado di sfruttare la memoria sintetica al fine di sviluppare avanzate abilità

affettive per i *social robot* che vanno oltre il corrente stato dell'arte, con l'obiettivo di creare approcci innovativi per la modellazione computazionale e la simulazione di processi affettivi simili a quelli umani e che siano in grado di basare la loro influenza sul *decision making* così come sui comportamenti del robot.

Tradizionalmente c'è sempre stato un forte dislivello tra lo studio scientifico delle emozioni (*affective science*) e i sistemi d'intelligenza artificiale in grado di riprodurre ed esprimere emozioni (*affective computing*). I ricercatori hanno tentato di concettualizzare i modelli di scienza affettiva per lo sviluppo dei modelli computazionali per l'emozioni ma hanno lasciato ancora molte domande aperte.

In questa tesi mi sono occupato di modellare, implementare ed integrare, sfruttando un sistema di memoria sintetica, il ciclo dei processi emotivi che copre l'elicitazione dei processi affettivi, l'interazione tra i differenti fenomeni affettivi, l'influenza che i differenti processi cognitivi hanno sull'affetto, le conseguenze dell'affetto sulle diverse componenti cognitive (attenzione, memoria, etc.) e sui sistemi motivazionali e di pianificazione dei robot e l'espressione multimodale dell'affetto.

L'obiettivo è implementare il sistema in un *framework* modulare che integra i vari meccanismi dei sistemi affettivi umani così da permettere allo sviluppatore di utilizzare i modelli affettivi legati a teorie consolidate (ad es. la teoria di valutazione di Roseman's, i modelli di processamento dei componenti di Scherer's, gli OCC model, i modelli di Gross's per la regolazione delle emozioni) specifici per ogni meccanismo affettivo. Per esempio, le valutazioni delle componenti del modello delle emozioni possono essere usate per informare il modulo d'elicitazione delle emozioni (Scherer, 2009), mentre il modello di Gross può fornire le basi per la regolazione delle emozioni (Gross, 2013), e quello di Ekman e Matsumoto's per l'espressioni facciali.

III – Implementazione di un’architettura per l’integrazione uomo – macchina tramite social robots

L’Architettura oggetto della tesi

Gli algoritmi, i modelli e le architetture implementate nel lavoro di tesi consentiranno di sviluppare abilità affettive per i robot con l’obiettivo di permettere loro di diventare più sociali e *human-friendly* affrontando e fornendo le risposte alle carenze sopra identificate.

Il mio approccio si focalizza sui seguenti temi:

- Implementare un *framework* innovativo per simulare le abilità affettive nei robot, sviluppando le tecnologie di *middleware* in grado di collegare differenti moduli che emulano meccanismi affettivi: elicitazione delle emozioni, dinamiche affettive, memoria affettiva, regolazione affettive, effetti esterni e interpretazione del contesto, oltre alla generazione di comportamenti affettivi stessi. Un approccio modulare consente l’implementazione computazionale di differenti teorie scientifiche riguardanti le modalità affettive, per ogni componente, consentendo quindi all’utente di selezionare un differente approccio in ogni componente del *framework*. Saranno create chiare e distinte definizioni delle operazioni, unificando le proprietà di interscambio fra le varie teorie e permettendo la standardizzazione dell’interfaccia software per ogni concetto affettivo computazionale.
- Sviluppare un modello computazionale di affetto basato su memoria sintetica per l’interazione a lungo termine. Le nuove abilità affettive saranno sviluppate esplorando le sue tecnologie con strutture credibili. Basate su apprendimento da

dati reali, possibilmente estendendo e migliorando le architetture usate nel settore.

- Progettare la soluzione tecnologica da sviluppare attraverso la tesi con la realizzazione di un prototipo funzionante, la cui sperimentazione finale avverrà in collaborazione con il CNR Isasi di Messina, che ha come focus la ricerca applicata ai bambini affetti da disturbi dello spettro autistico. Gli algoritmi in sviluppo mirano a far diventare i robot più sociali e amichevoli, fornendo loro una propria memoria e delle abilità affettive che consentiranno loro di sostenere interazioni a lungo termine. Seguendo i paradigmi presenti nello stato dell'arte, la mia tesi è, prima di tutto, sviluppata sulle conoscenze esistenti in modo da includere le versioni avanzate di caratteristiche già studiate e sviluppate nel passato, principalmente nei domini dei robots sociali, elicitazione emozionale, dinamiche affettive, regolazione affettiva, memoria affettiva, generazione di comportamenti affettivi e dell'interpretazione del contesto.

Il progetto, avendo come punto di partenza le caratteristiche come quelle sopra indicate e come base l'analisi della letteratura riguardante le scienze affettive e *l'affective computing*, ha l'obiettivo di avanzare lo stato dell'arte nelle capacità affettive per i *social robots*, sviluppando un *framework* modulare basato sull'architettura concettuale sotto illustrata.

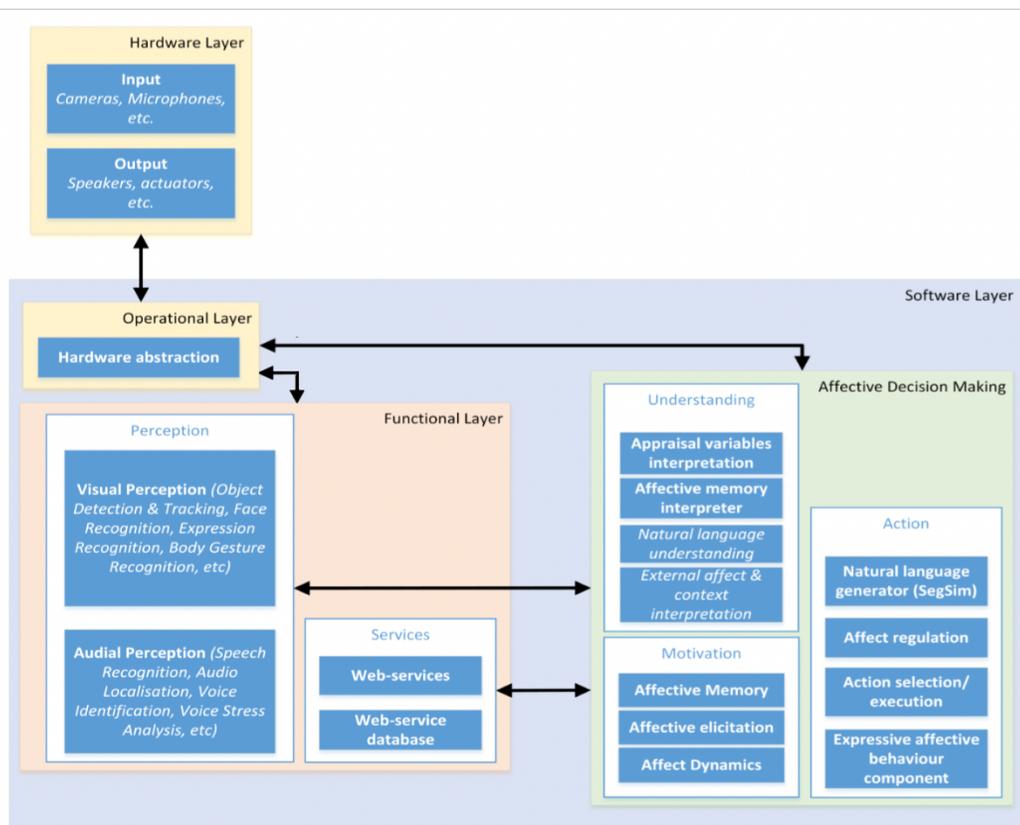


Figura 3 - L'architettura del sistema cognitivo di interazione a lungo termine

I principali componenti dell'architettura possono essere descritti come segue:

La **componente della percezione** sarà responsabile per il controllo del sistema robotico di percezione del contesto e consisterà di due sotto-componenti:

La componente dell'ambiente e dei comportamenti umani, la quale sarà responsabile per il rilevamento del contesto operativo del robot, nei termini sia del luogo in cui opera sia della sua controparte umana. Una struttura gerarchica conetterà la mappa metrica dell'ambiente (*3D*) con le informazioni semantiche, codificando così gli spazi in casa, gli oggetti e le loro relazioni, nonché le relazioni che gli oggetti possono avere con le attività tipiche dell'utente robot e renderà il robot in grado di conoscere la sua posizione nello spazio *3D*. Inoltre, un *framework* gerarchico probabilistico sarà ideato per codificare come le azioni dell'utente possono formulare le attività, che sono di interesse nel campo

d'applicazione dei casi d'uso dell'autismo, in relazione agli oggetti in casa, e successivamente come le attività che fanno parte della routine dell'utente sintetizzano i comportamenti giornalieri dell'utente. Lo sviluppo si basa su metodi esistenti in *computer vision* per il rilevamento umano ed il riconoscimento delle attività, che saranno estesi ove necessario al fine di migliorarne la robustezza. Questo componente sarà in grado di riconoscere una serie di attività e di comportamenti umani.

La componente che influenza la percezione esterna, la quale sarà in grado di abilitare il robot a percepire il corrente stato affettivo dell'utente. Alla fine di questo, i metodi esistenti per il riconoscimento affettivo basati sulla *computer vision* saranno investigati principalmente utilizzando il riconoscimento delle espressioni facciali, così come sulle affezioni vocali e sulle espressioni corporee (cambi nell'intonazione e nel modo in cui l'utente esegue le attività), relativamente alle emozioni di interesse primario nell'utente. Ad un altro livello saranno investigati anche i metodi per il riconoscimento delle emozioni basati sui biosegnali, basandosi sulla potenziale disponibilità delle misure dei biosegnali dell'utente, che possono essere forniti dal robot (per es. un orologio *wireless* che monitora i segnali del *GSR* e *dell'HR*), concentrandosi specificatamente su un riconoscimento emotivo più robusto degli stati affettivi con una forte componente della dimensione d'eccitazione. Questo componente, attraverso la fusione dei metodi dei suddetti campi, permetterà la percezione dello stato affettivo dell'utente la quale sarà successivamente alimentata, durante le operazioni del robot, dal *framework* cognitivo.

La **componente dell'interpretazione delle variabili di valutazione** è una parte del meccanismo affettivo di supporto alle decisioni e riceverà le caratteristiche degli stimoli o li percepirà esternamente dai componenti della percezione o li invocherà internamente dai componenti cognitivi, e calolerà i valori delle variabili di valutazione (novità, valenza, rilevanza per l'obiettivo, consistenza dell'obiettivo, certezza, controllo, rilevanza del valore) analizzandone le caratteristiche e basandosi anche sulle informazioni cognitive ulteriori (ad es. memorie, obiettivi, desideri, sforzi, credenze, contesto). Gli algoritmi innovativi basati sulle teorie consolidate saranno responsabili di calcolare automaticamente il valore di ogni variabile di valutazione e saranno sviluppati per il

corretto funzionamento di questa componente. Inoltre, in futuro il calcolo delle variabili di valutazione potrà anche essere basato su algoritmi di apprendimento in modo da considerare gli aspetti etici relativi alle abilità affettive dei robots. L'*output* della componente di interpretazione delle variabili di valutazione sarà successivamente usata per stimolare nel robot comportamenti emotivi.

La **componente d'elicitazione delle emozioni** rappresenta anche una parte del meccanismo affettivo di supporto alle decisioni e valuterà le variabili di valutazione basate sulla teoria dell'affetto selettivo, oltre a iniziare il processo emozionale. Questo componente, più specificatamente, riceverà le variabili d'approvazione dalla componente d'interpretazione di queste variabili ed eliciterà il processo emozionale corrispondente, basato su una varietà di teorie di valutazione (Scherer, Frijda, Roseman, etc.). Il componente di elicitazione dell'affetto interagirà anche con gli altri componenti affettivi per determinarne le proprietà dei processi emozionali (es. forza, durata, messa a fuoco, etc.). Tradizionalmente, le componenti *software* per l'elicitazione delle emozioni sono costruite per ritornare emozioni ben distinte (ad es. l'emozione della gioia con intensità al 70%). Tuttavia, un oggetto emotivo identificato dalla componente d'elicitazione dell'affetto può anche includere sia altre informazioni rilevanti, sia variabili d'approvazione necessarie per altre componenti.

L'**Influenza delle dinamiche** sarà responsabile della simulazione delle dinamiche tra differenti concetti affettivi. Le precedenti esperienze fatte dal CNR Isasi nell'implementazione di simili componenti nei loro rispettivi *framework*, contribuiranno all'ideazione di algoritmi robusti per la simulazione delle dinamiche che saranno alla base per lo sviluppo della componente di influenza delle dinamiche.

La **componente di Memoria Affettiva** memorizzerà e richiamerà le varie informazioni che vengono dai vari componenti e inoltre costruirà anche le associazioni tra i concetti. Questi processi di memoria sono fortemente influenzati dall'effetto risultante della memorizzazione o del richiamo soggettivo. La componente di memoria affettiva simulerà elementi della memoria semantica umana, la quale memorizzerà le informazioni fattuali,

mentre la memoria episodica umana memorizzerà le esperienze del robot. L'idea è che devono essere elaborati algoritmi che copriranno l'assunzione che l'affetto può influenzare la chiarezza e la durabilità della memoria; i ricordi simili allo stato d'animo dell'agente hanno più probabilità di essere richiamati (memoria umore- congruente) e richiamare gli eventi emozionali può influenzare lo stato affettivo attuale del robot.

La **componente di Valutazione delle Conseguenze** sarà responsabile del valutare e determinare le conseguenze delle azioni. Tutte le potenziali azioni e le strategie di regolazione (che possono servire come risposta sugli stimoli percepiti) saranno valutate e priorizzate, basandosi sulle conseguenze aspettate); questa informazione sarà fornita alla componente motivazionale. Questa componente dovrà essere vista come il livello in cima della memoria episodica, dove l'informazione statistica sulle associazioni è derivata dalle esperienze precedenti e la possibile conseguenza e i loro valori emozionali sono automaticamente determinati. Per questo componente occorrono algoritmi per trarre queste associazioni dalle informazioni memorizzate in memoria e mandarne i *feedback* al sistema motivazionale del robot.

La **componente di regolazione dell'Affetto** sarà responsabile di regolare lo stato affettivo corrente o di anticipare lo stato affettivo futuro abilitando un'azione esterna per affrontare la causa dell'emozione o modulando l'affetto interno e le componenti cognitive. Questa componente si baserà su teorie consolidate e valuterà l'effetto del robot alle situazioni emotive in corso, gli obiettivi del robot e le altre informazioni contestuali, e darà avvio alla migliore strategia affettiva in grado di affrontare la nuova situazione.

La **componente di Selezione dell'azione** può essere vista come parte del sistema motivazionale del robot o come parte del meccanismo di elicitazione delle emozioni. Questo componente sarà responsabile per provocare la tendenza a una determinata azione basandosi sugli eventi esterni, interni o associati con il processo di elicitazione delle emozioni.

La **componente espressiva dell'affetto** sarà responsabile della generazione di comportamenti affettivi legati alle espressioni facciali e corporee del robot, basandosi sulle informazioni derivanti dal componente di selezione dell'azione e dal componente di influenza delle dinamiche. Metodi computazionali saranno analizzati per consentire al robot di prendere decisioni specifiche e compiere azioni guidate dall'affetto o relative al cambio del posizionamento del robot in base all'utente, o relative al cambio della gestualità o espressione. L'effetto sulla voce può essere considerato come un canale addizionale per i comportamenti che esprimono affetto. Per questo canale è auspicabile impiegare software di *text-to-speech* disponibili in commercio, che supportino anche un elemento affettivo cruciale come il cambio dell'intonazione vocale.

Sviluppo dei singoli componenti

Definizione dei requisiti

Sono in via di definizione i requisiti delle seguenti relazioni: (1) come i concetti e le funzioni psicologiche (motivazione, valutazione, etc. etc.) interagiscono l'uno con l'altro, (2) come gli stimoli interni (stati fisiologici e cognitivi) e esterni (la percezione nell'ambiente) sono processati (negli episodi emotivi, questo è comunemente definito processo di valutazione), (3) come il risultato di questo processo di valutazione influenza o inizia le azioni e i comportamenti, (4) come producono o influenzano le espressioni facciali, corporee e vocali, e (5) cosa sono i processi di regolazione che sono attivati sui processi di valutazione, sui comportamenti e sulle espressioni.

Sono state progettate le interfacce e definite le funzionalità dei componenti del *software* per *l'affective computing*, così come le strutture dati e le ontologie che saranno scambiate attraverso i vari componenti del *software* e le componenti del *software* cognitivo / motivazionale dell'architettura robotica. Le specifiche devono coprire la maggioranza

delle teorie principali che riguardano la scienza affettiva di ogni componente, rendendo così possibile il cambio tra le varie teorie nel *framework* modulare / computazionale. Per esempio, la definizione delle variabili di approvazione / elicitazione delle emozioni coprirà la maggior parte delle teorie principali di questo argomento (Scherer, Fridja, Roseman, etc.), rendendo così possibile per lo sviluppatore o l'utente selezionare la teoria desiderata per l'elicitazione dell'emozione senza preoccuparsi circa gli ingressi e le uscite di questo componente del *software*.

Simulazione dell'effetto dei componenti

Componente d'interpretazione delle Variabili di Valutazione

Questo compito si focalizza sullo sviluppo del componente d'interpretazione delle variabili di approvazione, la quale riceve le caratteristiche degli stimoli, o la percezione esterna dai componenti di percezione, o la invoca internamente grazie alle componenti cognitive, e calcola il valore delle variabili di valutazione (es. novità, valenza, rilevanza per l'obiettivo, consistenza dell'obiettivo, certezza, controllo, rilevanza del valore) analizzandone le caratteristiche, e basandosi anche sulle informazioni cognitive ulteriori (ad es. memorie, obiettivi, desideri, sforzi, credenze, contesto). Lo scopo principale è considerare le raccomandazioni e le specifiche concettuali, e progettare e sviluppare innovativi algoritmi, basati su teorie consolidate, responsabili di calcolare automaticamente il valore di ogni variabile di valutazione. Il calcolo di alcune di queste variabili potrà richiedere nel futuro anche lo sviluppo di algoritmi di apprendimento in grado di considerare gli aspetti etici. *L'output* di questo componente sarà successivamente usato per attivare i comportamenti emotivi nel robot.

Componente di Elicitazione dell'Emozione

Questo compito si focalizza sullo sviluppo del componente d'elicitazione, la quale riceve dal componente dell'interpretazione delle variabili di valutazione tali variabili e elicita il corrispondente processo emotivo. Questo componente implementerà una varietà di teorie d'approvazione (Scherer, Frida, Roseman, etc) e renderà più facili per lo sviluppatore o l'utente passare da una teoria ad un'altra. Considerando le raccomandazioni e le specifiche

concettuali, questa funzione progetterà e svilupperà innovativi algoritmi, basati su teorie consolidate, responsabili di analizzare le variabili di valutazione e automaticamente elicitarne i processi emozionali. La componente di elicitazione dell'affetto interagirà anche con gli altri componenti affettivi per determinarne le proprietà dei processi emozionali (es. forza, durata, messa a fuoco, etc.). La componente di valutazione è responsabile della codifica affettiva degli stimoli interni ed esterni, che sono necessari per l'espressione dei comportamenti emotivi, così come per il supporto emotivo alle decisioni. Tuttavia, un oggetto emotivo identificato dalla componente d'elicitazione dell'affetto può anche includere sia altre informazioni rilevanti, sia variabili d'approvazione necessarie per altre componenti.

I componenti Affettivi di Memoria

Questo compito si focalizzerà sullo sviluppo dei componenti affettivi di memoria, la quale saranno responsabili del richiamo e della memorizzazione delle varie informazioni (utilizzate nel tempo) usate dai componenti di supporto alle decisioni. Il compito per questa componente sarà quello di progettare e sviluppare innovativi algoritmi che copriranno l'assunzione che l'affetto può influenzare la chiarezza e la durabilità della memoria; i ricordi simili allo stato d'animo dell'agente hanno più probabilità di essere richiamati (memoria umore- congruente) e richiamare gli eventi emozionali può influenzare lo stato affettivo attuale del robot. L'*output* di questo componente di memoria sarà successivamente usato per la maggioranza dei componenti di questa architettura. Le precedenti esperienze fatte dal CNR, nell'implementare componenti simili nei loro *framework*, contribuiranno allo sviluppo di algoritmi robusti per la gestione della memoria dichiarativa.

La componente d'interpretazione della memoria affettiva

Questo compito si focalizzerà sullo sviluppo di un componente responsabile per la valutazione e la determinazione delle conseguenze delle azioni. Questo componente è responsabile della valutazione e della prioritizzazione di tutte le potenziali azioni, oltre alle strategie di regolazione (che possono servire come risposta agli stimoli percepiti),

basandosi sulle conseguenze che ci si attendono. Questa informazione sarà fornita ai componenti motivazionali. Essa dovrà essere vista come il livello in cima della memoria episodica, dove l'informazione statistica sulle associazioni è derivata dalle esperienze precedenti e la possibile conseguenza e i loro valori emozionali sono automaticamente determinati. Occorre trarre queste associazioni dalle informazioni memorizzate in memoria e mandarne i *feedback* al sistema motivazionale del robot. Le precedenti esperienze fatte dal CNR nel derivare automaticamente informazioni statistiche dai comportamenti e dal monitoraggio dei dati dell'ambiente contribuiranno allo sviluppo di algoritmi robusti per la valutazione delle conseguenze.

Componente per la regolazione dell'affetto e della cognizione.

Questo compito si focalizza sullo sviluppo del componente responsabile per la regolazione dell'affetto. Questa componente valuterà l'effetto del robot rispetto alle situazioni emotive. Successivamente, la componente d'interpretazione della memoria affettiva prende il controllo e valuta le conseguenze potenziali delle strategie. Questo può portare nuovamente indietro alle componenti di regolazione se la precedente strategia di regolazione è stata valutata negativamente. Questo compito ha lo scopo di progettare e sviluppare algoritmi per derivare queste associazioni dalle informazioni memorizzate e per mandare *feedback* al componente d'interpretazione della memoria affettiva. Le precedenti esperienze fatte dal CNR nella regolazione delle emozioni nei loro rispettivi *frameworks* contribuiranno allo sviluppo di algoritmi robusti per la valutazione delle conseguenze.

Sistema motivazionale del Robot

Il sistema motivazionale integra le abilità affettive nel processo di selezioni delle sue azioni e viene guidato dagli eventi di questo componente, inoltre influenza gli altri componenti quando necessario. Le precedenti esperienze fatte dal CNR nei sistemi di supporto alle decisioni e nei sistemi motivazionali per i robots nei loro rispettivi *frameworks* ha contribuito allo sviluppo del sistema motivazionale.

Architettura

Sulla base degli studi fino a qui ho quindi disegnato la seguente architettura, che consentirà l'interazione uomo-macchina a breve ed a lungo termine.

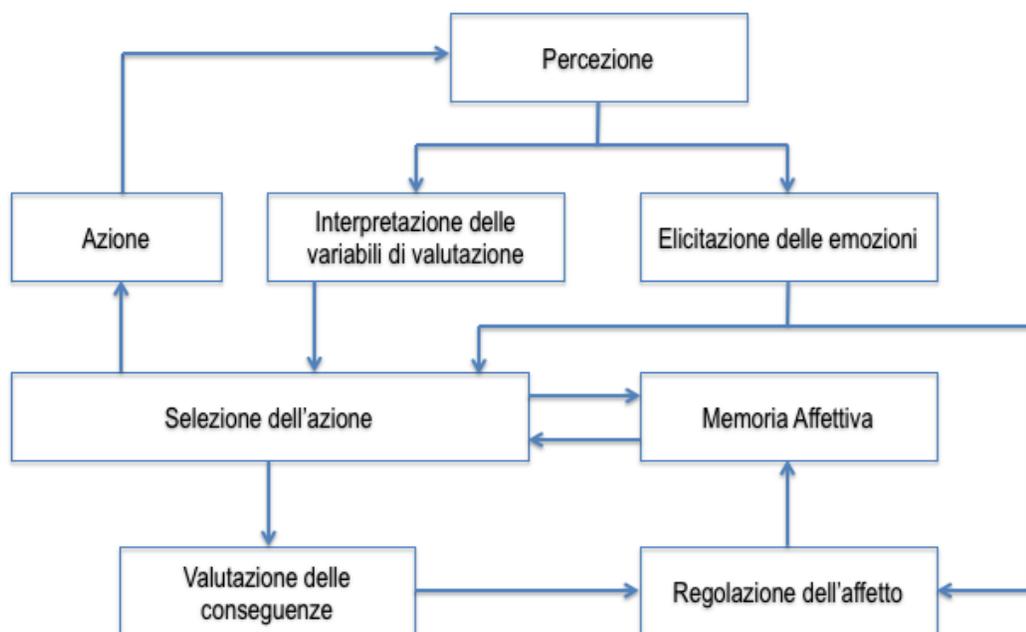


Figura 4 - Framework

Ho scelto di utilizzare la componente della percezione per il controllo del sistema robotico di acquisizione di dati sul contesto. Questa componente consente al robot di interpretare il contesto e di valutare il livello affettivo esterno.

Per poter applicare tale componente nella pratica, si devono stabilire due requisiti di base relativamente alla percezione che ha il robot del suo contesto operativo, ponendo l'enfasi sul monitoraggio del comportamento dell'utente e della sua incidenza. A tal fine, secondo questo approccio che ho deciso di seguire, la possibilità di valutare un "profilo utente" completo e quindi la sua modellazione si basa su due assi principali: la classe comportamentale e quella affettiva. La prima componente si focalizza sulle attività di analisi, ad es. la modalità con cui un individuo esegue una serie di azioni in relazione alla sua interazione delle variabili ambientali, come gli oggetti o le altre persone. La seconda componente si concentra sul riconoscimento affettivo, un fattore che indica la corrente situazione emotiva dell'utente.

Le attività di analisi comprendono quindi una componente di base per il "profilo utente" che in accordo con la letteratura attuale può essere modellata in due modi: i) attraverso l'analisi della modalità con cui le persone eseguono le attività e ii) attraverso la modellazione dell'interazione della persona con l'ambiente. La maggior parte delle metodologie di riconoscimento dell'attività si basano sull'abilità di dedurre l'attività in corso. I sistemi moderni sfruttano quasi esclusivamente la codifica dei colori *RGB* (Poppe, 2010) e la loro profondità (Chen et al., 2013) al fine di realizzare le attività e modellare i comportamenti, mentre il *focus* sull'interazione coinvolge due o più persone e/o oggetti. I metodi che raggiungono maggiori risultati nello stato dell'arte si basano sull'estrazione di *features* semantiche (Ziaeeafard et al., 2015) così da poter riconoscere attività di alto livello. In aggiunta, Gupta e Davis (Gupta e Davis, 2007) usano la coerenza tra tipi di oggetti e la postura del corpo dell'utente per migliorare l'accuratezza dell'analisi. Il metodo si concentra sulla postura della parte superiore del corpo dell'utente e più precisamente sulla traiettoria delle mani. Data la postura rilevata, il metodo estrae le caratteristiche di movimento relative agli arti, come la velocità e l'accelerazione. Dopo questa prima fase una *belief network* associa queste caratteristiche agli oggetti nella scena. Gupta et al. (2009) suggerisco anche l'impiego di un *framework Bayesiano* per integrare le posture umane e i cambi dinamici in relazione agli oggetti. Molti metodi sfruttano la correlazione tra le posizioni e le loro connessioni per estrarre *features*. In particolare, Han

et al. (2009) mettono in evidenza come un sistema, al fine di rendere possibile il riconoscimento, debba necessariamente usare non solo correlazioni spazio-temporali a basso livello estratte dal video, ma anche descrittori relazionali fra persone ed oggetti nella scena. Ho deciso di seguire questo approccio nella mia tesi. Impiegano quindi una serie di descrittori contestuali della scena ed un metodo di apprendimento di tipo Bayesiano basato su un processo Gaussiano per riconoscere le azioni dell'interlocutore in video complessi non strutturati. Nello specifico, i descrittori contestuali della scena catturano i *pattern* comportamentali mentre i processi di tipo Bayesiano classificano le interazioni tra le persone o tra le persone e gli oggetti. Ho segnalato questo lavoro perché lo ritengo chiave nello sviluppo della componente della percezione, anche se è importantissimo tenere conto del fatto che il metodo, come altri simili, fallisce quando il tracciamento degli oggetti non va a buon fine.

Per quanto riguarda l'interazione tra gli individui, Meng et al. (2012) tracciano le locazioni dei punti relativi alle parti del corpo e determinano i *joints* per tutti gli individui. Il metodo analizza le relazioni fra i *joints* estratti sia su un livello intra-personale che su un livello inter-personale e ne estrae le *features* semantiche spaziali. Grazie a quest'ultime, un modello di classificatore è addestrato per discriminare le attività correnti. Mukherjee et al. (2014) suggeriscono un approccio grafo-teorico per catturare le relazioni tra persone, dove i nodi del grafo sono le loro posizioni dominanti. Più precisamente, per ogni partecipante all'attività, un descrittore di posizione è calcolato in base a un dizionario di posizioni; le posizioni dominanti sono quindi estratte secondo una metrica per tutti gli individui che contribuisce alla formazione di un grafo bipartito. Tale grafo può fornire una descrizione dell'interazione, che modella l'attività eseguita.

Chen et al. (2011) propongono un *framework* per la comprensione basato sulla stima del corpo e dei movimenti della testa. A seguito del tracciamento della postura del corpo, il loro algoritmo ha due fasi. Nella prima viene fatta un'analisi per ogni frame tesa a localizzare la testa ed estrarre *features* dalla testa stessa e del corpo. Nella seconda si valuta la probabilità che queste *features* appartengano ad una particolare posa. Le

probabilità sono quindi fuse all'interno di un filtro temporale che stima nel complesso la posizione del corpo e la posa del corpo e della testa sfruttando l'accoppiamento fra la posizione del corpo (in termini di direzione del movimento), la posa del corpo e la posa della testa. In pratica, la descrizione del comportamento è dedotta da un modello dinamico di primo ordine che accoppia i *pattern* di comportamento con la direzione del corpo e della testa. Tuttavia, anche se interessante, tale sistema non può essere utilizzato in questa tesi, poiché è privo di informazioni che sono essenziali per la percezione; non si ottiene una raffigurazione completa dell'utente e dell'ambiente.

Per quanto riguarda la componente affettiva, oggetto di studio di varie discipline, in letteratura troviamo molte ricerche, in particolare i metodi di analisi automatizzata, che tentano di trovare soluzioni in grado di gestire spontaneamente i pattern acquisiti. Numerosi studi suggeriscono come sia il canale visuale, in particolare l'espressione facciale, l'indizio cruciale, soprattutto se correla adeguatamente con il corpo e la voce (Ambady e Rosenthal, 1992). Le caratteristiche facciali più comunemente estratte sono quelle geometriche o quelle basate sulle apparenze, come rughe, rigonfiamenti e solchi. Le metodologie ben consolidate che si basano su metodi di estrazione di *feature* geometriche sono quelli di Chang et al. (2006) che utilizza un modello facciale costituito da 58 *landmarks* e quelli di Pantic et al. (2006) e Pantic e Rothkrantz (2004) che utilizzano una *set* di punti facciali ben distinti circostanti l'area di occhi, bocca, naso, mento, sopracciglia. Esempi tipici di *features* basate sull'aspetto sono quelle di Bartlett et al. (2005;2006) e di Guo e Dyer (2005) che utilizzano le *wavelets* di Gabor e le *eigenfaces* (dall'analisi degli autovettori estratti dall'immagine). Anderson e McOwen (2006) utilizzano un modello di volto che si adatta al volto estratto su base spaziale, mentre Chang et al. (2004) hanno formulato un algoritmo di riconoscimento probabilistico basato su alcuni sottospazi costituiti da facce allineate. Alcuni autori usano approcci ibridi che utilizzano *feature* geometriche e *template* di volti. Zhang e Ji (2005) hanno determinato e processato 26 punti facciali attorno agli occhi, le sopracciglia e la bocca e dei punti di transizione come le cosiddette zampe di gallina o i solchi naso-labbra. Lucey et al. (2005) utilizzano un modello, indicato nel lavoro come l'*Active Model Appearance*, in grado di

catturare le caratteristiche di aspetto facciale e la forma delle espressioni facciali. Sempre nell'ambito dell'approccio ibrido, molti studi si focalizzano sul riconoscimento emotivo multi-modale e più precisamente sull'analisi di effetti audio visivi. Zeng et al. (2007) usano i dati raccolti nelle interviste di ricerche psicologiche e si focalizzano sul riconoscimento di emozioni positive e negative.

Gesti e posture, in effetti, possono essere considerati quali parti di un più ampio sistema semiotico che è alla base della comunicazione umana. Si può, infatti, sostenere che l'attitudine, l'intenzione e, in generale, il significato, vengono espressi solo in parte dal contenuto verbale, e spesso, se non la maggior parte delle volte, da canali non verbali. In questa prospettiva, i canali non verbali possono essere interpretati come gli elementi non detti di rappresentazione degli stati interni (Ekman & Oster, 1979; Amighi, 1999; Taylor et al., 2011). Su questa linea, alcuni ricercatori lavorano nell'estrazione di parametri gestuali da sequenze video; riportano come sia un approccio molto interessante per il riconoscimento automatico delle emozioni (Glowinski et al., 2011).

L'espressione dell'emozione potrebbe essere inferita non esclusivamente da sequenze video/audio ma con un approccio particolare, l'acquisizione e l'analisi di correlati fisiologici tramite sensori indossabili. Questi approcci che si basano su biosegnali vengono tipicamente adottati nell'analisi dell'*arousal*. Mauri et al. (2010) riportano quantitativamente la correlazione fra stati affettivi e segnali acquisiti da sensori indossabili; i segnali riportati nel lavoro sono: trend della respirazione, la risposta galvanica della pelle, il volume pulsato del sangue, l'elettrocardiogramma e l'encefalogramma. Betella et al. (2015), hanno disegnato un esperimento dove, in contrasto ai tradizionali protocolli, i partecipanti erano liberi di muoversi in uno spazio fisico e di essere stimolati a differenti livelli di *arousal*, in modo da poter studiare la flessibilità del riconoscimento affettivo in ambienti realistici. Altri lavori del gruppo di ricerca CNR che segue il mio lavoro di tesi, riportano ottimi risultati in *setting* terapeutici strutturati reali (Billeci et al., 2016). Tuttavia, proprio come condiviso con il gruppo di ricerca CNR, queste tecnologie, quando utilizzate al fine del riconoscimento affettivo in ambienti non strutturati, hanno dei limiti che soltanto grandi gruppi industriali e di ricerca a livello internazionale e grazie ad enormi investimenti, stanno cercando di superare. Si

può concludere che l'approccio da seguire è senza dubbio multi-modale che comprende scale differenti, strutture temporali e livelli metrici; in ogni caso, si può affermare che il riconoscimento emotivo è ancora un quesito aperto.

Basandomi sull'attenta analisi sopra riportata, ho quindi individuato e sviluppato le seguenti due sotto-componenti:

a) *Componente di acquisizione di dati relativi all'ambiente ed al comportamento dell'interlocutore.* Per quanto riguarda l'interlocutore, questo modulo ne rileva la presenza, individua i segmenti corporei ed inferisce i principali comportamenti. Per quanto riguarda l'ambiente, l'algoritmo si basa sui sensori integrati nel robot per rilevare il contesto nel quale opera. Grazie alle telecamere ed all'algoritmo da me sviluppato è possibile estrarre dati dall'ambiente con un sistema di riferimento egocentrico. In termini di analisi dell'ambiente circostante, l'algoritmo da me sviluppato integra librerie open source di *Image captioning in TensorFlow* (sviluppate da Google Brain team); si basano su un sistema di machine learning in grado di produrre automaticamente frasi che descrivono i contenuti di un'immagine (Vinyals et al., 2016). Il modello è basato su una *deep convolution neural network (CNN)* che prendendo in ingresso l'intera immagine è in grado di rappresentarla tramite vettori a lunghezza fissa che rappresentano una "codifica" dell'immagine stessa. Il processo di codifica è seguito da un *task* di classificazione dell'immagine, la cui uscita confluisce in un "decodificatore" basato su una *Recurrent Neural Network (RNN)*. Con l'aiuto dei ricercatori del CNR, ho collezionato un database di video acquisiti nel corso di terapie di trattamento cognitivo-comportamentale svolte da terapisti esperte su bambini affetti da disturbi dello spettro autistico. Ho addestrato quindi il sistema sviluppato con questi video. Successivamente, dall'analisi delle frasi ottenute quale *output* del sistema, ho sviluppato un algoritmo che classifica possibili relazioni fra gli oggetti nella scena ed il robot, nonché una possibile relazione fra gli oggetti nella scena e l'interlocutore.

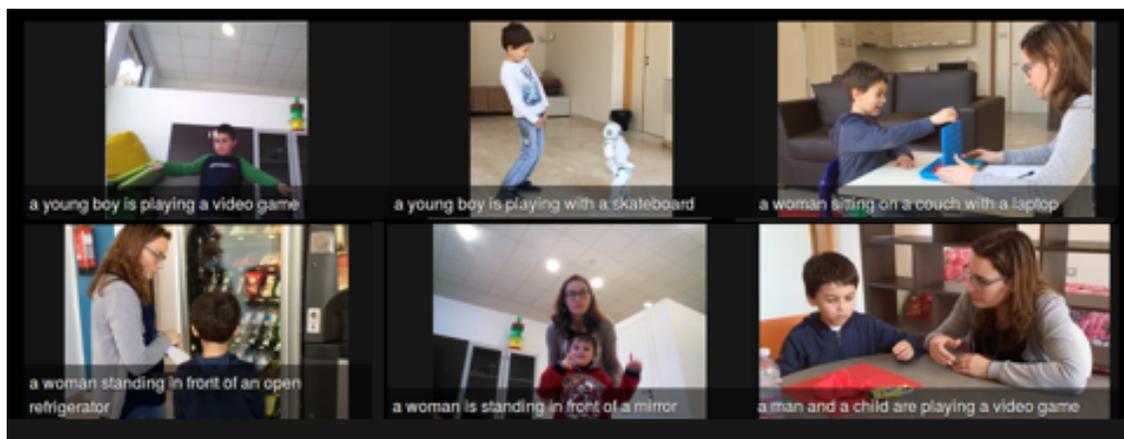


Figura 5 - Riconoscimento delle azioni nella scena

In termini di rilevazione ed analisi del comportamento dell'interlocutore, ho progettato un algoritmo che integra la libreria open source *OpenPose*, una libreria *multi-thread* per la rilevazione di *keypoint* in tempo reale su più persone (Cao et al., 2017). La libreria è scritta in C++ utilizzando *OpenCV* e *Caffe*, i cui autori sono Gines Hidalgo, Zhe Cao, Tomas Simon, Shih-En Wei, Hanbyul Joo e Yaser Sheikh. *OpenPose* è il primo sistema in grado di rilevare in tempo reale i giunti corporei, le mani e *keypoints* facciali (in totale 130 *keypoints*) su una singola immagine. Inoltre, cosa fondamentale per lo scenario applicativo di questa tesi, le performance computazionali del sistema consentono alla stima di rimanere invariante rispetto al numero di interlocutori presenti nella scena. Grazie a queste informazioni, il mio algoritmo consente di classificare le azioni dell'interlocutore in relazione alla scena; classifica inoltre il numero di interlocutori presenti nella scena, la loro posizione e gli eventuali oggetti presenti. Considerando lo scenario applicativo della mia tesi, si è deciso di classificare le azioni nelle seguenti classi di macro-azioni osservabili: azioni di comunicazione espressiva; azioni di comunicazione ricettiva; azioni di interazione sociale; azioni di imitazione; azioni cognitive; azioni di gioco.

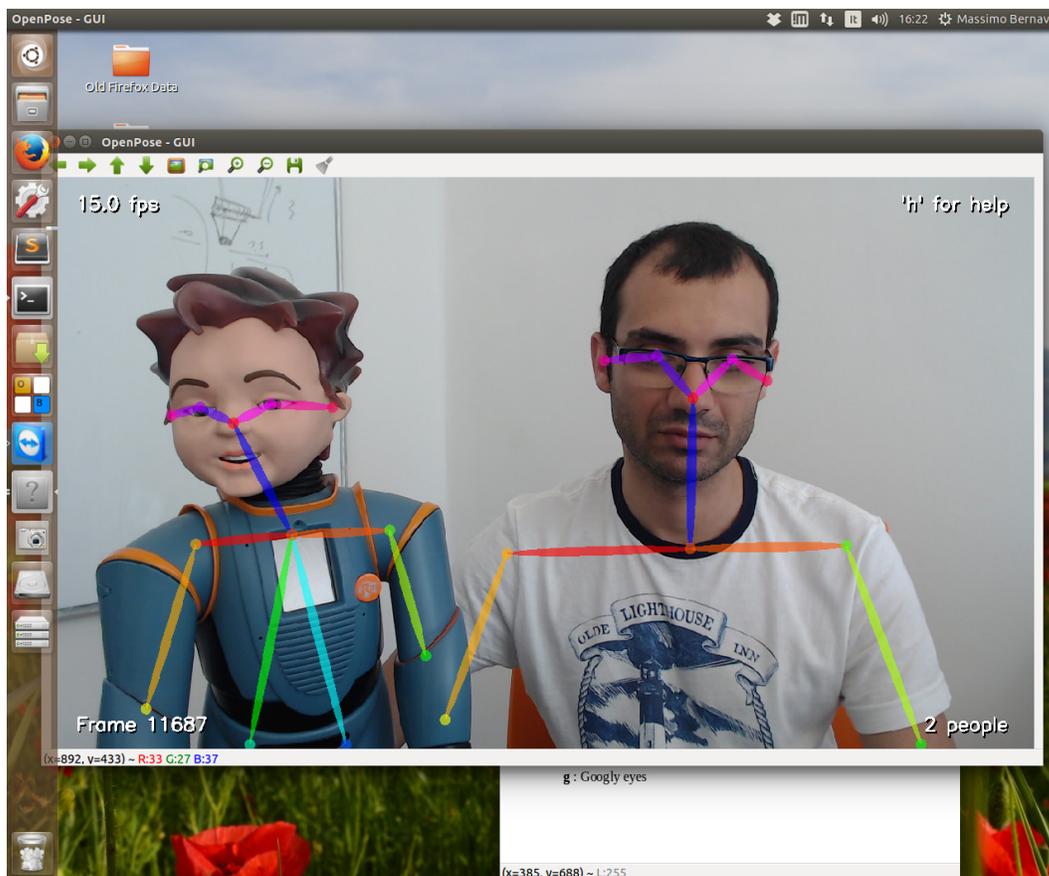


Figura 6 - OpenPose

b) *Componente di percezione dello stato affettivo corrente dell'interlocutore.* Basandomi su librerie allo stato dell'arte della *computer vision* ho sviluppato un algoritmo di riconoscimento delle espressioni facciali. L'algoritmo integra una libreria già sviluppata e testata da CNR Isasi. La libreria è stata sviluppata su un modulo di analisi di video basata un algoritmo riportato in T. Baltrušaitis et al. (2016). Si compone di tre componenti principali: un blocco di rilevazione e tracciamento di *keypoints* facciali (vedi sopra), un sistema di stima della posa, un blocco di tracciamento del punto di mira visiva (*gaze*) ed un sistema di valutazione della presenza e dell'intensità delle *Facial Action Unit* (Ekman et al., 2013). Il sistema rileva il volto, sulla base delle informazioni illustrate nel punto precedente, e procede con la rilevazione dei *keypoints* facciali; grazie a questi ultimi il

sistema calcola la posa della testa e stima il punto di mira visiva, che a loro volta vengono utilizzati per calcolare l'intensità delle *Facial Action Unit*. L'affidabilità di un classificatore di azioni dipende largamente dai dati di training e dalla sua abilità a stimare le espressioni facciali di un soggetto quando la sua espressione neutrale è sconosciuta. La soluzione proposta sfrutta l'idea di introdurre un'unità di azione facciale real-time e il sistema di rilevazione di eventi in base a geometria e aspetto. Il primo passo per un corretto rilevamento della presenza e dell'intensità di un AU è il *mapping* di una faccia rilevata in un contesto di riferimento comune. A tal fine i punti di riferimento attualmente rilevati vengono trasformati in una rappresentazione del punto di riferimento frontale da un'espressione naturale. La dimensione di un vettore finale di 1391 elementi che descrivono il volto e susseguentemente ridotto al 1391 elementi dall'approccio del *Principal Component Analysis (PCA)*. Il vettore completo delle *features* è fatto quindi dalla concatenazione delle *features* di geometria e aspetto. Per tener conto delle differenze personali, il valore mediano delle *features* è sottratto dai valori stimati nel corrente *frame*. L'ultimo passo per la stima dell'intensità e della rilevazione di un AU, viene ottenuto, rispettivamente, con un *Support Vector Machines (SVM)* e un *Support Vector Regression (SVR)*. In entrambi i casi sono impiegati *kernel* lineari. I modelli usati in questo approccio sono addestrati su *dataset DISFA* (S. M. Mavadati et al., 2013), *SEMAINE* (G. McKeown et al., 2010) and *BP4D* (Zhang et al., 2014). Quando le etichette dell'*AU* si sovrappongono a *dataset* multipli l'apprendimento è effettuato congiuntamente.



Figura 7

In aggiunta a questo lavoro sto integrando segnali provenienti da un braccialetto commerciale al fine di fornire anche l'informazione fisiologica dell'interlocutore al quale è richiesto di indossare il braccialetto. A questo fine ho integrato algoritmi già presenti e testati nel centro di ricerca CNR di Messina che segue il mio lavoro di tesi. Questi metodi consentono di ottenere un livello di agitazione/stress dell'interlocutore.

Ho sviluppato la componente dell'interpretazione delle variabili di valutazione nella forma di un sistema di supporto alle decisioni gerarchico. Il sistema riceve quale ingresso le variabili estratte dalla componente di acquisizione di dati relativi all'ambiente ed al comportamento dell'interlocutore, nonché le variabili estratte dalla componente di percezione dello stato affettivo corrente dell'interlocutore (entrambe esplicitate nella sezione precedente). Il sistema si compone di un primo livello costituito da una *Support Vector Machines (SVM)* al quale fa seguito un modulo basato su un modello *Support Vector Regression (SVR)*. Il sistema, nell'ambito della macro-azione osservata, elabora le intensità e frequenza delle seguenti variabili di valutazione: contatto oculare con il robot; allineamento del corpo verso il robot; distanza fra gli interlocutori ed il robot; alternanza del turno nella relazione; rivolgere oggetti al robot; esplorare gli oggetti; latenza tra lo

stimolo e la risposta del bambino allo stimolo; coerenza della risposta allo stimolo; seguire lo spostamento dello sguardo dell'interlocutore verso un oggetto (attenzione congiunta); *pointing* (un interlocutore indica un oggetto nella scena); uso funzionale degli oggetti; uso simbolico degli oggetti; stereotipie gestuali; condivisione del divertimento. Ho iniziato ad addestrare (sto continuando) con i suddetti modelli sul *database* di video acquisiti nel corso di terapie di trattamento cognitivo-comportamentale con il robot *Nao* svolte da terapisti esperte su bambini affetti da disturbi dello spettro autistico. Per ogni video è stato effettuato un processo di *labeling*, cioè l'annotazione strutturata del video.

Ho progettato la componente d'elicitazione delle emozioni che, sempre nella forma di sistema di supporto alle decisioni gerarchico, si occupa della valutazione di come far agire il robot sulla base di un processo emozionale basato sull'applicazione: strategia abilitativa di minori con disturbi del neurosviluppo. È stato fondamentale, nel corso degli studi, riuscire a capire come la scienza affettiva sia frammentata in varie scuole di pensiero sull'elicitazione di emozioni, che includono: (1) le teorie delle emozioni di base, che sostengono l'esistenza di un numero di emozioni universali ed altre in forma di prototipi che si evolvono nel tempo (ad es. Ekman, Izard); (2) le teorie costruttiviste, la quale vedono le emozioni come una valutazione soggettiva basata sui segnali fisiologici neurobiologici (ad es. Russell, Barrett); (3) le teorie valutative (*appraisal theories*), che si focalizzano appunto su un processo valutativo delle emozioni, dove gli stimoli sono elaborati tramite diverse variabili di valutazione (ad es. Scherer, Roseman, modello *OCC*). Nel nostro caso, l'idea è che gli stati emotivi siano causati da una valutazione (o appunto *appraisal*) delle situazioni in cui il robot è coinvolto.

Tra queste, le *appraisal theories* sono le più popolari nell'ambito della ricerca in robotica o nell'ambito della realtà virtuale, mentre solo un piccolo numero di ricercatori ha usato le teorie di base o quelle costruttiviste (ad es. i robot Kismet (Breazeal 2003) e Max (Becker et al. 2007)), sulla base dell'ipotesi dei *markers* somatici di Damasio (Damasio 1995). Per quanto riguarda le *appraisal theories*, il modello *OCC* (Ortony et al., 1988) sembra essere il più popolare tra i ricercatori informatici, con la maggior parte delle

implementazioni che utilizzano questo modello (ad es. *GAMYGDALA* (Popescu et al. 2013), *GALAAD* (Adam 2007), *EVA* (Kasap et al. 2009), *FAtiMA* (Dias et al. 2011), il modello di Dastani (Steunebrink et al. 2012)). Il modello *OCC* indica come la forza di una data emozione dipenda principalmente dagli eventi, dagli agenti o oggetti nell'ambiente degli agenti che esibiscono emozioni. Il modello specifica circa 22 categorie di emozioni e consiste di cinque processi che definiscono l'intero sistema: dalla categorizzazione iniziale di un evento al comportamento finale di un agente (nel nostro caso di un robot). Questi processi sono: 1) classificazione di un evento, azione o oggetto; 2) quantificazione dell'intensità delle emozioni rilevate; 3) interazione fra l'emozione generata con l'emozione esistente; 4) *mapping* dello stato emozionale rispetto ad una espressione emotiva; 5) espressione dello stato emozionale. La semplicità e la struttura ben definita di questo modello, lo rendono accessibile ai ricercatori informatici e le sue proprietà possono essere facilmente modellate in qualsiasi linguaggio di programmazione. Tuttavia molti psicologi considerano questo modello antiquato. Questa impopolarità in psicologia del modello *OCC* ha messo sotto pressione molti informatici per sviluppare altri modelli di valutazione emotiva. Il modello di Scherer's è stato usato da *GRACE* (Dang et al. 2008), *iGrace* (Saint-Aim et al. 2009), *Soar-Emote* (Marinier et al. 2009); Il modello di Roseman è stato usato da *Cathexis* (Velásquez 1996) e *Mary* (Fagel et al. 2016); e il modello di Lazarus è stato usato da *Coppèlius'Concoction* (Hoorn et al. 2012), *EMA* (Marsella & Gratch 2009), e *MAMID* (Hudlicka 2008a).

La frammentazione nelle teorie sull'elicitazione delle emozioni, e la complessità di molti modelli causano varie differenze di interpretazione tra l'ingegneria informatica, le scienze cognitive e la psicologia, così che molti ricercatori preferiscono seguire l'interpretazione delle teorie fatte da colleghi nel loro dominio, invece di seguire il corrente stato dell'arte. Questo trend ha fatto sì che molti informatici hanno scelto di default il modello *OCC*, senza affrontarlo con il giusto approccio critico. D'altra parte ciò è facilmente comprensibile, in quanto risulta l'unico modello che fornisce delle chiare specifiche sul processo di elicitazione delle emozioni. Altri ricercatori hanno tentato di sviluppare modelli computazionali basati su altre teorie ben conosciute (Scherer, Roseman, Lazarus).

Tuttavia a causa della mancanza di chiare e distinte definizioni operative e di specifiche dei concetti affettivi e delle loro conseguenze, molte implementazioni risultando troppo semplicistiche perdono in credibilità; nei fatti quindi, molte assunzioni possono comunque essere ricondotte al modello *OCC*.

Ho cercato di affrontare questa sfida raggiungendo un consenso con alcune delle psicologhe che lavorano al CNR e all'Università di Messina su come specificare i valori delle valutazioni delle emozioni che possono essere calcolati esclusivamente durante il trattamento che si focalizza sul protocollo terapeutico. In questo senso, un approccio ispirato essenzialmente sul modello *OCC*, con gli opportuni aggiustamenti, può risultare invece vincente in termini di interazioni a lungo termine che hanno come scopo il supporto ad una terapia.

I tipi di emozioni che il robot può elicitarne sono state organizzate nel seguente modo:

1. emozioni da attrazione verso l'interlocutore: l'emozione scaturisce dalla valutazione dell'espressione facciale degli interlocutori (bambino in trattamento e genitori) come piacevole o meno (PIACEVOLE SI/NO e intensità) rispetto alle attività valutate dalla componente di acquisizione di dati relativi all'ambiente ed al comportamento dell'interlocutore;
2. emozioni basate su eventi: l'emozione scaturisce dalla valutazione, da parte del robot, di un evento, che può essere desiderabile o indesiderabile (DESIDERABILE SI/NO e intensità) rispetto ai suoi obiettivi terapeutici all'interno della macro-area rilevata;
3. emozioni da attribuzione: l'emozione scaturisce dalla valutazione dell'azione come apprezzabile o meno (APPREZZABILE SI/NO e intensità), rispetto all'insieme dei comportamenti del robot nella macro-area in oggetto;

La componente di memoria affettiva memorizza e richiama le informazioni che vengono dai vari componenti, nonché dal protocollo di trattamento. Questi processi di memoria

sono fortemente influenzati dall'effetto risultante della memorizzazione o del richiamo soggettivo. La componente di memoria affettiva simula elementi della memoria semantica umana, che memorizza le informazioni fattuali; al contempo svolge il ruolo (che è tipico della memoria episodica umana) memorizzerà le esperienze del robot.

La componente di valutazione delle conseguenze è responsabile del valutare e determinare le conseguenze delle azioni. Tutte le potenziali azioni e le strategie di regolazione (che possono servire come risposta sugli stimoli percepiti) sono valutate e priorizzate, basandosi sulle conseguenze aspettate.

La componente di regolazione dell'affetto è responsabile della regolazione dello stato affettivo corrente o di anticipare lo stato affettivo futuro abilitando un'azione esterna per affrontare la causa dell'emozione o modulando l'affetto interno e le componenti cognitive. La componente di selezione dell'azione può essere vista come parte del sistema motivazionale del robot o come parte del meccanismo di elicitazione delle emozioni. Questo componente sarà responsabile per provocare la tendenza a una determinata azione basandosi sugli eventi esterni, interni o associati con il processo di elicitazione delle emozioni. La componente di generazione dell'azione è responsabile della generazione dei comportamenti del robot legati alle espressioni facciali e corporee del robot, basandosi sulle informazioni derivanti dal componente di selezione dell'azione.

IV – Protocolli utilizzati per l’acquisizione dei dati di apprendimento del robot

Dopo aver progettato l’architettura in grado di sostenere l’interazione uomo - macchina, mi sono concentrato sul sistema di acquisizione dei dati del *training set*, cioè l’insieme dei dati che consentirà al robot di apprendere. Ho quindi deciso, con l’ausilio delle terapisti del CNR, di creare un applicativo mobile che potesse guidare il robot tramite vari protocolli psicologici e raccogliere dati di apprendimento per il robot. Prima di procedere ho ritenuto utile fare una riflessione sull’utilizzo dei robot nella terapia dei bambini affetti da disturbi del neurosviluppo.

Perché i robot?

La domanda che ci si pone nell’ambito dei robot assistivi è: sono dei validi motivatori? Svolgono effettivamente il compito richiesto?

La motivazione è un aspetto fondamentale nell’approccio ad una terapia o ad un particolare *task*. Vi sono due tipi di motivazione, quella intrinseca e quella estrinseca. Quella intrinseca viene da una persona, mentre quella estrinseca viene da una fonte esterna. Tuttavia quella intrinseca può essere influenzata da fattori esterni come ad esempio dei *feedback* tattili o vocali.

Ad esempio un *feedback* vocale positivo stimola positivamente la motivazione intrinseca dell’utente, tuttavia se quest’ultimo impara ad eseguire il *task* lo stesso *feedback* non ha più nessun effetto sulla motivazione.

Questo da un ruolo fondamentale al motivatore che generalmente è costituito da un terapeuta, che oltre a monitorare la risposta al task, lo modifica in base alle necessità del paziente. In questa tesi si propone l'utilizzo di *robot sociali* come motivatori, non in sostituzione del terapeuta, ma come complemento. Tuttavia l'utilizzo di robot umanoidi in terapia è fonte di domande.

L'argomento nell'ultima decade è stato molto discusso, in quanto si è visto uno sviluppo, grazie alle nuove tecniche per l'intelligenza artificiale (come il *deep learning*), significativo da parte dei robot e un incremento massimo delle ricerche a tal fine. Ad esempio Fasola e Mataric (2010) hanno descritto l'implementazione di un robot sociale in grado di monitorare le performance di utenti universitari combinando *task* fisici e *task* cognitivi allo scopo di migliorare le performance sui *task* richiesti.

Vi sono molti fattori tuttavia che contribuiscono alla creazione di un rapporto, fattori descritti da Bickmore e Picard (2005) in relazione al rapporto umano – computer, dettagli che si possono applicare anche all'interazione umano – robot.

Questi includono caratteristiche come l'empatia, l'umorismo, la fiducia, la gentilezza, la conoscenza reciproca. Su questi caratteri poniamo molta importanza per la costruzione di relazioni e conseguentemente hanno un uguale valenza per costruire un'interazione credibile umano – robot.

Eliminare la ripetitività nelle istruzioni o nelle risposte del robot diventa fondamentale per far sì che l'utente non percepisce quest'ultimo come ripetitivo, prevedibile e poco credibile, questo può portare ad una diminuzione della percezione, da parte dell'utente dell'intelligenza del robot, tuttavia poiché il campo d'uso, nel nostro specifico, è rapportato con soggetti affetti da disturbo del neuro sviluppo, la ripetitività e la prevedibilità possono essere dei fattori che giocano un ruolo importante nello sviluppo del *task*, tuttavia la ripetitività come precedentemente detto può portare ad una perdita di credibilità e il bilanciamento tra questi fattori diventa quindi fondamentale.

Indagare sui principi di progettazione per un'interazione a lungo-termine diventa un *must* per costruire relazioni importanti, poiché i bambini che necessitano di attenzioni particolare possono potenzialmente beneficiare di un sostegno sotto forma di robot, purché questo diventi un loro “amico” e instauri con loro un rapporto di fiducia, come ad esempio

è stato svolto da Jansen et al (2011) nel progetto *Aliz-e*, dove è stato introdotto il robot come *caregiver*, nel caso di bambini diabetici che dovevano passare una settimana in ospedale e quindi lontano da casa.

Robben (2011) ha eseguito una ricerca sul ruolo di amico o “*buddy*” del robot. Una parte essenziale dell’interazione tra bambino e il robot è il “*bonding*”. Per *bonding* ci si riferisce alla creazione di una sensazione di amicizia e fiducia tra il bambino e il robot. Mentre il bambino gioca con il robot, quest’ultimo tramite il conteggio del numero di sorrisi cercherà di capire se il bambino sta passando del tempo piacevole. Il robot tuttavia va adattato sul carattere del bambino. Nello specifico il robot si deve adattare al livello di introversione o estroversione mostrato dal bambino.

Gli indicatori per l’introversione o estroversione sono il contatto visivo, il livello di energia o la velocità di movimento. Il riconoscimento del volto viene utilizzato per stabilire se il bambino ama giocare. Il robot in base a questi parametri deve adattare il dialogo e modula la propria personalità, come ad es la velocità e il contenuto dello *speaking* e la velocità di movimento, per adattarsi al bambino con il quale interagisce. La ragione che sta alla base di questa ricerca è che i bambini si legano meglio con il robot quando questo presenta caratteristiche caratteriali simili. Questo fenomeno è noto come ‘*similarity attraction*’. Fasola e Mataric (2010) indicano varie variabili che contribuiscono alla motivazione intrinseca: la competizione, la gratificazione, il *feedback* in tempo reale delle prestazioni, l’efficacia e la determinazione. Vallerand et al (1986) descrivono diverse variabili che invece riducono la motivazione intrinseca e dovrebbero pertanto essere evitate. Tra queste variabili si annoverano: ricompense materiali, sorveglianza, scadenze, mancanza di autodeterminazione e *feedback* negativi sulle prestazioni. La sfida ottimale è un fattore che dobbiamo tenere più in considerazione: quando un’attività è troppo facile, l’utente si annoierà mentre se un *task* è troppo difficile l’utente diventerà frustrato o ansioso. Andrade e Corruble (2005) descrivono questo fenomeno come “*game balancing*”. Questo bilanciamento consiste nel cambiare parametri, scenari e comportamenti per evitare il caso estremo del giocatore che diventa frustrato perché il compito assegnato è troppo difficile o che si annoia perché il *task* è troppo semplice.

Le ricerche sul ruolo educativo del robot allo stato dell'arte sono infatti ancora in via di sviluppo. Jansen et al (2012; 2013) hanno investigato se la motivazione intrinseca nello svolgere un gioco cresce con un robot adattivo alla performance rispetto ad un robot che non è adattivo.

Un aspetto rivelatosi particolarmente importante in questa ricerca è stato l'adattività nei dialoghi. I dialoghi sociali includono cose come i saluti, gli argomenti di tipo generale come il tempo (argomento particolarmente gradito agli inglesi) o come scambiarsi preferenze personali (Higashinaka et al, 2008). La *self-disclosure* e l'*empatia* sono riconosciuti come fattori che contribuiscono notevolmente alla conversazione (Altman and Taylor, 1973). Questo effetto è più forte se viene utilizzato un robot umanoide rispetto ad un agente virtuale. Ci sono svariati fattori, come precedentemente detto, che influenzano una conversazione, ma se un robot mostra empatia, *self-disclosure* ed *embodiment* fisico si ha un buon punto di partenza per una valutazione formativa e questi aspetti possono essere integrati facilmente nel dialogo e nel comportamento di un robot. Nello specifico svariati studi hanno mostrato che la *self-disclosure* ha un ruolo centrale nello sviluppo delle relazioni (Collins e Miller, 1994; Laurenceau, 1998). La *self-disclosure* è definita come "la condivisione di informazione con altri che normalmente non sapevano o non scopriranno". Quando qualcuno si impegna in questa condivisione si pretende implicitamente che anche l'altro partner di conversazione divulghi informazioni (norma di reciprocità). Quest'effetto di condivisione tuttavia dipende dalla simpatia che si prova per una persona. Le persone che si impegnano in informazioni intime tendono ad essere più gradite ed affidabili di quelle che invece rivelano informazioni meno personali o più superficiali. Le persone tendono a preferire i robot quando si rivelano informazioni affettive e non relative ai task in attività collaborative (Siino, 2008). La *self-disclosure* reciproca tra il bambino ed il robot può contribuire alla profondità e alla qualità della loro relazione.

Un altro ruolo chiave nelle terapie incentrare sul paziente è l'empatia perché implica l'apprendimento del mondo intimo dell'altro e la comprensione comune delle emozioni (Tapus e Mataric, 2007). Una delle definizioni più complete dell'empatia è la seguente: "La capacità di assumere il ruolo dell'altro, di adottare prospettive alternative a se stessi

e di comprendere le reazioni emotive degli altri in relazione al contesto fino al punto di eseguire movimenti corporei simili a quelli altrui”. I robot per ovvi motivi non possono sentire empatia ma la possono emulare nel loro comportamento come ad esempio mostrando preoccupazione o esprimendo sentimenti negativi in risposta ad un malessere altrui. Un’ultima caratteristica in grado di influenzare un dialogo è *l’embodiment*. I robot sociali non necessitano obbligatoriamente di un corpo fisico per interagire con i propri utenti, poiché un dialogo può essere eseguito anche da un *avatar virtuale 3D* ma avere una forma fisica offre vantaggi sostanziali poiché un robot “fisico” è più attraente e percepibile dal mondo esterno e l’impressione dei partecipanti è notevolmente influenzata dalla presenza fisica. Wainer et al (2007) hanno condotto uno studio utilizzando un robot-allenatore dedicato agli anziani e i partecipanti hanno fortemente preferito un robot fisicamente incarnato rispetto a quello simulato, è stato inoltre osservato che il robot fisico ha incontrato un maggior grado di impegno da parte degli utenti (Deshmukh, 2012). È stato anche dimostrato che l’effetto di avere un corpo fisico è stato prevalente anche quando il robot è stato mostrato in remoto tramite telecamera. La presenza sociale di un robot fisico remoto è stata quasi identica ad un robot fisicamente vicino all’utente (Powers et al, 2007). Il tutto dimostra la differenza positiva che *l’embodiment* ha sulla relazione sociale.

Nel progetto Aurora è stato investigato l’uso potenziale del robot come terapeuta o “gioco educativo” con un focus particolare ai bimbi affetti da sindrome autistica.

In accordo con la *National Autistic Society* nei soggetti affetti da ASD sono compromesse l’interazione sociale, la comunicazione sociale e l’immaginazione (Wing, 1996).

Inoltre, le persone affette da autismo mostrano uno scarso contatto oculare e raramente si lasciano coinvolgere in giochi interattivi. Un dato statistico significativo indica che i soggetti affetti da autismo sono in prevalenza uomini.

L’idea di utilizzare un robot in terapia nasce dal fatto che i soggetti affetti da ASD si sentono più a loro agio in ambienti ben definitivi e con comportamenti prevedibili.

Murray (1997) ha affermato che le persone con autismo tendono a fissare lo sguardo su oggetti isolati dall’area circostante. I *computer*, secondo la tesi di Murray, possono irrompere in questo mondo mettendo il focus dell’attenzione sullo schermo, permettendo

in tal modo all'utente di ignorare eventi esterni. Murray argomenta che i computer durante l'educazione e la terapia possono essere un beneficio per l'autoconsapevolezza e l'autostima. Inoltre sono in grado di motivare un individuo a parlare, leggere o condividere informazioni. Hershkowitz (2000) ha fortemente argomentato l'uso dei computer nell'insegnare il linguaggio e capacità scolastiche nei bambini con autismo (Howlin et al., 1999).

Il dialogo non è l'unica componente dell'interazione, questa comprende anche il linguaggio del corpo o la gestualità e molti di questi sono espressi in maniera subconscia. L'evoluzione umana e il nostro sviluppo hanno sintonizzato il nostro sistema cognitivo e percettivo a captare una moltitudine di segnali sociali.

Differentemente da come noi interagiamo con gli oggetti fisici durante i nostri primi 4 anni di vita sviluppiamo la capacità di "leggere la mente", imparando a predire ed interpretare i comportamenti umani in termini di stati mentali.

I *deficit* in quest'abilità, come è stato dimostrato nelle persone con autismo, rendono i comportamenti sociali molto imprevedibili.

L'utilizzo dei robot o di specifici *software* con soggetti affetti da ASD è stato investigato molto nell'ultima decade. I sistemi *software* includono ambienti virtuali altamente strutturati e vengono utilizzati dai terapisti e dagli insegnanti come strumenti in grado di insegnare *skills* sociali o altri tipi di *skills* (come il riconoscimento di emozioni, o come imparare ad attraversare la strada (Strickland, 1996).

Dai primi anni 80 l'utilizzo dei robot nel campo educativo è divenuto popolare.

L'utilizzo dei robot nelle terapie dei bambini affetti da autismo è stato invece studiato più di recente (Dautenhahn e Weey, 2000; Werry et al., 2001; Weir, 1976). Ad esempio un primo lavoro svolto da Weir e Emanuel (28) ha suggerito effetti positivi su un bambino di 7 anni affetto da autismo. Michaud e Theberge-Turmel hanno studiato vari *designs* con differenti capacità di interazione in grado di interagire con i bambini, ad esempio un elefante, una sfera robotica, o un robot a mezzo busto (Michaud et Théberge-Turmel, 2002).

Tuttavia nonostante tutti gli studi ancora troppo poche le evidenze circa gli effetti terapeutici che possono essere collegati ai robot.

Usare un computer o un ambiente virtuale tuttavia è differente dall'utilizzare un robot poiché questo incorpora gli aspetti dell'interazione "viso a viso" tra umani, come investigato da Dautenhahn e Werry (2004), che hanno studiato vantaggi e svantaggi dell'utilizzo di diverse tecnologie.

I vantaggi nell'utilizzo dei robots sono che quest'ultimi forniscono un ambiente sicuro, semplificato, prevedibile e affidabile dove la complessità dell'interazione può essere controllata e gradatamente incrementata. Similarmente studi psicologici hanno mostrato che i bambini con autismo preferiscono giocattoli dall'aspetto semplice ed ambienti prevedibili (Ferrara e Hill, 1980) che possono fornire dei buoni punti di partenza per gli interventi terapeutici. Tuttavia il ruolo del robot è quello del mediatore, che ha come scopo primario quello di non rimpiazzare il contatto umano ma facilitarlo. Robins et. al (2004) hanno mostrato come un piccolo umanoide può fornire un divertente *focus* d'attenzione in grado di rivelare competenze sociali e comunicative dei bambini con autismo. Alla luce di queste ricerche l'uso clinico dei robot interattivi promette uno sviluppo significativo e dimostra che gli individui con ASD presentano punti di forza nella comprensione del mondo fisico mentre mostrano debolezze nella comprensione del mondo sociale (Klin, 2000), sono più sensibili alle risposte, anche a quelle sociali, quando sono somministrate tramite la tecnologia piuttosto che ad un essere umano (Ozonoff, 1995), sono intrinsecamente più interessati al trattamento quando questo coinvolge componenti elettronici o robotici (Robins et al., 2005). Tuttavia la maggior parte del supporto fino ad oggi per il suo utilizzo in terapia si basa su prove aneddotiche e manca il sostegno alla generalizzazione delle competenze (Ricks e Colton, 2010). Molta attenzione è stata data a quale tipo di robot (umanoidi vs non umanoidi, potrebbe essere efficace, ma l'accento non è stato posto sui modi migliori per integrare i robot in sessioni di terapia. Vi sono diverse domande aperte su come poterlo fare, ad esempio quali possono essere i migliori ruoli del robot in terapia e quali soggetti affetti da ASD si adattano meglio ad un trattamento del genere. Una questione fondamentale è se gli individui con ASD preferiscono i robot o le caratteristiche robotizzate alle caratteristiche umane o ai giocattoli ed in caso affermativo cosa rende attraente queste caratteristiche? Alcuni ricercatori hanno sostenuto che anche se i robot non umanoidi sembrano avere un *appeal* più generale in

relazioni ai soggetti ASD, i robot simili all'uomo avrebbero maggiori potenzialità di generalizzazione. Considerare queste questioni diventa fondamentale, a causa dell'eterogeneità della popolazione ASD, poiché alcuni potrebbero non preferire i robot o i robot umanoidi.

È stato mostrato da un piccolo studio condotto da Pioggia et al. (2005) che i soggetti con ASD mostrano risposte al robot sia di carattere positivo che di carattere negativo, successivamente uno studio *follow-up* ha trovato dei miglioramenti nella comunicazione sociale. Dautenhahn et al hanno condotto una serie di studi che li ha portati alla conclusione che alcuni individui preferiscono i robot ai giocattoli non robotici o agli umani. Due studi (Bird et al, Pierno et al, 2008; Bird et al., 2007) hanno trovato un miglioramento del movimento da parte dei bambini subito dopo che lo stesso era stato eseguito dal robot, questo suggerisce che si possono avere dei benefit da *task* di imitazione che coinvolgono robot.

Una potenziale applicazione prevede la creazione di un ambiente dove un robot potrebbe modellare comportamenti specifici per il bambino (Dautenhahn, 2003) o il bambino potrebbe praticare competenze specifiche con il robot (Heyes, 2001). L'obiettivo sarebbe quello di insegnare un'abilità che il bambino potrebbe imitare o imparare, e trasferirla alle interazioni con gli esseri umani, il robot diventa quindi attivo nell'insegnamento.

Imitazione

Il primo protocollo implementato tramite la mia *app* è un protocollo d'imitazione.

Gli psicologi non usano una definizione univoca d'imitazione. Ad esempio, sostengono che i neonati imitano perché sono in grado di eseguire la sequenza bimodale percezione – azione (reagiscono ad un movimento visto eseguendo un movimento che lo copia), e i bambini con autismo non imitano perché non sono in grado di riprodurre sequenze

complesse di azioni. Quello che quindi viene chiamato imitazione è l'abilità di imitare nuove strategie e sequenze di azioni. (Heyes, 2001).

Una definizione usata prima dell'introduzione di una prospettiva di sviluppo nello studio dell'imitazione (Aronfreed, 1969).

Queste esitazioni su come definire l'imitazione viene in sostegno dell'idea che l'imitazione non è un fenomeno unitario.

C'è un grande corpo di esperimenti psicologici e di *neuroimaging* che dimostrano che l'azione della percezione condivide alcuni meccanismi neurali e cognitivi con azioni di creazione, simulazione, riconoscimento e in una certa misura d'imitazione (Decety & Grezes, 1999). Questi risultati sono molto rilevanti per capire i meccanismi coinvolti nell'imitazione. Sulla base di queste informazioni, i neuroscienziati hanno proposto il concetto di rappresentazioni motorie condivise (Georgieff & Jeannerod, 1998).

Esaminando con attenzione questa ipotesi si può dire che esiste una gerarchia di meccanismi coinvolti in differenti tipi di imitazione che hanno tutti in comune il dover reagire alla percezione di un movimento mirato ad un obiettivo o all'azione di produzione di movimenti simili. Questa visione prende in considerazione l'implicazione dei nuovi risultati nella neurobiologia scoperti da Rizzolatti e colleghi (Fadiga, Fogassi, Pavesi & Rizzolatti, 1995; Iacoboni et al, 1999; Rizzolatti, Fadiga, Fogassi & Gallese, 2002) la quale hanno scoperto nella scimmia e successivamente nella corteccia umana premotoria, una classe di neuroni che hanno denominato "neuroni specchio" poiché vengono attivati quando un'azione viene eseguita o osservata su un oggetto. Questa capacità neurale di rispondere all'azione può essere coinvolta nell'imitazione ad alto livello, infatti essa non necessariamente porta alla riproduzione dell'azione osservata. Quando i neuroni specchio si attivano per l'azione, viene prodotta un'attività neurale che corrisponde alla rappresentazione dell'attività neurale generata dalla produzione effettiva dell'azione osservata. Secondo Rizzolatti e colleghi (2002) i neuroni specchio che si attivano per le azioni con oggetti (F5) sono il miglior esempio conosciuto del sistema di risonanza dello specchio ma ci sono dei neuroni che si attivano quando semplici movimenti sono eseguiti dopo essere stati osservati. La risonanza di quest'ultimi può spiegare imitazioni a basso livello come la facilitazione sociale o l'imitazione neonatale.

La distinzione proposta da Rizzolatti et al. (2002) ci porta a disegnare un continuum tra le risposte incontrollate e quelle intenzionali come risposta ad un'azione, piuttosto che escludere dalla definizione di imitazione quei comportamenti che non sono informati dall'intenzione di imitare. La recente prospettiva di Byrne e Russon (1998) evita anche una chiara distinzione. Loro non negano la definizione di imitazione ad alcuni comportamenti ma piuttosto distinguono tra due livelli di imitazione: uno a basso livello che raggruppa i comportamenti primari ed uno ad alto livello che raggruppa l'intuizione creativa circa gli obiettivi interessati. Dautenhahn e Nehaniv (2002) e Mitchell (2002) condividono la visione di un continuum in una gerarchia dei livelli di imitazione. L'imitazione è stata a lungo vista dai comportamentalisti come richiesto dall'obiettivo dell'apprendimento: la procedura "guarda me e fai come me" è una tecnica chiave, l'imitazione è stata definita come apprendimento senza incentivi e senza prove ed errori (Bandura, 1971).

Un'interessante primo passo verso l'idea che gli usi dell'imitazione possono differire in base ai vincoli di sviluppo e alle attuali esigenze di adattamento dello sviluppo dell'infante è stato fatto da Yando, Seitz e Zigler (1976). Questi autori hanno proposto una teoria a 2 fattori dove il livello di sviluppo cognitivo e la motivazione sono stati i fattori essenziali nello sviluppo dell'imitazione. I motivi per imitare differiscono significativamente con l'età, quello che rimane da scoprire è per quale motivo. Nei primi anni 80 alcune voci cominciarono a proporre che l'imitazione ha 2 funzioni: una cognitiva e sociale o una funzione d'apprendimento e comunicativa (Nadel-Brulfert & Baudonnière, 1982). Numerosi ricercatori sono stati coinvolti nell'esplorazione della possibile origine della funzione comunicativa dell'imitazione. Tra loro Maratos (1973), Pawlby (1977) e Uzgiris (1981) sono stati sicuramente pionieri in questo campo. Hanno mostrato che solo poche settimane dopo la nascita, l'imitazione include una turnazione tra i partner (Maratos, 1973; Uzgiris, Broome & Kruper, 1989). Alla nascita la frequenza dell'imitazione è stata suggerita avere un valore predittivo per l'interazione faccia a faccia a 3 mesi (Heimann, 1991) Kugiumutzakis e colleghi (1999), ispirandosi a Trevarthen (1999) hanno dimostrato l'efficienza di un contesto interattivo nella prima produzione dell'imitazione. Nadel (1986) ha mostrato che i bambini molto piccolo approfittano del fatto che l'imitazione

coinvolge due ruoli: imitatore e modello. I bambini di due anni usano questi due ruoli in modo alternato (Nadel, 2002). I *partner* coordinano il loro tempo per ottenere la sincronia tra il modello e l'attività dell'imitatore. Questo porta a scambi di lunga durata che non sono possibili in nessun altro modo prima dell'insorgenza del linguaggio. La distinzione tra le due funzioni dell'imitazione può sembrare inutile dal momento che l'imitazione è di natura sociale e richiede sempre un incorporamento sociale, tuttavia Nadel (2004) ha mostrato come i risultati dell'imitazione come apprendimento sono radicalmente diversi dai risultati dell'imitazione come comunicazione. Imparare tramite l'imitazione è un beneficio per l'individuo ma non implica la condivisione di qualsiasi cosa con il modello, al contrario quando qualcuno comunica con un interlocutore tramite l'imitazione entrambi ne beneficiano in quanto l'imitazione porta cambiamenti in entrambi (Nadel, 1999). In altre parole, colui che viene imitato e colui che imita formano un nuovo sistema dinamico, un sistema evolutivo di somiglianze costruite sulla base di due repertori diversi. Questa nuova prospettiva di sviluppo ha rivelato la necessità di allargare la messa a fuoco dell'interesse per l'imitazione nella robotica.

Fino a poco tempo fa i robotici sono stati principalmente interessati all'apprendimento della funzione dell'imitazione: un modo semplice per un robot per imparare da un altro e conseguentemente aumentare la velocità di apprendimento in una popolazione di robot (Berthouze et al., 1998; Kuniyoshi, 1994; Schaal, Atkenson & Vijayakumar, 2000). Se membri diversi di una popolazione robotica possono imparare non soltanto da se stessi ma anche da altri, questo riduce drasticamente la complessità del computo di apprendimento consentendo la diffusione di conoscenze su un dato compito nell'intera popolazione. Così ogni membro può trarre vantaggio della conoscenza degli altri e integrare questa conoscenza nella propria capacità. Questo è un caso di apprendimento osservazionale che a sua volta può accelerare l'apprendimento delle attività successive da parte dell'individuo. Tuttavia anche se incorporato in un ambiente sociale l'imitazione non mette in collegamento gli individui tramite l'interazione. Uno dei pionieri di un approccio focalizzato sull'interazione robotica è Dautenhahn (1995), che ha proposto un'architettura in cui un robot segue un "insegnante" (un altro robot) usando una semplice regola sensoriale-motoria per ridurre al minimo la sua perdita di energia. Questa regola consiste

semplicemente nel mantenere il contatto con un oggetto in movimento. Agendo in questo modo il robot esegue lo stesso movimento del robot “insegnante” e quindi sembra imitarlo. I robotici hanno cominciato a esaminare come due delle loro architetture genetiche possono sviluppare imitazioni a basso livello. Contemporaneamente gli evoluzionisti sono stati coinvolti in studi comparativi dello sviluppo imitativo nei neonati, nei bambini sani e nei bambini con autismo. Se ci sono condizioni biologiche che costringono lo sviluppo dell’imitazione, allora dovremmo trovare una gerarchia unica di comportamento in due casi contrastanti di sviluppo. Ad un basso funzionamento i bambini con autismo, così come i neonati, possono produrre azioni di percezione, accoppiare e imitare i movimenti semplici che vedono senza avere l’intenzione di farlo, senza tener conto del fatto che questi comportamenti siano intenzionali o no. Allo stesso modo possono avere risposte segrete o non specificate al loro essere imitati. Ad un livello più alto di funzionamento i bambini con autismo possono essere in grado di riprodurre l’obiettivo di un modello. Generalmente un bambino di 5 anni presenta degli interessi sensoriali simili a bambini di 8-9 mesi d’età. Ad ogni modo il loro *pattern* sensoriale è più ripetitivo, l’imitazione è selettiva ed è usato con lo scopo di incrementare gli stimoli (Lemay 2004). Essi presentano abilità non utilizzate (ad esempio pianificare e indurre comportamenti), e mostrano anche *deficit* nella condivisione dell’attenzione (evitano il contatto oculare e non rispondono ai sorrisi) e nelle convenzioni (come una scarsa imitazione di espressioni facciali e di gesti come la testa per dire no o agitare la mano per dire ciao) per comunicare interessi comuni (Lemay 2004). Questi deficit sono spiegati da una difficoltà nella percezione e nel comprendere gli stimoli dall’ambiente che influenzano la loro comprensione dei segnali sociali (gesti, parole e intenzioni altrui) (Zilbovicius 2004). Nello sviluppo tipico i bambini di 8-9 mesi dimostrano intenti di comunicazione dagli occhi e dall’imitazione dei gesti. Pertanto supponendo che la complessità dell’interazione umana limiti la comunicazione reciproca da parte dei bambini autistici a causa delle loro disabilità sensoriali, si ritiene che i bambini autistici abbiano bisogno di interventi che tengono conto dei loro personali interessi e dei loro *deficit* nella comprensione utilizzando un semplice e prevedibile mezzo in grado di attirare la loro attenzione e facile da comprendere.

I robot mobili mostrano delle potenzialità in questo senso perché sono prevedibili, semplici e facili da comprendere (Nadel 2004) e possono essere progettati in conformità agli standard particolari e ai deficit di comprensione nei bambini con autismo. Possono generare più interesse ed un'ampia varietà di situazioni d'interesse rispetto agli oggetti statici e portare in gioco competenze di interazione sociale (contatti visivi, imitazione) (Robins et al., 2004b).

Nello specifico ho creato una serie di comportamenti all'interno del robot chiamati randomicamente, suddivisi per modalità d'imitazione:



Figura 8

- imitazione eseguita dopo il movimento del robot
- imitazione eseguita insieme al movimento del robot
- imitazione mista

Al termine di ogni movimento il terapeuta assegna un punteggio che viene registrato dall'App

Queste tre modalità sono state somministrate in ordine crescente durante sessioni di terapia differenti ed è stato monitorato il miglioramento da parte del bambino nell'eseguire i movimenti.

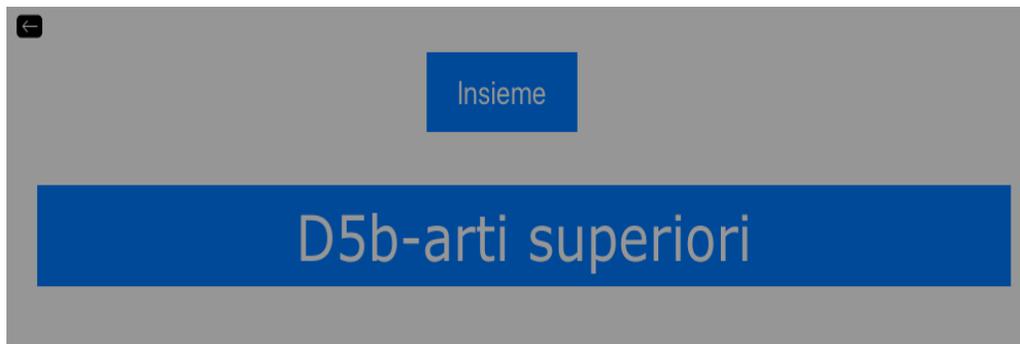


Figura 9



Figura 10

Possiamo vedere una tabella con i risultati del test che evidenzia come la risposta al *task* sia stata soddisfacente (il risultato massimo ottenibile è 4), da notare inoltre i campi evidenziati che riguardano i bambini affetti da ASD a basso funzionamento

D4	D4b	D5	D5b	D17	D17b	D20	D20b	D22	D22b
3	3	3	3	2	2	3	3	2	2
1	1	2	2	1	2	2	2	1	1
1	0	0	0	0	1	1	1	0	1
3	3	3	3	2	2	1	2	2	2
0	0	0	0	0	1	0	0	1	0
2	2	3	2	2	2	2	2	1	1
3	3	3	3	2	2	3	3	2	2
3	3	3	3	2	3	1	1	2	2
3	3	3	2	2	2	3	3	2	3
0	0	0	0	0	1	0	0	1	1
3	3	3	2	2	3	2	2	2	2
3	3	3	3	2	2	3	3	2	2
3	3	1	3	2	2	3	3	2	2
3	3	3	3	2	2	2	3	2	2
1	1	1	0	1	1	0	0	0	1
0	0	0	0	0	0	0	0	0	0

Figura 11 - Risultati task imitazione



Figura 12

Contatto oculare

Il secondo protocollo inserito è relativo al contatto oculare.

Il contatto oculare a volte viene indicato come un faccia a faccia o uno scambio di sguardi (Mirenda et al., 1983) che supporta migliori comunicazioni verbali e non verbali (Carbone et al., 2013). Nei primi stadi di sviluppo i bambini impiegano il contatto oculare per regolare l'interazione sociale faccia a faccia. Più tardi, si occupa di coordinare l'attenzione visiva tra un altro individuo e un oggetto di interesse (Arnold et al., 2000). È stato dimostrato che la regolazione del contatto oculare ha importanti effetti e influenza anche nelle informazioni linguistiche e verbali (Podrouzek et al., 1988). Uno dei primi e più evidenti indicatori dei ritardi nello sviluppo e del disturbo dello spettro autistico è il

disavanzo diadico (occhio – volto) e triadico (attenzione congiunta rivolta ad un terzo o ad un oggetto) nei comportamenti comunicativi sociali. Poiché il contatto oculare coinvolge importanti ruoli sociali, e la mancata emissione di questo importante segnale può avere gravi inconvenienti e implicazioni nella vita educativa, relazione e sociale.

Le prime indagini comportamentali sulle risposte al contatto visivo dimostrano che se i bambini non riescono ad orientarsi verso l'istruttore, probabilmente non riescono a rispondere e apprendere un nuovo concetto (Fox et al., 1977; Helgenson et al., 1989; Greer et al., 2007). Beattie (1981), Lalljee e Cook (1972), Stephenson e Rutter (1974) hanno fatto alcune analisi sull'effetto dello sguardo sulla fluidità del discorso; hanno suggerito che la conversazione è più fluida quando le persone possono vedere le espressioni facciali altrui. Inoltre, lo sguardo sarebbe necessario per far sentire i soggetti a proprio agio durante una conversazione. Tuttavia è anche vero che l'eccessivo contatto visivo riduce la qualità della conversazione.

Considerando i deficit di attenzione di contatto oculare e attenzione congiunta nei soggetti con ASD sono stati condotti diversi studi per migliorare le risposte al contatto oculare di questi soggetti. Gli interventi sono stati principalmente basati su sessioni di terapia "umane". L'utilizzo di *prompt* fisici e vocali può avere successo su alcuni individui ma su altri può avere degli svantaggi tra cui dei comportamenti inferenti all'utilizzo di *prompt* fisici e necessità di rinforzi nei *prompt* vocali che porta ad un rallentamento nell'apprendimento.

All'inizio degli anni 80 una nuova area di ricerca ha indirizzato gli studi sul contatto oculare dove l'obiettivo è stato fissato nell'insegnamento delle abilità di riconoscimento dello sguardo tramite varie strategie sociali interattive (Jones et al., 2004). Sono state utilizzate un insieme di procedure (il ruolo del gioco, l'imitazione contingente, il ritardo temporale e le tecniche di modifica del comportamento naturalistico) e sono state in grado di mostrare degli aumenti moderati in un'ampia varietà di comportamenti sociali, inclusi il contatto oculare e l'attenzione congiunta. In alcuni studi recenti alcune variabili motivazionali ed estrinseche rafforzano (sotto forma di lode sociale o ricompense) sono

state utilizzate per incoraggiare i bambini a fornire più risposte visive. Tuttavia è stato mostrato che anche questo tipo di rinforzo può essere un fattore limitante negli studi sullo sguardo quando si insegnano le risposte all'attenzione condivisa (Whalen & Schreibman, 2003), nonostante ciò questo metodo è ancora ampiamente utilizzato nelle sessioni di terapia. In un recente studio condotto nel 2012, Plavnick e Ferreri (2012) hanno proposto di utilizzare il *mind training* per migliorare le interazioni sociali dei bambini con ASD. Questo un altro approccio che è possibile applicare per migliorare le risposte oculari dei bambini affetti da ASD. L'approccio preso in considerazione in questa tesi è invece quello effettuato tramite robot. Nonostante i molti studi effettuati, come precedentemente descritto, solo pochi di essi utilizzano robot umanoidi per insegnare o praticare competenza di comunicazione sociale. Un recente studio condotto da Ricks (2010) suggerisce che questo approccio può avere un'utilità clinica, rilegando però il limite d'efficacia al periodo dell'infanzia. Bekele e colleghi (2014) hanno studiato lo sviluppo e l'applicazione di un sistema robotico umanoide in grado di gestire in modo intelligente le richieste di attenzione congiunta e rispondere in modo adeguato sulla base delle misurazioni del sistema dello sguardo e dell'attenzione. Hanno scoperto che i bambini con ASD in età prescolare hanno un contatto oculare più frequente verso l'agente robotico umanoide e rispondono con più attenzione agli stimoli congiunti. Ciò suggerisce che i sistemi robotici hanno una maggiore probabilità di successo nel migliorare l'attenzione coordinata.

Si è trattato di un *task* a difficoltà crescente. Tramite un *set* di 5 comportamenti differenti è stato testato l'aggancio oculare del bambino nei confronti del robot.

Il robot in ognuno dei comportamenti richiedeva tramite varie tecniche l'aggancio visivo da parte del bambino. In particolare sono stati utilizzati i *led* oculari presenti nel robot e tramite l'utilizzo di rinforzi nella richiesta e nella risposta alternativamente somministrati dal terapeuta la difficoltà del *task* aumentava.

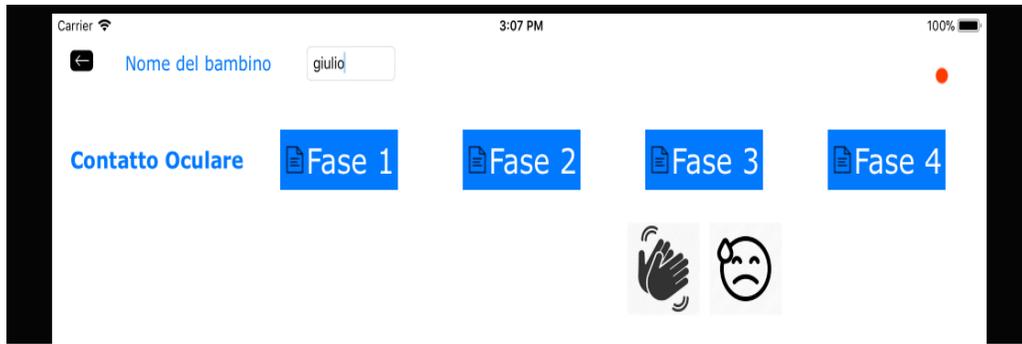


Figura 13



Figura 14

CAA

Comunicare è una funzione fondamentale nelle nostre vite, specialmente se vista in termini di interazione sociale, tuttavia molti bambini hanno difficoltà con questa abilità. Questi bambini possono avere una capacità verbale assente o limitata e spesso richiedono l'ausilio della comunicazione aumentativa assistita (Alant, 1994). E' stato studiato che questi soggetti sono deficitari sia nell'acquisizione del linguaggio che nella funzione di espressione dello stesso. Lo scopo della CAA è quella di aiutare tali bambini ad aumentare il loro output verbale fornendo un modo alternativo per esprimere le loro necessità e i loro pensieri. Tra i soggetti che a volte necessitano di questa tecnica sono presenti anche i bambini affetti da ASD.

Nel corso del tempo sono state utilizzate differenti forme di AAC e inclusi differenti *devices*, con dei target precisi per le persone di tutte le età e con differenti gradi di disabilità o *deficit* comunicativi (McNaughton, 2013; Alant, 1994; Light, 2007; Sigafos et al., 2004). Tuttavia studi recenti hanno mostrato che non vi sono *device* più specifici di altri per ottenere risultati migliori.

Ho potuto notare tramite l'utilizzo di un sistema di CAA presente nell'APP quanto sia fondamentale una fase iniziale di familiarizzazione da parte del robot per l'esecuzione dei tasks.

Inizialmente alcuni bambini che non avevano mai visto *Nao* provavano diffidenza verso di lui, mentre altri lo vedevano come un giocattolo. Durante il trattamento ho potuto notare che il *delay* tra il tempo di scrittura sul *tablet* e l'esecuzione da parte del robot poteva provocare un deficit d'attenzione nell'atteggiamento del bambino. Ho quindi implementato un sistema CAA tramite il quale il terapeuta è in grado di eseguire una conversazione in tempi rapidi e ho potuto notare come i bambini diventano immediatamente "amici" di *Nao*, da qui il sistema di memoria indotto, poiché a ogni interazioni memori delle conversazioni precedenti il CAA veniva ampliato con nuove frasi e risposte e le conversazioni hanno preso continuità nel corso delle settimane.

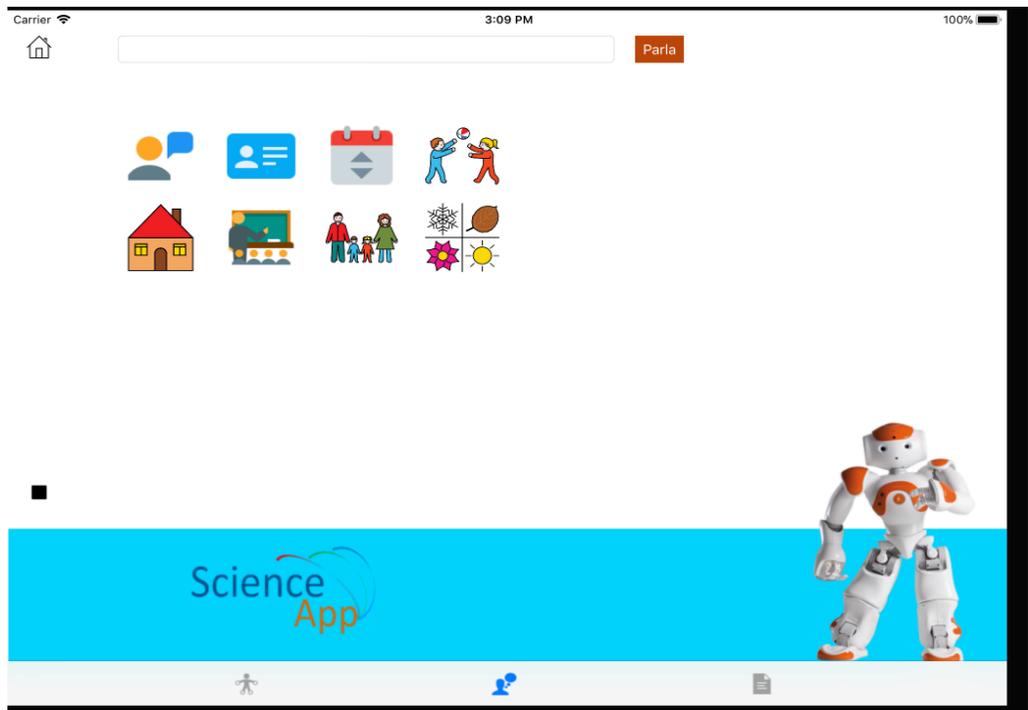


Figura 15

T-Nao

Server

Ho realizzato l'App *T-Nao* per il controllo remote tramite un Sistema client – server. Il server è stato realizzato in *Python*, un linguaggio di *scripting* orientato agli oggetti con licenza *open-source*.

Nel dettaglio questo *server* si basa sul modulo *BaseHTTPServer* la quale definisce due classi per l'implementazione di un web server, nello specifico *SimpleHTTPServer* e *CGIHTTPServer*.

Questa classe è accessibile da un *Handler* che resta in ascolto di eventuali connessioni tramite un *ip address* e una porta *TCP* che possono essere diverse in relazione al robot che si usa. Nella mia sperimentazione ho utilizzato due robot *NAO*, (Leonardo e Romeo) che avevano come *ipaddress* rispettivamente *Romeo.local* e *Leonardo.local*, la porta utilizzata è stata la 3001.

L'*Handler* contiene il metodo *doGet* che gestisce una richiesta *REST*, nello specifico una richiesta *GET*. Tramite la decodifica di questa richiesta vengono letti i parametri che avvieranno uno o più *behaviours* all'interno del robot.

Tra questi parametric possiamo trovare quelli per i comportamenti motori, i parametri per l'avvio di un comportamento di *speaking* o le istruzioni per avviare o fermare le varie modalità di registrazione o semplicemente interrompere i *behaviors* in corso.

Il *server* è stato successivamente installato sul robot e viene avviato automaticamente all'avvio del robot.

Client

il client è stato implementato in un app per sistemi operativi *iOS*.

Ho scelto di creare l'applicativo per *tablet (iPad)* a causa delle dimensioni maggiori che ne permettono una migliore visibilità ed in particolare questo sistema operativo a causa della maggiore standardizzazione dei *device* presenti sul mercato.

Per la progettazione *software* ho utilizzato l'ide *XCode* presente sui sistemi *OSX* e il linguaggio di programmazione *Swift*, in particolare è stato eseguito in corso d'opera un porting dell'applicativo da *Swift 2* a *Swift 3*

Per la connessione con il robot ho utilizzato la libreria opensource *CocoaAsyncSocket*. Si tratta di una libreria per la creazione di *socket* asincroni in grado di comunicare con l'*Handler* del *server* trattato in precedenza

Nel dettaglio la mia implementazione scandaglia la rete alla ricerca di robot presenti sullo stesso *network* e li mostra tramite una tabella presente all'avvio dell'app.

L'utilizzatore potrà quindi riconoscere il robot (in caso vi fossero più robot connessi alla stessa rete *LAN*) e selezionarlo. Questo processo è implementato nella classe *NaoServiceBrowser* presente nell'appendice tecnica.

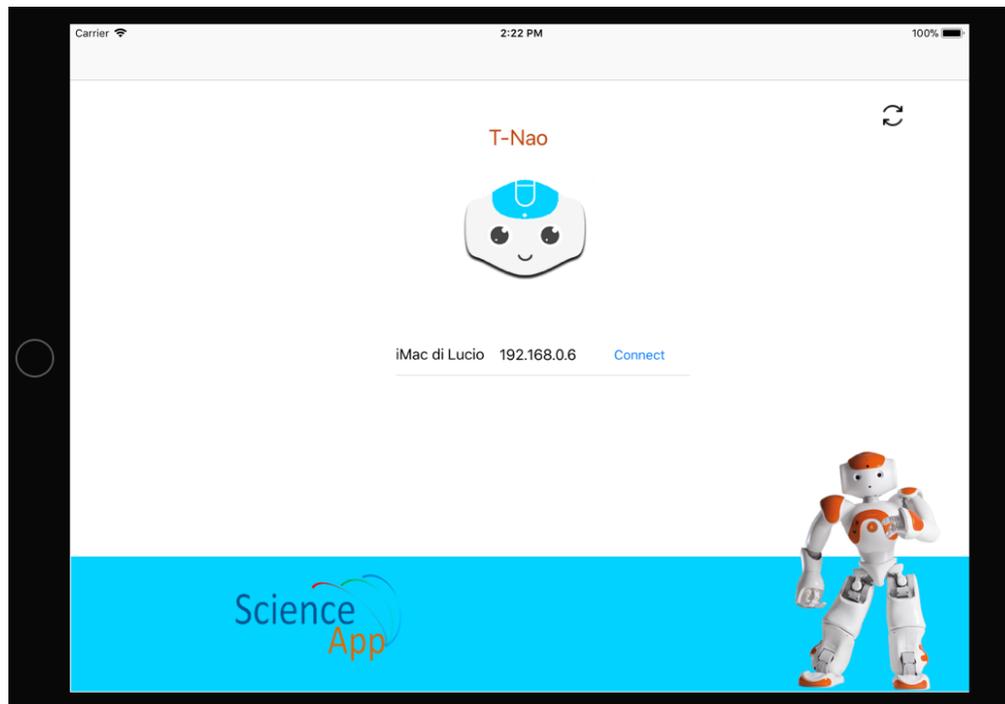


Figura 16

Non appena il robot viene connesso, si ottiene un *feedback* vocale da parte del robot e si apre un menu da dove è possibile scegliere le varie voci del menu presenti

- protocolli
 - FLL (protocollo per il contatto oculare)
 - PhD (protocollo di imitazione)
 - HL (protocollo di imitazione con differenziazione tra disturbi dello spettro autistico ad alto e a basso funzionamento)
- comunicazione rapida
- behaviour di “intrattenimento”, spesso utilizzati alla fine del trattamento

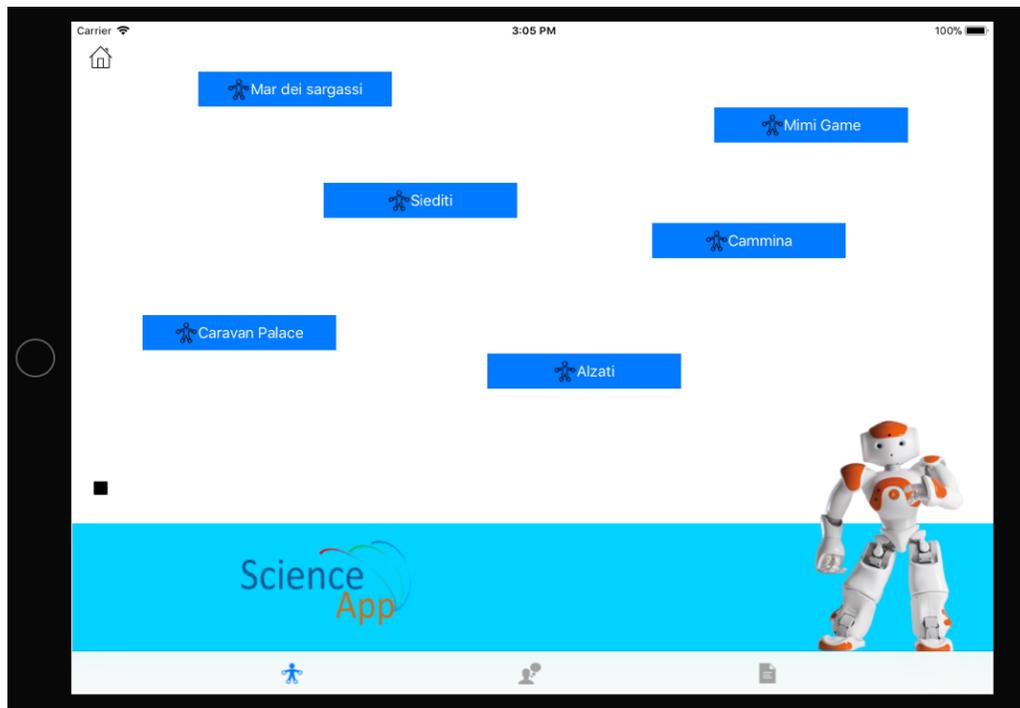


Figura 17

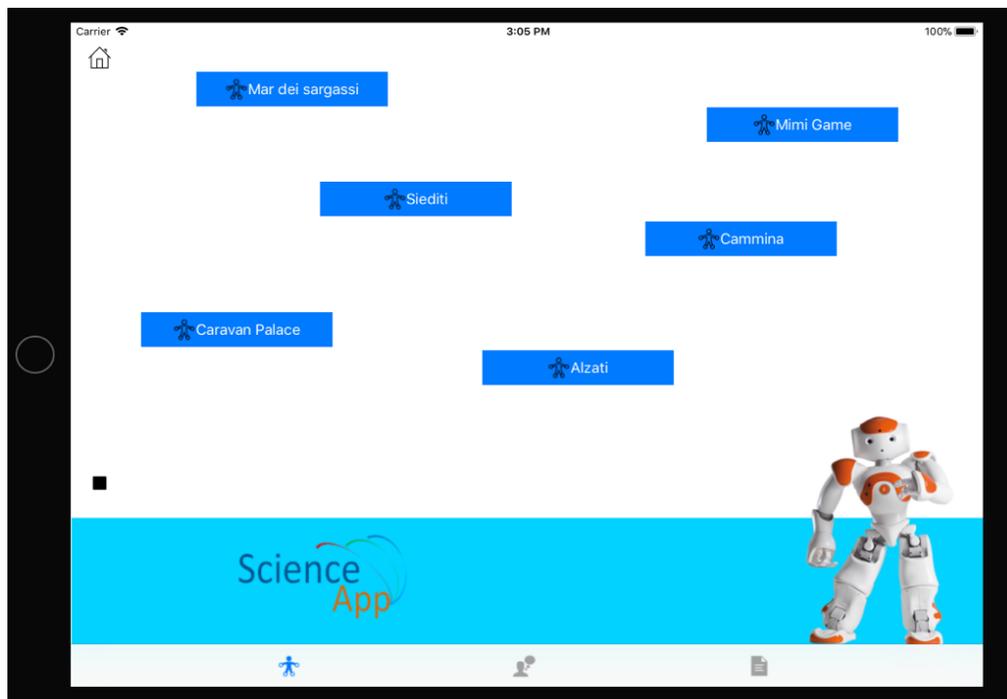


Figura 18

Le istruzioni al server vengono gestite tramite chiamate *REST*, ove per *REST* possiamo intendere un insieme di principi architetturali che hanno nelle risorse l'elemento fondamentale per i *web server RESTful*. Nel caso specifico della mia applicazione le risorse sono rappresentate dai *Behaviours*.

La raccolta dati viene gestita sia tramite l'app che tramite il robot. L'applicativo memorizza in file testuali le risposte fornite nel *task* mentre il robot tramite i propri microfoni e le proprie telecamere memorizza filmati in formato *3gp* e file audio in formato *wav*, successivamente uniti da me in *postprocessing* in formato *Mov*

Conclusioni

Il lavoro di tesi ha previsto sia la progettazione di un'architettura robotica in grado di sostenere l'interazione uomo-macchina, sia l'implementazione di un *framework* software con relative librerie il cui algoritmo di apprendimento è basato su una campagna sperimentale sul campo.

Il framework è interoperabile, integra librerie allo stato dell'arte, i diversi moduli della suddetta architettura ed i relativi algoritmi di apprendimento.

Il controllo dell'architettura si basa sul monitoraggio del comportamento dell'interlocutore e della sua incidenza. Si tratta di una caratterizzazione dell'interlocutore basata su due pilastri, come si comporta l'interlocutore e quanto è affettiva la relazione con lui.

L'approccio ha previsto la misura automatica delle modalità con cui l'interlocutore esegue un'azione in relazione alla sua interazione con gli oggetti e la scena e le altre persone presenti. Per valutare quando sia affettiva la relazione, il sistema stima la situazione emotiva dell'interlocutore sfruttando il canale visuale, in particolare l'espressione facciale. Partendo da librerie di *pattern recognition* sviluppate nei laboratori CNR Isasi, l'architettura consente di rilevare il volto e *keypoints* facciali, da questi la posa della testa e la stima del punto di mira visivo, che a loro volta vengono utilizzati per calcolare l'intensità delle espressioni facciali su una scala di azioni tipo denominate *Facial Action Unit*. Considerando che l'affidabilità di un classificatore dipende dai dati di *training*, è stato sviluppato un sistema di acquisizione di un database di espressioni di interlocutori composte da bambini, terapisti, genitori e medici.

L'interpretazione delle variabili di valutazioni si basa sul sistema di supporto alle decisioni di tipo gerarchico. Per risolvere il problema di come fare agire il robot sulla base di un processo emozionale - problema attualmente aperto nel campo della robotica - è stata sviluppata una memoria 'affettiva' ed un sistema di supporto alle decisioni gerarchico, che consentono di selezionare l'azione da eseguire utilizzando dati di *training* consistenti

nella strategia abilitativa cognitivo - comportamentale usata maggiormente dai terapisti nel corso della terapia con minori affetti da disturbi del neurosviluppo. La componente di memoria affettiva memorizza e richiama le informazioni che vengono dai vari componenti, nonché dal protocollo di trattamento, memorizzando al contempo le esperienze del robot

Appendice Tecnica

PyServer.py

```
1. from BaseHTTPServer import BaseHTTPRequestHandler,HTTPServer
2. from SocketServer import ThreadingMixIn
3. import urllib
4. import json
5. from urlparse import urlparse,parse_qs
6. from naoqi import ALProxy
7. import threading
8. import logging
9.
10.
11.
12. class MyClass(GeneratedClass):
13.     def __init__(self):
14.         GeneratedClass.__init__(self)
15.         self.logger.debug("Server start on port 3001")
16.         self.tts = ALProxy('ALTextToSpeech')
17.         self.tts.post.say("Sono in attesa")
18.         self.server = ThreadingServer(("iMac-di-
Lucio.local", 51342), NaoHandler)
19.         self.server_thread = threading.Thread(target=self.server.serve_forever)
20.
21.         self.server_thread.daemon = True
22.         self.server_thread.start()
23.
24.     def onLoad(self):
25.         #self.memory = None
26.         pass
27.
28.     def onUnload(self):
29.         self.server.shutdown()
30.         self.server.server_close()
31.         pass
32.
33.     def onInput_onStart(self):
34.         #self.onStopped() #activate the output of the box
35.         #self.recording()
36.         pass
37.
38.     def onInput_onStop(self):
39.         self.server.shutdown()
40.         self.server.server_close()
41.
42.
```

```

43. class NaoHandler(BaseHTTPRequestHandler):
44.
45.     def do_GET(self):
46.         """Responde to a GET request."""
47.         self.protocol_version = 'HTTP/1.1'
48.         self.bm = ALProxy("ALBehaviorManager")
49.         self.tts = ALProxy('ALTextToSpeech')
50.         self.vr = ALProxy('ALVideoRecorder')
51.         o = urlparse(self.path)
52.         qs = parse_qs(o.query)
53.         type = o.path.strip('/')
54.         if type == 'behavior':
55.             self.bm.runBehavior(qs['name'][0])
56.         elif type == 'stop' :
57.             self.bm.stopBehavior(qs['name'][0])
58.         elif type == 'say' :
59.             self.tts.post.say(qs['text'][0])
60.         elif type == 'stopSay':
61.             self.tts.post.stopAll()
62.         elif type=='motion' :
63.             str = qs['movement'][0]
64.             #motion(qs['movement'][0],float(qs['x'][0]),float(qs['y'][0]))
65.             self.mot = ALProxy("ALMotion")
66.             self.posture = ALProxy("ALRobotPosture")
67.             if str=='toward' :
68.                 self.posture.goToPosture("StandInit", 0.5)
69.                 self.mot.moveToward(float(qs['x'][0]),float(qs['y'][0]), float(
qs['w'][0]), [{"Frequency", 0.5}])
70.             elif str=='rotate' :
71.                 x = 0.2;
72.                 y = 0.2;
73.                 theta = 1.5709;
74.                 self.mot.moveTo(x, y, theta);
75.             self.send_response(200)
76.             self.send_header('Content-Type', "application/json")
77.             self.end_headers()
78.             self.wfile.write(str)
79.             return
80.
81.     def motion(movement,x,y):
82.         mot = ALProxy("ALMotion")
83.         if movement=='toward' :
84.             mot.moveToward(x,y, 0.0, [{"Frequency", 1.0}])
85.         #if movement=='rotate' :
86.             #float(qs['x'][0]),float(qs['y'][0])
87.
88.
89. class ThreadingServer(ThreadingMixIn, HTTPServer):
90.     allow_reuse_address = True
91.
92.     def shutdown(self):
93.         self.socket.close()
94.         HTTPServer.shutdown(self)

```

Client.swift

```
1. //
2. // client.swift
3. // T-Nao
4. //
5. // Created by Lucio Busà on 11/10/2016.
6. // Copyright © 2016 Lucio Busà. All rights reserved.
7. //
8.
9. import Foundation
10. import Alamofire
11.
12. class Client {
13.
14.     var behaviour: String = ""
15.     var clicked: Bool! = false
16.
17.     func connect(_ ip_address:String, sender:UIButton!,calling_class: UIViewCon
18.         troller)
19.     {
20.         let parameters: Parameters = ["text": "connesso"]
21.         Alamofire.request("http://" + ip_address + ":3001/say", method: .get, paramet
22.             ers: parameters).responseData { response in
23.                 switch response.result {
24.                 case .success( _):
25.                     calling_class.performSegue(withIdentifier: "ipaddress", sender:
26.                         sender)
27.                     print(response.result)
28.                 case .failure( _):
29.                     let alertController = UIAlertController(title: "", message: "Con
30.                         nection refused", preferredStyle: UIAlertControllerStyle.alert)
31.                     alertController.addAction(UIAlertAction(title: "Ok", style: UIA
32.                         lertActionStyle.default, handler: nil))
33.                     calling_class.present(alertController, animated: true, completi
34.                         on: nil)
35.                 }
36.             }
37.     }
38.
39.     func callSay(_ parameters: [String : AnyObject], ip_address:String, calling
40.         _class: UIViewController){
41.
42.         Alamofire.request("http://" + ip_address + ":3001/say", method: .get, paramet
43.             ers: parameters).responseData { response in
44.                 switch response.result {
45.                 case .success( _):
46.                     print("SUCCESS")
47.                 case .failure( _):
48.                     let alertController = UIAlertController(title: "", message: "Rip
49.                         rova", preferredStyle: UIAlertControllerStyle.alert)
50.                     alertController.addAction(UIAlertAction(title: "Ok", style: UIA
51.                         lertActionStyle.default))
```

```

43.         calling_class.present(alertController, animated: true, completi
on: nil)
44.     }
45. }
46. }
47.
48. func callBehaviour(_ behaviour: String,parameters: [String : AnyObject], ip
_address:String, calling_class: UIViewController){
49.     if(self.behaviour != "recording") {
50.         self.stopAction(ip_address, calling_class: calling_class)
51.     }
52.     self.behaviour = behaviour
53.     Alamofire.request("http://" + ip_address + ":3001/behavior", method: .get, p
arameters: parameters).responseData { response in
54.         NSLog(response.debugDescription)
55.         switch response.result {
56.         case .success( _):
57.             print("ok")
58.         case .failure( _):
59.             let alertController = UIAlertController(title: "", message: "Rip
rova", preferredStyle: UIAlertControllerStyle.alert)
60.             alertController.addAction(UIAlertAction(title: "Ok", style: UIA
lertActionStyle.default))
61.             calling_class.present(alertController, animated: true, completi
on: nil)
62.         }
63.     }
64. }
65.
66. func stopAction(_ ip_address:String, calling_class: UIViewController) {
67.     Alamofire.request("http://" + ip_address + ":3001/stop", method: .get, parame
ters: ["name": self.behaviour]).responseData { response in
68.         NSLog(response.debugDescription)
69.         switch response.result {
70.         case .success( _):
71.             print("ok")
72.         case .failure( _):
73.             let alertController = UIAlertController(title: "", message: "Los
e Connection", preferredStyle: UIAlertControllerStyle.alert)
74.             alertController.addAction(UIAlertAction(title: "Ok", style: UIA
lertActionStyle.default){
75.                 action -
76.                 > Void in calling_class.dismiss(animated: true, completion: nil)
77.             })
78.             calling_class.present(alertController, animated: true, completi
on: nil)
79.         }
80.     }
81.
82. func callRecording(_ parameters: [String : AnyObject], ip_address:String, c
alling_class: UIViewController){
83.     Alamofire.request("http://" + ip_address + ":3001/recording", method: .get, p
arameters: parameters).responseData { response in
84.         NSLog(response.debugDescription)
85.         switch response.result {
86.         case .success( _):
87.             print("ok")
88.         case .failure( _):

```

```

89.         let alertController = UIAlertController(title: "", message:"Rip
rova", preferredStyle: UIAlertControllerStyle.alert)
90.         alertController.addAction(UIAlertAction(title: "Ok", style: UIA
lertActionStyle.default))
91.         calling_class.present(alertController, animated: true, completi
on: nil)
92.     }
93. }
94. }
95.
96. func record(ip_address:String, recording: Bool,calling_class:UIViewControll
er) {
97.     if recording == false {
98.         self.clicked = false
99.         Alamofire.request("http://" + ip_address + ":3001/stop", method:
.get, parameters: ["name": "recording"]).responseData { response in
100.             NSLog(response.debugDescription)
101.             switch response.result {
102.                 case .success( _):
103.                     print("ok")
104.                 case .failure( _):
105.                     let alertController = UIAlertController(title: "",
message:"Riprova", preferredStyle: UIAlertControllerStyle.alert)
106.                     alertController.addAction(UIAlertAction(title: "Ok"
, style: UIAlertActionStyle.default))
107.                     calling_class.present(alertController, animated: tr
ue, completion: nil)
108.                 }
109.             }
110.         }
111.     }
112.     else {
113.         self.clicked = true
114.         self.callBehaviour("recording", parameters: ["name": "recordi
ng" as AnyObject], ip_address: ip_address, calling_class: calling_class)
115.     }
116. }
117.
118.
119.
120. }

```

NaoServiceBrowser.swift

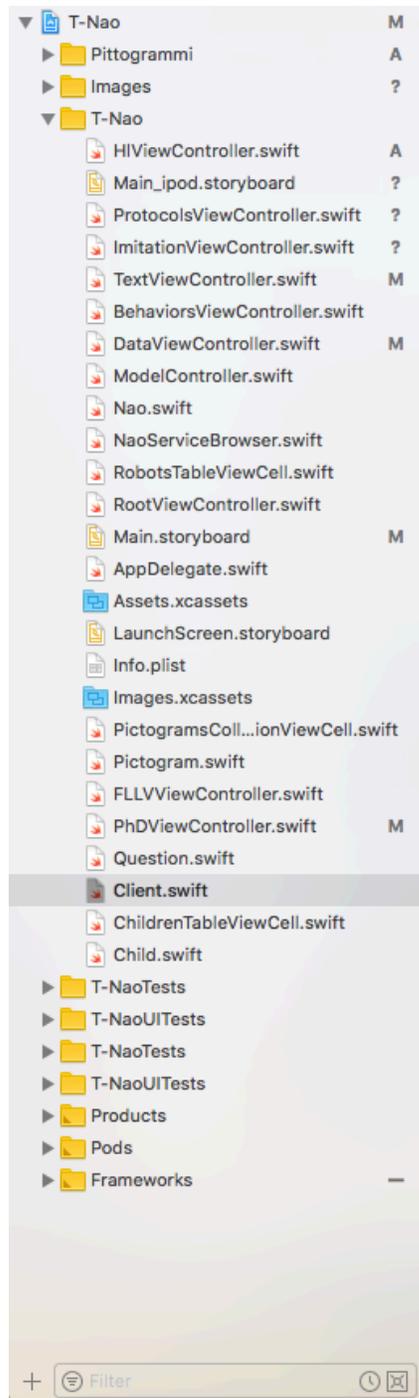
```
1. //
2. // NaoServiceBrowser.swift
3. // playNao
4. //
5. // Created by Lucio Busà on 05/09/16.
6. // Copyright © 2016 Lucio Busà. All rights reserved.
7. //
8.
9. import Foundation
10. import CocoaAsyncSocket
11.
12. extension NSData {
13.     func castToCPointer<T>() -> T {
14.         let mem = UnsafeMutablePointer<T>.allocate(capacity: MemoryLayout<T.Type>.size)
15.         self.getBytes(mem, length: MemoryLayout<T.Type>.size)
16.         return mem.move()
17.     }
18. }
19.
20.
21. class NaoServiceBrowser: NSObject, NetServiceDelegate, NetServiceBrowserDelegate, GCDAsyncSocketDelegate {
22.     var socket: GCDAsyncSocket!
23.     var services: NSMutableArray!
24.     var serviceBrowser: NetServiceBrowser!
25.     var service: NetService!
26.     var naoList = [Nao]()
27.
28.     func startBrowsing(){
29.         if (services != nil) {
30.             services.removeAllObjects()
31.         } else {
32.             services = NSMutableArray()
33.         }
34.         self.serviceBrowser = NetServiceBrowser()
35.         self.serviceBrowser.delegate = self
36.         self.serviceBrowser.searchForServices(ofType: "_naoqi._tcp", inDomain:
37.         "")
38.     }
39.     func netServiceBrowser(_ browser: NetServiceBrowser, didFind service: NetService, moreComing: Bool) {
40.         services.add(service)
41.         service.delegate = self
42.         service.resolve(withTimeout: 30.0)
43.     }
44.
45.     func netServiceBrowser(_ browser: NetServiceBrowser, didRemove service: NetService, moreComing: Bool) {
46.         services.remove(service)
47.         refreshService()
48.     }
49. }
```

```

49.
50. func refreshService()
51. {
52.     naoList.removeAll()
53.     for object in services {
54.         if let sender = object as? NetService {
55.             resolveAddress(sender)
56.         }
57.     }
58. }
59.
60. func resolveAddress(_ sender: NetService){
61.     let data = sender.addresses!.first! as NSData
62.
63.     let inetAddress: sockaddr_in = data.castToCPointer()
64.     if inetAddress.sin_family == __uint8_t(AF_INET) {
65.         if let ip = String(cString: inet_ntoa(inetAddress.sin_addr), encoding:
        .ascii) {
66.             // IPv4
67.             naoList.append(Nao(name: sender.name, ipAddress: ip))
68.         }
69.     } else if inetAddress.sin_family == __uint8_t(AF_INET6) {
70.         let inetAddress6: sockaddr_in6 = data.castToCPointer()
71.         let ipStringBuffer = UnsafeMutablePointer<Int8>.allocate(capacity:
        Int(INET6_ADDRSTRLEN))
72.         var addr = inetAddress6.sin6_addr
73.
74.         if let ipString = inet_ntop(Int32(inetAddress6.sin6_family), &addr,
        ipStringBuffer, __uint32_t(INET6_ADDRSTRLEN)) {
75.             if let ip = String(cString: ipString, encoding: .ascii) {
76.                 // IPv6
77.                 naoList.append(Nao(name: sender.name, ipAddress: ip))
78.             }
79.         }
80.
81.         ipStringBuffer.deallocate(capacity: Int(INET6_ADDRSTRLEN))
82.     }
83. }
84. }
85.
86. func netServiceDidResolveAddress(_ sender: NetService) {
87.     resolveAddress(sender)
88. }
89.
90. func netService(_ sender: NetService, didNotResolve errorDict: [String : NS
        Number]) {
91.     service.delegate = nil
92. }
93.
94.
95. }

```

Struttura delle classi



Bibliografia

- Ackermann, E. (1996). Perspective-taking and object construction: Two keys to learning. *Constructionism in practice: designing, thinking, and learning in a digital world*, Lawrence Erlbaum, Mahwah, NJ, 25-35.
- Adam, C. (2007). Emotions: from psychological theories to logical formalization and implementation in a BDI agent (Doctoral dissertation).
- Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Holt, Rinehart & Winston.
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis.
- Amighi, J. K. (1999). *The meaning of movement: Developmental and clinical perspectives of the Kestenberg Movement Profile*. Taylor & Francis.
- Anderson, J. R. (2013). *Language, memory, and thought*. Psychology Press.
- Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, *111*(4), 1036.
- Anderson, K., & McOwan, P. W. (2006). A real-time automated system for the recognition of human facial expressions. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *36*(1), 96-105.
- Andrade, G., Ramalho, G., Santana, H., & Corruble, V. (2005, August). Extending reinforcement learning to provide dynamic game balancing. In *Proceedings of the Workshop on Reasoning, Representation, and Learning in Computer Games, 19th International Joint Conference on Artificial Intelligence (IJCAI)* (pp. 7-12).
- Argyle, M., Lalljee, M., & Cook, M. (1968). The effects of visibility on interaction in a dyad. *Human relations*, *21*(1), 3-17.
- Aronfreed, J. (1969). The problem of imitation. *Advances in child development and behavior*, *4*, 209-319.
- Arnold, A., Semple, R. J., Beale, I., & Fletcher-Flinn, C. M. (2000). Eye contact in children's social interactions: What is normal behaviour?. *Journal of Intellectual and Developmental Disability*, *25*(3), 207-216.
- Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. *Psychology of learning and motivation*, *2*, 89-195.
- Baddeley, A. D., & Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, *8*, 47-89.
- Baddeley, A., Cocchini, G., Della Sala, S., Logie, R. H., & Spinnler, H. (1999). Working memory and vigilance: Evidence from normal aging and Alzheimer's disease. *Brain and cognition*, *41*(1), 87-108.

- Baddeley, A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, 4(11), 417-423.
- Baddeley, A. D., Baddeley, H. A., Bucks, R. S., & Wilcock, G. K. (2001). Attentional control in Alzheimer's disease. *Brain*, 124(8), 1492-1508.
- Baltrušaitis, T., Robinson, P., & Morency, L. P. (2016, March). Openface: an open source facial behavior analysis toolkit. In *Applications of Computer Vision (WACV), 2016 IEEE Winter Conference on* (pp. 1-10). IEEE.
- Bandura, A. (1971). Psychological modelling conflicting theories. Chicago: Aldine
- Bar, M. (2007). The proactive brain: using analogies and associations to generate predictions. *Trends in cognitive sciences*, 11(7), 280-289.
- Bar, M., Aminoff, E., & Ishai, A. (2007). Famous faces activate contextual associations in the parahippocampal cortex. *Cerebral Cortex*, 18(6), 1233-1238.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2005, June). Recognizing facial expression: machine learning and application to spontaneous behavior. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on* (Vol. 2, pp. 568-573). IEEE.
- Bartlett, M. S., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., & Movellan, J. (2006, April). Fully automatic facial action recognition in spontaneous behavior. In *Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on* (pp. 223-230). IEEE.
- Baxter, P., & Browne, W. (2009, September). Memory-based cognitive framework: a low-level association approach to cognitive architectures. In *European Conference on Artificial Life* (pp. 402-409). Springer, Berlin, Heidelberg.
- Baxter, G., & Sommerville, I. (2011). Socio-technical systems: From design methods to systems engineering. *Interacting with computers*, 23(1), 4-17.
- Baxter, P., Wood, R., Morse, A. F., & Belpaeme, T. (2011, November). Memory-Centred Architectures: Perspectives on Human-Level Cognitive Competencies. In *AAAI Fall Symposium: Advances in Cognitive Systems*.
- Beattie, G. W. (1981). A further investigation of the cognitive interference hypothesis of gaze patterns during conversation. *British Journal of Social Psychology*, 20(4), 243-248.
- Becker, C., Kopp, S., & Wachsmuth, I. (2007). Why emotions should be integrated into conversational agents. *Conversational informatics: an engineering approach*, 49-68.
- Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism*, 18(5), 598-608.
- Belpaeme, T., Baxter, P. E., Read, R., Wood, R., Cuayáhuitl, H., Kiefer, B., ... & Looije, R. (2012). Multimodal child-robot interaction: Building social bonds. *Journal of Human-Robot Interaction*, 1(2), 33-53.

- Berthouze, L., Shigematsu, Y., & Kuniyoshi, Y. (1998). Dynamic categorization of explorative behaviors for emergence of stable sensorimotor configuration. In Pfeifer, R., Blumberg, B., Meyer, J., & Wilson, S., (Eds.), *Proceeding of the Fifth International Conference on Simulation of Adaptive Behaviour*, (pp. 67–72). Zurich: ETHZ
- Betella, A., Zucca, R., Cetnarski, R., Greco, A., Lanatà, A., Mazzei, D., ... & Verschure, P. F. (2014). Inference of human affective states from psychophysiological measurements extracted under ecologically valid conditions. *Frontiers in neuroscience*, 8.
- Bickel, W. K., Yi, R., Landes, R. D., Hill, P. F., & Baxter, C. (2011). Remember the future: working memory training decreases delay discounting among stimulant addicts. *Biological psychiatry*, 69(3), 260-265.
- Bickhard, M. H. (2009). The interactivist model. *Synthese*, 166(3), 547-591.
- Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 12(2), 293-327.
- Billeci, L., Tonacci, A., Tartarisco, G., Narzisi, A., Di Palma, S., Corda, D., ... & Pioggia, G. (2016). An Integrated Approach for the Monitoring of Brain and Autonomic Response of Children with Autism Spectrum Disorders during Treatment by Wearable Technologies. *Frontiers in neuroscience*, 10.
- Bird, G., Leighton, J., Press, C., & Heyes, C. (2007). Intact automatic imitation of human and robot actions in autism spectrum disorders. *Proceedings of the Royal Society of London B: Biological Sciences*, 274(1628), 3027-3031.
- Birdwhistell, R. L. (2010). *Kinesics and context: Essays on body motion communication*. University of Pennsylvania press.
- Blake, A., & Yuille, A. (1992). *Active vision* (No. 006.37 B5).
- Blanc, R., Gomot, M., Gattegno, M. P., Barthélémy, C., & Adrien, J. L. (2002). Les troubles de l'activité symbolique chez des enfants autistes, dysphasiques et retardés mentaux et l'effet de l'étayage de l'adulte. *Revue Québécoise de Psychologie*.
- Bluethmann, W., Ambrose, R., Diftler, M., Askew, S., Huber, E., Goza, M., ... & Magruder, D. (2003). Robonaut: A robot designed to work with humans in space. *Autonomous robots*, 14(2), 179-197.
- Byrne, R.W., & Russon, A.E. (1998). Learning by imitation: A hierarchical approach. *DEST "nad-r8"> Behavioral and Brain Sciences*, 21, 667–721.
- Branigan, H. P., Pickering, M. J., Pearson, J., & McLean, J. F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics*, 42(9), 2355-2368.
- Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, 59(1), 119-155.
- Bull, L. (Ed.). (2004). *Applications of learning classifier systems* (Vol. 150). Springer Science & Business Media.

- Bull, L., & Kovacs, T. (2005). Foundations of learning classifier systems: An introduction. *Foundations of Learning Classifier Systems*, 913-913.
- Burgess, N., Donnett, J. G., Jeffery, K. J., & John, O. (1997). Robotic and neuronal simulation of the hippocampus and rat navigation. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 352(1360), 1535-1543.
- Cao, Z., Simon, T., Wei, S. E., & Sheikh, Y. (2016). Realtime multi-person 2d pose estimation using part affinity fields. *arXiv preprint arXiv:1611.08050*.
- Carbone, V. J., O'Brien, L., Sweeney-Kerwin, E. J., & Albert, K. M. (2013). Teaching eye contact to children with autism: A conceptual analysis and single case study. *Education and Treatment of Children*, 36(2), 139-159.
- Chang, Y., Hu, C., Feris, R., & Turk, M. (2006). Manifold based analysis of facial expression. *Image and Vision Computing*, 24(6), 605-614.
- Chang, Y., Hu, C., & Turk, M. (2004, June). Probabilistic expression analysis on manifolds. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on* (Vol. 2, pp. II-II). IEEE.
- Chen, L., Wei, H., & Ferryman, J. (2013). A survey of human motion analysis using depth imagery. *Pattern Recognition Letters*, 34(15), 1995-2006.
- Chen, C., Heili, A., & Odobez, J. M. (2011, November). A joint estimation of head and body orientation cues in surveillance video. In *Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on* (pp. 860-867). IEEE.
- Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: a meta-analytic review. *Psychological bulletin*, 116(3), 457.
- Costella, J. P. (1995). A Beginner's Guide to the Human Field of View. *School of Physics, The University of Melbourne*.
- Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological bulletin*, 104(2), 163.
- Cowan, N. (1999). An embedded-processes model of working memory. *Models of working memory: Mechanisms of active maintenance and executive control*, 20, 506.
- Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior*, 11(6), 671-684.
- Dang, T. H. H., Letellier-Zarshenas, S., & Duhaut, D. (2008, September). GRACE-Generic robotic architecture to create emotions. In *11th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines-CLAWAR 2008*.
- Dautenhahn K (1999) Robots as social actors: Aurora and the case of autism. Proc. CT99, The Third International Cognitive Technology Conference, August, San-Francisco, p 359-374

- Dautenhahn K, Werry I (2000) Issues of robot-human interaction dynamics in the rehabilitation of children with autism. Proc. From Animals to Animats, The Sixth International Conference on the Simulation of Adaptive Behavior (SAB2000). 11 - 15 September 2000. Paris, France, p 519-528
- Dautenhahn, K. (2003). Roles and functions of robots in human society: implications from research in autism therapy. *Robotica*, 21(4), 443-452.
- Dautenhahn K, Werry I (2004) Towards interactive robots in autism therapy: Background, motivation and challenges. *Pragmatics and Cognition* 12(1):1-35
- De Jaegher, H., Di Paolo, E., & Gallagher, S. (2010). Can social interaction constitute social cognition?. *Trends in cognitive sciences*, 14(10), 441-447.
- Decety, J., & Grezes, J. (1999). Neural mechanisms subserving the perception of human actions. *Trends in Cognitive Sciences*, 3, 172–178.
- de Sá, A. G., & Rix, J. (2000). Virtual prototyping: The Integration of Design and Virtual Reality. In *CAD Tools and Algorithms for Product Design* (pp. 128-150). Springer Berlin Heidelberg.
- Derbinsky, N., & Laird, J. E. (2010, March). Extending soar with dissociated symbolic memories. In *Symposium on Human Memory for Artificial Agents, AISB* (pp. 31-37).
- Deshmukh, A., Aylett, R., Kriegel, M., & Vargas, P. A. (2012). Multiple embodiments for robots in heritage applications. In *First international conference robotic innovation for cultural heritage 2012*.
- Dias, J., Mascarenhas, S., & Paiva, A. (2014). Fatima modular: Towards an agent architecture with a generic appraisal framework. In *Emotion Modeling* (pp. 44-56). Springer International Publishing.
- Duda, R. O., & Shortliffe, E. H. (1983). Expert systems research. *Science*, 220(4594), 261-268.
- Ekman, P., & Oster, H. (1979). Facial expressions of emotion. *Annual review of psychology*, 30(1), 527-554.
- Ekman, P., Friesen, W. V., & Ellsworth, P. (2013). *Emotion in the human face: Guidelines for research and an integration of findings*. Elsevier.
- Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., & Rasmussen, D. (2012). A large-scale model of the functioning brain. *science*, 338(6111), 1202-1205.
- Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, 73, 2608–2611.
- Fagel, S., Moussa, M. B., & Cereghetti, D. (2016, May). How avatars in care context should show affect. In *Proceedings of the 10th EAI International Conference on Pervasive Computing Technologies for Healthcare* (pp. 262-267). ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering).
- Fasola, J., & Matarić, M. J. (2010, August). Robot motivator: Increasing user enjoyment and performance on a physical/cognitive task. In *Development*

- and Learning (ICDL), 2010 IEEE 9th International Conference on (pp. 274-279). IEEE.
- Faust, J., Simon, C., & Smart, W. D. (2006, October). A video game-based mobile robot simulation environment. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 3749-3754). IEEE.
- Ferrara C, Hill SD (1980) The responsiveness of autistic children to the predictability of social and non-social toys. *Autism and Developmental Disorders* 10(1):51-57.
- Foxx, R. M. (1977). Attention training: The use of overcorrection avoidance to increase the eye contact of autistic and retarded children. *Journal of Applied Behavior Analysis*, 10(3), 489-499.
- Fuster, J. M. (1997). Network memory. *Trends in neurosciences*, 20(10), 451-459.
- Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, 119(2), 593-609.
- Georgieff, N., & Jeannerod, M. (1998). Beyond consciousness of external reality: A who system for consciousness of action and self-consciousness. *Consciousness and Cognition*, 7, 465-477
- Glenberg, A. M. (1997). What memory is for. *Behavioral and brain sciences*, 20(1), 1-19.
- Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a minimal representation of affective gestures. *IEEE Transactions on Affective Computing*, 2(2), 106-118.
- Golosio, B., Cangelosi, A., Gamotina, O., & Masala, G. L. (2015). A cognitive neural architecture able to learn and communicate through natural language. *PloS one*, 10(11), e0140866.
- Goodrich, M. A., & Schultz, A. C. (2007). Human-robot interaction: a survey. *Foundations and trends in human-computer interaction*, 1(3), 203-275.
- Goertzel, B., Lian, R., Arel, I., De Garis, H., & Chen, S. (2010). A world survey of artificial brain projects, Part II: Biologically inspired cognitive architectures. *Neurocomputing*, 74(1), 30-49.
- Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, memory, and cognition*, 11(3), 501.
- Greer, D. R., & Ross, D. E. (2007). Verbal behavior analysis. New York, NY: Pearson Education.
- Guo, G., & Dyer, C. R. (2005). Learning from examples in the small sample case: face expression recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 35(3), 477-488.
- Gupta, A., & Davis, L. S. (2007, June). Objects in action: An approach for combining action understanding and object perception. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on* (pp. 1-8). IEEE.

- Gupta, A., Kembhavi, A., & Davis, L. S. (2009). Observing human-object interactions: Using spatial and functional compatibility for recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10), 1775-1789.
- Haith, M. M. (1980). Rules that babies look by: The organization of newborn visual activity. Lawrence Erlbaum Associates.
- Han, D., Bo, L., & Sminchisescu, C. (2009, September). Selection and context for action recognition. In *Computer Vision, 2009 IEEE 12th International Conference on* (pp. 1933-1940). IEEE.
- Hayes-Roth, B. (1995). An architecture for adaptive intelligent systems. *Artificial Intelligence*, 72(1-2), 329-365.
- Hebb, D. O. (1949). The organization of behavior; a neuropsychological theory.
- Heimann, M. (1989). Neonatal imitation, gaze aversion, and mother infant interaction. *Infant Behavior and Development*, 12, 495-505.
- Helgeson, D. C., Fantuzzo, J. W., Smith, C., & Barr, D. (1989). Eye-contact skill training for adolescents with developmental disabilities and severe behavior problems. *Education and Training in Mental Retardation*, 56-62.
- Hershkowitz V (2000) Computer based therapy for individuals with autism. *Advance Magazine*, January 10.
- Heyes, C. (2001). Causes and consequences of imitation. *Trends in Cognitive Sciences*, 6, 253-261.
- Higashinaka, R., Dohsaka, K., & Isozaki, H. (2008, December). Effects of self-disclosure and empathy in human-computer dialogue. In *Spoken Language Technology Workshop, 2008. SLT 2008. IEEE*(pp. 109-112). IEEE.
- Holland, L. Z., Carvalho, J. E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S. M., & Yu, J. K. (2013). Evolution of bilaterian central nervous systems: a single origin?. *EvoDevo*, 4(1), 27.
- Howlin P, Baron-Cohen S, Hadwin J (1999) Teaching Children with Autism to Mind-read. New York, John Wiley and Sons
- Hsu, S. W., & Li, T. Y. (2006, October). Third-Person Interactive Control of Humanoid with Real-Time Motion Planning Algorithm. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 4845-4850). IEEE.
- Hudlicka, E. (2008). Affective computing for game design. In *Proceedings of the 4th Intl. North American Conference on Intelligent Games and Simulation* (pp. 5-12).
- Iacoboni, M., Woods, R., Brass, M., Bekkering, H., Mazziotta, J.C., & Rizzolatti, G. (1999). DEST "nad-r19"> Cortical mechanisms of human imitation. *Science*, 286, 2526-2528
- Janssen, J. B., van der Wal¹², C. C., & Neerincx¹³, M. Motivate to learn: Effects of performance adaptation on child motivation of robot interaction.
- Johnsson, M., & Balkenius, C. (2009). Experiments with self-organizing systems for texture and hardness perception. *Robotics and Autonomous Systems*, 4, 53-62.

- Jones, E. A., & Carr, E. G. (2004). Joint attention in children with autism: Theory and intervention. *Focus on Autism & Other Developmental Disabilities*, 19, 13–26.
- Kasap, Z., Moussa, M. B., Chaudhuri, P., & Magnenat-Thalmann, N. (2009). Making them remember—Emotional virtual characters with memory. *IEEE Computer Graphics and Applications*, 29(2), 20-29.
- Kasderidis, S., & Taylor, J. G. (2008). The GNOSYS Cognitive Architecture. In *International Conference on Cognitive Systems*.
- Klin, A., Lang, J., Cicchetti, D. V., & Volkmar, F. R. (2000). Brief report: Interrater reliability of clinical diagnosis and DSM-IV criteria for autistic disorder: Results of the DSM-IV autism field trial. *Journal of autism and Developmental disorders*, 30(2), 163-167.
- Kohonen, T., Schroeder, M. R., Huang, T. S., & Maps, S. O. (2001). Springer-Verlag New York. Inc., Secaucus, NJ, 43, 2.
- Krichmar, J. L., & Edelman, G. M. (2002). Machine psychology: autonomous behavior, perceptual categorization and conditioning in a brain-based device. *Cerebral Cortex*, 12(8), 818-830.
- Krichmar, J. L., Nitz, D. A., Gally, J. A., & Edelman, G. M. (2005). Characterizing functional hippocampal pathways in a brain-based device as it solves a spatial memory task. *Proceedings of the National Academy of Sciences of the United States of America*, 102(6), 2111-2116.
- Kugiumutzakis, J. (1999). Genesis and development of early infant mimesis to facial and vocal models. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 36–59). Cambridge, MA: Cambridge University Press.
- Kuniyoshi, Y. (1994). The science of imitation — towards physically and socially grounded intelligence. Special Issue, RWC Technical Report, TR-94001, Real World Computing Project Joint Symposium, Tsukuba-shi, Ibaraki-ken
- Laird, J. E. (2001, May). It knows what you're going to do: adding anticipation to a Quakebot. In *Proceedings of the fifth international conference on Autonomous agents* (pp. 385-392). ACM.
- Lang, F. (2007). Mechanisms and significance of cell volume regulation. *Journal of the American college of nutrition*, 26(sup5), 613S-623S.
- Langley, P., Laird, J. E., & Rogers, S. (2009). Cognitive architectures: Research issues and challenges. *Cognitive Systems Research*, 10(2), 141-160.
- Laurenceau, J. P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: the importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of personality and social psychology*, 74(5), 1238.
- Lee, M. H., Meng, Q., & Chao, F. (2007). Staged competence learning in developmental robotics. *Adaptive Behavior*, 15(3), 241-255.
- Lee, P. U., & Tversky, B. (2001, January). Costs of switching perspectives in route and survey descriptions. In *Proceedings of the Cognitive Science Society* (Vol. 23, No. 23).
- Lemay, M. (2004). *L'autisme aujourd'hui*. Paris: Éd. Odile Jacob.

- Logie, R., Baddeley, A., Mané, A., Donchin, E., & Sheptak, R. (1989). Working memory in the acquisition of complex cognitive skills. *Acta psychologica*, 71(1), 53-87.
- Lucey, S., Ashraf, A. B., & Cohn, J. F. (2007). Investigating spontaneous facial action recognition through aam representations of the face. In *Face recognition*. InTech.
- Luciw, M., & Weng, J. (2010). Top-Down Connections in Self-Organizing Hebbian Networks: Topographic Class Grouping. *IEEE Transactions on Autonomous Mental Development*, 2(3), 248-261.
- Maratos, O. (1973). The origin and development of imitation in the first six months of life. Paper presented at the British Psychological Society Annual Meeting, Liverpool.
- Marinier, R. P., Laird, J. E., & Lewis, R. L. (2009). A computational unification of cognitive behavior and emotion. *Cognitive Systems Research*, 10(1), 48-69.
- Marsella, S. C., & Gratch, J. (2009). EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1), 70-90.
- Martin, R. M. (2004). *Electronic structure: basic theory and practical methods*. Cambridge university press.
- Mauri, M., Magagnin, V., Cipresso, P., Mainardi, L., Brown, E. N., Cerutti, S., ... & Barbieri, R. (2010, August). Psychophysiological signals associated with affective states. In *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE* (pp. 3563-3566). IEEE.
- Mavadati, S. M., Mahoor, M. H., Bartlett, K., Trinh, P., & Cohn, J. F. (2013). Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4(2), 151-160.
- McKeown, G., Valstar, M. F., Cowie, R., & Pantic, M. (2010, July). The SEMAINE corpus of emotionally coloured character interactions. In *Multimedia and Expo (ICME), 2010 IEEE International Conference on* (pp. 1079-1084). IEEE.
- McKinstry, C., Dale, R., & Spivey, M. J. (2008). Action dynamics reveal parallel competition in decision making. *Psychological Science*, 19(1), 22-24.
- Menchaca-Brandan, M. A., Liu, A. M., Oman, C. M., & Natapoff, A. (2007, March). Influence of perspective-taking and mental rotation abilities in space teleoperation. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on* (pp. 271-278). IEEE.
- Meng, L., Qing, L., Yang, P., Miao, J., Chen, X., & Metaxas, D. N. (2012, November). Activity recognition based on semantic spatial relation. In *Pattern Recognition (ICPR), 2012 21st International Conference on* (pp. 609-612). IEEE.
- Michaud F, Théberge-Turmel (2002) Mobile robotic toys and autism: Observations of interactions. In K. Dautenhahn, A. Bond, Canamero L, Edmonds B (eds): *Socially Intelligent Agents- Creating Relationships with Computers and Robots*. Boston, Dordrecht & London, Kluwer Academic Publishers, p 125-132.

- Mirenda, P. L., Donnellan, A. M., & Yoder, D. E. (1983). Gaze behavior: A new look at an old problem. *Journal of Autism and Developmental Disorders*, 13(4), 397-409.
- Mirza, N. A., Nehaniv, C. L., Dautenhahn, K., & Te Boekhorst, R. (2007). Grounded sensorimotor interaction histories in an information theoretic metric space for robot ontogeny. *Adaptive Behavior*, 15(2), 167-187.
- Mitchell, R. (2002). Pretending and imagination in animals and children. Cambridge: Cambridge University Press.
- Morse, A., & Aktius, M. (2009). Dynamic liquid association: Complex learning without implausible guidance. *Neural Networks*, 22(7), 875-889.
- Morse, A., Belpaeme, T., Cangelosi, A., & Smith, L. (2010, January). Thinking with your body: Modelling spatial biases in categorization using a real humanoid robot. In *Proceedings of the Cognitive Science Society* (Vol. 32, No. 32).
- Mukherjee, S., Biswas, S. K., & Mukherjee, D. P. (2014). Recognizing interactions between human performers by 'Dominating Pose Doublet'. *Machine Vision and Applications*, 25(4), 1033-1052.
- Müller, P., Wonka, P., Haegler, S., Ulmer, A., & Van Gool, L. (2006, July). Procedural modeling of buildings. In *Acm Transactions On Graphics (Tog)* (Vol. 25, No. 3, pp. 614-623). ACM.
- Murray D (1997) Autism and information technology:therapy with computers. In Powell S, Jordan R (eds): Autism and learning: a guide to good practice. London, David Fulton Publishers, p 100-117
- Nadel-Brulfert, J. & Baudonnière, P.M. (1982). The social function of reciprocal imitation in 2-year-old peers. *International Journal of Behavioral Development*, 5, 95-109.
- Nadel I, J., & Decety, J. (2002). Imiter pour découvrir l'humain. Paris: PUF.
- Natale, L., Orabona, F., Metta, G., & Sandini, G. (2005, June). Exploring the world through grasping: a developmental approach. In *Computational Intelligence in Robotics and Automation, 2005. CIRA 2005. Proceedings. 2005 IEEE International Symposium on* (pp. 559-565). IEEE.
- Neath, I., Bireta, T. J., & Surprenant, A. M. (2003). The time-based word length effect and stimulus set specificity. *Psychonomic Bulletin & Review*, 10(2), 430-434.
- Needham, A., & Baillargeon, R. (1997). Object segregation in 8-month-old infants. *Cognition*, 62(2), 121-149.
- Needham, A. (1999). The role of shape in 4-month-old infants' object segregation. *Infant Behavior and Development*, 22(2), 161-178.
- Needham, A., Cantlon, J. F., & Holley, S. M. O. (2006). Infants' use of category knowledge and object attributes when segregating objects at 8.5 months of age. *Cognitive Psychology*, 53(4), 345-360.
- O'Reilly, R. C., Noelle, D. C., Braver, T. S., & Cohen, J. D. (2002). Prefrontal cortex and dynamic categorization tasks: representational organization and neuromodulatory control. *Cerebral cortex*, 12(3), 246-257.

- O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in cognitive sciences*, 6(12), 505-510.
- Oberauer, K. (2002). Access to information in working memory: exploring the focus of attention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3), 411.
- Ortony, A., Clore, G., & Collins, A. (1988). *The Cognitive Structure of Emotions*: Cambridge Uni. Press, New York.
- Ozonoff, S. (1995). Reliability and validity of the Wisconsin Card Sorting Test in studies of autism. *Neuropsychology*, 9(4), 491.
- Pantic, M., & Patras, I. (2006). Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 36(2), 433-449.
- Pantic, M., & Rothkrantz, L. (2004, June). Case-based reasoning for user-profiled recognition of emotions from face images. In *Multimedia and Expo, 2004. ICME'04. 2004 IEEE International Conference on* (Vol. 1, pp. 391-394). IEEE.
- Pawlby, S. (1977). Imitative interaction. In H. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 203–223). London: Academic Press Inc.
- Pickering, M. J., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and brain sciences*, 27(2), 169-190.
- Pierno, A. C., Mari, M., Lusher, D., & Castiello, U. (2008). Robotic movement elicits visuomotor priming in children with autism. *Neuropsychologia*, 46(2), 448-454.
- Pioggia, G., Igliozzi, R., Ferro, M., Ahluwalia, A., Muratori, F., & De Rossi, D. (2005). An android for enhancing social skills and emotion recognition in people with autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 13(4), 507-515.
- Pioggia, G., Igliozzi, R., Sica, M. L., Ferro, M., Muratori, F., Ahluwalia, A., & De Rossi, D. (2008). Exploring emotional and imitational android-based interactions in autistic spectrum disorders. *Journal of CyberTherapy & Rehabilitation*, 1(1), 49-61.
- Plavnick, J. B., & Ferreri, S. J. (2012). Collateral effects of mind training for children with autism. *Research in Autism Spectrum Disorders*, 6, 1366-1376.
- Podrouzek, W., & Furrow, D. (1988). Preschoolers' use of eye contact while speaking: The influence of sex, age, and conversational partner. *Journal of Psycholinguistic Research*, 17(2), 89-98.
- Popescu, C. (2013, October). Wireless intelligent control of a mobile robot used for tracking and monitoring of patients. In *System Theory, Control and Computing (ICSTCC), 2013 17th International Conference* (pp. 704-709). IEEE.

- Poppe, R. (2010). A survey on vision-based human action recognition. *Image and vision computing*, 28(6), 976-990.
- Poucet, B., Lenck-Santini, P. P., Paz-Villagrán, V., & Save, E. (2003). Place cells, neocortex and spatial navigation: a short review. *Journal of Physiology-Paris*, 97(4), 537-546.
- Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007, March). Comparing a computer agent with a humanoid robot. In *Human-Robot Interaction (HRI), 2007 2nd ACM/IEEE International Conference on* (pp. 145-152). IEEE.
- Prescott, T. J., González, F. M. M., Gurney, K., Humphries, M. D., & Redgrave, P. (2006). A robot model of the basal ganglia: behavior and intrinsic processing. *Neural Networks*, 19(1), 31-61.
- Quinton, J. C., Buisson, J. C., & Perotto, F. (2008). Anticipative coordinated cognitive processes for interactivist and piagetian theories. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 171, 287.
- Repovš, G., & Baddeley, A. (2006). The multi-component model of working memory: explorations in experimental cognitive psychology. *Neuroscience*, 139(1), 5-21.
- Ricks, D. J., & Colton, M. B. (2010, May). Trends and considerations in robot-assisted autism therapy. In *Robotics and Automation (ICRA), 2010 IEEE International Conference on* (pp. 4354-4359). IEEE.
- Riegler, A. (2005). Constructive memory. *Kybernetes*, 34(1/2), 89-104.
- Rizzolatti, G., Fadiga, L., Fogassi, L. & Gallese, V. (2002). From mirror neurons to imitation: Facts and speculations, In A. Meltzoff & W. Prinz (Eds.), *The Imitative Mind: Development, Evolution and Brain Bases* (pp. 247–266). Cambridge, MA: Cambridge University Press.
- Robben, S. M. B., Looije, R., Haselager, P., & Neerincx, M. (2011). It's NAO or Never! Facilitate Bonding Between a Child and a Social Robot: Exploring the Possibility of a Robot Adaptive to Personality. *Unpublished master's thesis, Radboud Universiteit Nijmegen*.
- Robins B, Dickerson P, Stribling P, Dautenhahn K (2004) Robot-mediated joint attention in children with autism: A case study in a robot-human interaction. *Interaction studies: Social Behaviour and Communication in Biological and Artificial Systems*, John Benjamins Publishing Company, Amsterdam 5:2:161-198.
- Robins, B., Dautenhahn, K., Te Boekhorst, R., & Billard, A. (2005). Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills?. *Universal Access in the Information Society*, 4(2), 105-120.
- Saint-Aimé, S., Le-Pévédic, B., & Duhaut, D. (2009, December). Experimentation to evaluate EmotiRob interaction model. In *Robotics and Biomimetics (ROBIO), 2009 IEEE International Conference on* (pp. 1524-1529). IEEE.
- Schaal, S., Atkenson, C., & Vijayakumar, S. (2000). Real time robot learning with locally DEST "nad-r41"> weighted statistical learning. *IEEE International*

- Conference on Robotics and Automation, (pp.288–293). CA: San Francisco
- Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. *Approaches to emotion*, 2293, 317.
- Shi, J., & Malik, J. (2000). Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence*, 22(8), 888-905.
- Siino, R. M., Chung, J., & Hinds, P. J. (2008, August). Colleague vs. tool: Effects of disclosure in human-robot collaboration. In *Robot and Human Interactive Communication, 2008. RO-MAN 2008. The 17th IEEE International Symposium on* (pp. 558-562). IEEE.
- Squire, L. R. (2004). Memory systems of the brain: a brief history and current perspective. *Neurobiology of learning and memory*, 82(3), 171-177.
- Steunebrink, B. R., Dastani, M., & Meyer, J. J. C. (2012). A formal model of emotion triggers: an approach for BDI agents. *Synthese*, 185(1), 83-129.
- Strickland D (1996) A virtual reality application with autistic children. *Presence: Teleoperators and Virtual Environment* 5(3):319-329.
- Sun, R. (2004). Desiderata for cognitive architectures. *Philosophical Psychology*, 17(3), 341-373.
- Sun, R. (2007). The importance of cognitive architectures: An analysis based on CLARION. *Journal of Experimental & Theoretical Artificial Intelligence*, 19(2), 159-193.
- Tapus, A., & Mataric, M. J. (2007, March). Emulating Empathy in Socially Assistive Robotics. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics* (pp. 93-96).
- Tiegerman, E., & Primavera, L. H. (1984). Imitating the autistic child: Facilitating communicative gaze behavior. *Journal of autism and developmental disorders*, 14(1), 27-38.
- Trafton, G., Hiatt, L., Harrison, A., Tamborello, F., Khemlani, S., & Schultz, A. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1), 30-55.
- Tulving, E. (2002). Episodic memory: from mind to brain. *Annual review of psychology*, 53(1), 1-25.
- Torta, E., van Dijk, E., Ruijten, P. A., & Cuijpers, R. H. (2013, October). The ultimatum game as measurement tool for anthropomorphism in human-robot interaction. In *International Conference on Social Robotics* (pp. 209-217). Springer, Cham.
- Tversky, B., Lee, P., & Mainwaring, S. (1999). Why do speakers mix perspectives?. *Spatial cognition and computation*, 1(4), 399-412.
- Užgiris, I.C. (1981), Two functions of imitation during infancy. *International Journal of Behavioral Development*, 4, 1–12.
- Užgiris, I.C., Broome, S. & Kruper, J. (1989). Imitation in mother-child conversations: A focus on the mother. In G.E. Speidel & K.E. Nelson (eds.), *The many faces of imitation in language learning* (pp. 91–120). New York: Springer-Verlag

- Vallerand, R. J., Gauvin, L. I., & Halliwell, W. R. (1986). Negative effects of competition on children's intrinsic motivation. *The Journal of Social Psychology, 126*(5), 649-656.
- Van Hoorn, A., Waller, J., & Hasselbring, W. (2012, April). Kieker: A framework for application performance monitoring and dynamic software analysis. In *Proceedings of the 3rd ACM/SPEC International Conference on Performance Engineering* (pp. 247-248). ACM.
- Velásquez, J. (1998). Modeling emotion-based decision-making. *Emotional and intelligent: The tangled knot of cognition*, 164-169.
- Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2017). Show and tell: Lessons learned from the 2015 mscoco image captioning challenge. *IEEE transactions on pattern analysis and machine intelligence, 39*(4), 652-663.
- Zeng, Z., Hu, Y., Roisman, G., Wen, Z., Fu, Y., & Huang, T. (2007). Audio-visual spontaneous emotion recognition. *Artificial intelligence for human computing, 72-90*.
- Zhang, H., Manocha, D., Hudson, T., & Hoff III, K. E. (1997, August). Visibility culling using hierarchical occlusion maps. In *Proceedings of the 24th annual conference on Computer graphics and interactive techniques* (pp. 77-88). ACM Press/Addison-Wesley Publishing Co..
- Zhang, Y., & Ji, Q. (2005). Active and dynamic information fusion for facial expression understanding from image sequences. *IEEE Transactions on pattern analysis and machine intelligence, 27*(5), 699-714.
- Zhang, X., Yin, L., Cohn, J. F., Canavan, S., Reale, M., Horowitz, A., ... & Girard, J. M. (2014). Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image and Vision Computing, 32*(10), 692-706.
- Zilbovicius, M. (2004). Imagerie cérébrale et autisme infantile. *Revue Cerveau et Psychologie*.
- Ziaeeffard, M., & Bergevin, R. (2015). Semantic human activity recognition: a literature review. *Pattern Recognition, 48*(8), 2329-2345.
- Wainer, J., Feil-Seifer, D. J., Shell, D. A., & Mataric, M. J. (2007, August). Embodiment and human-robot interaction: A task-based perspective. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on* (pp. 872-877). IEEE.
- Weir, S., & Emanuel, R. (1976). *Using LOGO to catalyze communication in an autistic child*. University of Edinburgh. Department of Artificial Intelligence.
- Weng, W. C. (2007). U.S. Patent Application No. 11/710,951.
- Werry I, Dautenhahn K, Harwin W (2001) Investigating a robot as a therapy partner for children with autism. Proc. of the 6th European Conference for the Advancement of Assistive Technology (AAATE 2001), 3-6 September. Ljubljana, Slovenia.
- Werry I, Dautenhahn K, Ogden B, Harwin W (2001) Can Social Interaction Skills Be Taught by a Social Agent? The Role of a Robotic Mediator in Autism

- Therapy. In Beynon M, Nehaniv CL, Dautenhahn K (eds): Proc. CT2001, The Fourth International Conference on Cognitive Technology: Instruments of Mind, LNAI 2117. Berlin Heidelberg, Springer-Verlag, p 57-74.
- Wetherby, A. M., & Prutting, C. A. (1984). Profiles of communicative and cognitive-social abilities in autistic children. *Journal of Speech and Hearing Research*, 28, 364–377.
- Whalen, C., & Schreibman, L. (2003). Joint attention training for children with autism using behavior modification procedures. *Journal of Child Psychology and Psychiatry*, 44, 456–468.
- Wing L (1996) *The Autistic Spectrum*. London, Constable Press.
- Wyss, R., König, P., & Verschure, P. F. J. (2006). A model of the ventral visual system based on temporal stability and local memory. *PLoS biology*, 4(5), e120.
- Yando, R., Seitz, V., & Zigler, E. (1978). *Imitation: A developmental perspective*. Hillsdale, NJ: Lawrence Erlbaum.