

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

Campus di Cesena
Scuola di Ingegneria e Architettura
Corso di Laurea in Ingegneria Biomedica

**AUTISMO E NORMOTIPICITÀ:
ANALISI MODELLISTICA
DEI PROCESSI SENSORIALI**

Tesi in

Sistemi Neurali

Relatore:
Prof. Mauro Ursino

Presentata da:
Daniela Pisani

Correlatore:
Ing. Cristiano Cuppini

**Sessione I
Anno Accademico 2016-2017**

Sembra che per avere successo
nella scienza o nell'arte
un pizzico di autismo sia fondamentale.

Hans Asperger

Indice

Introduzione	1
1 L'integrazione Multisensoriale	5
1.1 Cos'è l'integrazione multisensoriale	5
1.2 Evoluzione di un cervello multisensoriale	6
1.3 Plasticità cerebrale	9
1.4 Apprendimento e Memoria in un contesto multisensoriale	10
1.5 Parametri per l'identificazione dello stimolo	12
1.6 Integrazione multisensoriale nei singoli neuroni	13
1.7 Integrazione multisensoriale nelle popolazioni neurali . .	17
1.8 Comprendere l'attività neurale attraverso comportamento e percezione	20
2 Analisi dei processi sensoriali nei soggetti autistici	23
2.1 L'autismo attraverso l'effetto McGurk	23
2.2 Sensibilità al McGurk effect nei soggetti con DSA(studio di Taylor)	25
2.3 Integrazione audiovisiva in condizioni rumorose(studio di Foxy)	27
2.3.1 Metodo e Analisi	27
2.3.2 Effetti del SNR, delle articolazione dello <i>speaker</i> e dell'età sul riconoscimento delle parole	29
2.3.3 Differenze tra normotipici e autistici	29
2.3.4 <i>Eye Tracking</i>	32

2.3.5	Discussione finale	33
2.4	Paradigma del ‘discorso nel rumore’ (studio di Smith e Bennetto)	34
2.4.1	Metodo e Analisi	34
2.4.2	Risultati a confronto	36
2.4.3	Discussione finale	37
2.5	Risultati discordanti degli studi condotti in presenza di rumore	38
3	Modello Matematico della Rete Neurale	41
3.1	Introduzione al modello	41
3.2	Descrizione qualitativa del modello	42
3.3	Descrizione matematica della rete neurale	45
4	Risultati	51
4.1	Scelta del caso d’interesse	52
4.2	Verso la simulazione del soggetto con DSA	57
4.2.1	Danneggiamento connessioni cross-modali	58
4.2.2	Danneggiamento sinapsi <i>feedforward</i>	60
4.2.3	Confronto fra le due ipotesi	61
4.3	Risposta del bambino autistico	64
	Conclusioni	69
	Bibliografia	73
	Ringraziamenti	75

Introduzione

Per molti anni, gli studi sulla percezione si sono focalizzati sul comprendere come le informazioni venissero elaborate, filtrate e processate per le singole modalità, apportando molteplici informazioni su come le singole modalità operassero.

Però nella realtà che ci circonda, le informazioni provenienti dagli oggetti o dagli eventi non attivano una singola modalità sensoriale, bensì attivano molteplici modalità. Ad esempio, quando un interlocutore ci sta parlando, noi lo ascoltiamo e allo stesso tempo lo stiamo guardando; di conseguenza le informazioni che riceviamo vengono elaborate in due modalità diverse, le quali a loro volta comunicano. È proprio per questo motivo che negli ultimi anni gli studi riguardanti i processi sensoriali hanno subito un cambio di direzione, analizzando come le informazioni vengano processate da una rete neurale multisensoriale, dinamica, capace di mettere in correlazione le diverse aree del cervello deputate all'elaborazione delle diverse modalità sensoriali. Attraverso questo tipo di elaborazione il cervello è capace di decidere quali informazioni mettere in correlazione e quali separare.

Questa capacità di comunicazione tra le diverse modalità sensoriali, permette di elaborare le informazioni, in modo ottimale, producendo alcuni effetti, che potremmo definire, di facilitazione o interferenza. Perciò in un contesto multisensoriale, in cui l'individuo riceve gli stimoli provenienti da diverse modalità, i processi di apprendimento, memoria ed elaborazione dell'informazione vengono facilitati.

È stato riscontrato che nei soggetti da disturbi dello spettro autistico (DSA) la capacità di integrare informazioni multisensoriali è ridotta rispetto a un soggetto normotipico, ossia un individuo le cui capacità intellettive e neurologiche possano essere considerate nella norma.

In particolare, se per un individuo normotipico, il contesto multisensoriale apporta un miglioramento nel processo di elaborazione, per un individuo affetto da disturbi dello spettro autistico, la stimolazione multisensoriale non comporta un miglioramento apprezzabile nella comprensione delle informazioni che lo circondano.

Di fatto la stimolazione multisensoriale è alla base delle abilità sociali e linguistiche degli individui. I soggetti affetti da questo disturbo manifestano grandi difficoltà proprio nei contesti sociali e nella comprensione del parlato.

In questo progetto di Tesi si vuole analizzare, attraverso l'uso di un modello neurocomputazionale, come i problemi derivanti dall'autismo possano essere influenzati da un indebolimento delle sinapsi tra le due regioni unisensoriali, visiva ed uditiva. Tali sinapsi prendono il nome di sinapsi cross-modali.

Per quantificare tale effetto, viene utilizzato uno dei metodi di illusione audiovisiva, l'effetto McGurk, il quale permette di spiegare come per i soggetti autistici, i processi multisensoriali avvengono in maniera differente rispetto ai soggetti normotipici.

Difatti gli effetti dell'illusione audiovisiva sono particolarmente ridotti nei soggetti autistici, ciò sta a dimostrare la ridotta capacità di integrare gli stimoli per quest'ultimi. In questo metodo gli stimoli presi in considerazione sono lo stimolo acustico e lo stimolo visivo.

Inoltre per poter effettuare questo studio viene utilizzato un modello matematico basato su una rete neurale; il modello include connessioni tra due regioni unisensoriali che a loro volta sono connesse con una regione multisensoriale. Attraverso questo modello viene simulata l'integrazione e l'elaborazione multisensoriale, sia a livello unisensoriale che a livello

multisensoriale.

Le simulazioni che sono state effettuate mirano ad identificare le differenze dei processi sensoriali nei soggetti affetti da disturbi dello spettro autistico rispetto ai soggetti normotipici.

La Tesi è articolata come segue:

Il Capitolo 1 fornisce un'introduzione al concetto di integrazione multisensoriale, spiegando come e dove avvengono i processi integrativi

Il Capitolo 2 descrive il disturbo dello spettro autistico, possibili cause di tale disturbo e successiva descrizione dell'effetto McGurk utilizzato per descrivere l'integrazione multisensoriale nei soggetti autistici

Il Capitolo 3 invece descrive il modello matematico e fornisce alcune nozioni sulle reti neurali

Infine Il Capitolo 4 conterrà le informazioni riguardanti le prove effettuate attraverso l'ausilio del modello matematico con le relative Conclusioni

Capitolo 1

L'integrazione Multisensoriale

1.1 Cos'è l'integrazione multisensoriale

Ogni giorno gli esseri umani vengono investiti da una moltitudine di informazioni, che possono essere correlate tra di loro o essere discordanti. Chi ha il compito di decidere quali informazioni devono essere integrate e quali invece separate è il cervello.

Per molti anni, la ricerca scientifica si è basata sullo studio delle singole modalità sensoriali, indagando su come le informazioni vi giungono e su come poi vengono elaborate e filtrate. Grazie a questi studi, la ricerca ha raccolto una enorme mole di dati che ci permettono di capire come funzionino singolarmente le modalità sensoriali.

Sebbene queste informazioni sono assai importanti, ci accorgiamo che il mondo esterno quando cerca di comunicare non lo fa attraverso un'unica modalità. È proprio per questo motivo che negli ultimi anni, gli studi riguardanti le modalità sensoriali hanno subito un cambio di rotta, basandosi sull'ipotesi che le informazioni che giungono al nostro cervello in realtà non provengono da singole modalità, ma attraverso molteplici modalità sensoriali.

Il cervello viene funzionare come una rete neurale multisensoriale, altamente dinamica e interattiva, in grado di mettere in correlazione le diverse

aree del cervello, al fine di integrare o separare le informazioni che vi giungono.

Il fenomeno per il quale le informazioni ci pervengono attraverso una rete sensoriale multipla, è detto convergenza cross-modale o integrazione multisensoriale. L'integrazione multisensoriale è un processo che viene generato nel momento in cui sopraggiungono, all'individuo, diverse informazioni da diverse modalità sensoriali e vengono identificate e successivamente elaborate, come se provenissero da un unico evento o un unico oggetto fisico.

Nonostante le regioni cerebrali, le quali sono soggette al fenomeno dell'integrazione multisensoriale, siano diverse, vi sono dei concetti fondamentali che le accomunano e che governano i processi integrativi. I concetti alla base dell'integrazione multisensoriale sono la vicinanza spaziale e il sincronismo temporale.

Infatti stimoli che sopraggiungono vicini, sia temporalmente che spazialmente, generano risposte ad alto guadagno. Mentre invece stimoli che giungono separatamente generano risposte deboli.

Un altro concetto riguarda l'efficacia degli stimoli stessi. L'accoppiamento di stimoli seppur debolmente efficaci, permette, in genere, di creare una risposta, in termini di guadagno, elevata.

Per questo motivo, possiamo affermare che l'integrazione delle informazioni comporta una vasta gamma di miglioramenti dal punto di vista comportamentale e percettivo, nonché delle funzioni cognitive.

1.2 Evoluzione di un cervello multisensoriale

Il mondo che ci circonda è costituito da una vasta gamma di informazioni, le quali vengono ogni giorno codificate dalle diverse strutture sensoriali. Per far ciò il nostro sistema nervoso è organizzato in architetture specializzate per poter integrare e combinare tutti questi dati multisensoriali. Grazie a tale struttura cerebrale, le informazioni vengono interpretate

attraverso il fenomeno del *binding* sensoriale, ossia il processo attraverso il quale informazioni differenti vengono codificate come provenienti da un unico evento o oggetto fisico, apportando notevoli benefici comportamentali e percettivi.

Il riconoscimento e l'apprendimento del linguaggio stesso sono il risultato dell'integrazione delle informazioni provenienti dalle modalità visive e acustiche.

Storicamente si pensava che le funzioni multisensoriali si verificassero nelle regioni associative di più alto livello, mentre gli studi recenti hanno invece dimostrato che regioni unisensoriali sono capaci di influenzarsi l'un l'altra sin nei primi stadi di elaborazione.

In base agli studi condotti sui modelli animali è stato possibile capire come avvengono le interazioni tra i vari sistemi sensoriali e come maturano durante il corso della vita.

Infatti, gli *input* provenienti da sistemi unisensoriali si presentano in strutture che successivamente diverranno multisensoriali, e man mano che gli ingressi dei differenti sensi convergono, si formano i cosiddetti neuroni multisensoriali e con il progredire dell'individuo, tali neuroni aumentano.

Basandoci sugli studi condotti sul collicolo superiore (CS) del gatto [1], possiamo notare come progredisce la struttura multisensoriale. La reattività somatosensitiva è presente sin dalla nascita, mentre quella acustica matura dopo una settimana di vita, invece la reattività visiva si presenta alcune settimane dopo. Le prime interazioni che si verificano sono quelle di tipo somatosensitive-acustiche, seguite dalle interazioni tra la vista e gli altri sensi. È importante notare che i neuroni multisensoriali immaturi, seppur capaci di rispondere agli stimoli provenienti da diverse modalità, sono assai diversi da quelli adulti. Man mano che il gatto cresce, i neuroni multisensoriali del CS aumentano di incidenza fino a rappresentare più della metà dei neuroni del CS. Questi neuroni

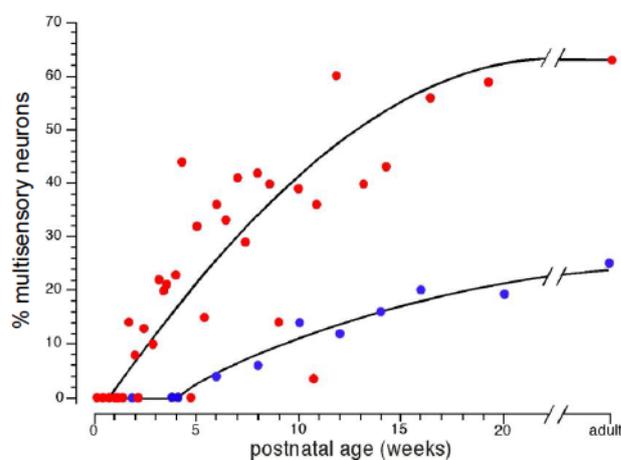


Figura 1.1: Cronologia dello sviluppo della presenza dei neuroni multisensoriali in due strutture cerebrali del gatto, il collicolo superiore (in rosso) del mesencefalo e la regione corticale che circonda il solco silviano anteriore (in blu).

gradualmente sviluppano la capacità di integrare le informazioni provenienti dalle diverse modalità potendo così variare i profili di risposta.

Le capacità apprese durante lo sviluppo postnatale gioca un ruolo fondamentale nello sviluppo multisensoriale.

Per quanto riguarda gli esseri umani, gli studi condotti sullo sviluppo cerebrale nel periodo postnatale, hanno evidenziato che i processi multisensoriali emergono in modo graduale e che, come per gli animali, le prime esperienze giocano anche qui un ruolo fondamentale.

Alla nascita, il neonato riesce a percepire caratteristiche multisensoriali di base come l'intensità e il sincronismo temporale. Durante la crescita, riesce ad acquisire capacità di elaborazione multisensoriale sempre più sofisticate. Solitamente, le proprietà di oggetti ed eventi che provengono da una sorgente comune, sono identificate in maniera sincrona a livello spaziale e temporale. Tali livelli di percezione più bassi, offrono una soluzione al fenomeno del *binding* sensoriale. È stato osservato che i neonati, per riconoscere un oggetto, comprendere un evento, o acquisire informazioni sul linguaggio, si affidano alle relazioni spazio-temporali.

Possiamo infatti dire che anche per gli esseri umani l'esperienza gioca un ruolo fondamentale per lo sviluppo dei processi multisensoriali.

Inoltre, l'acquisizione di un'esperienza percettiva in un ambiente multisensoriale nativo si basa su un processo di restringimento percettivo multisensoriale (RPM). Ad esempio, un bambino di 8-10 mesi, che linguisticamente è molto attivo, risponde alla ridondanza multisensoriale per tutti gli stimoli visivi-acustici di un discorso, invece un bambino di 12 mesi, il quale è andato incontro ad un processo di restringimento percettivo, e quindi ha già acquisito informazioni sul discorso nativo, tenderà a concentrarsi sugli *input* audio-visivi salienti del linguaggio.

Per quanto riguarda invece lo sviluppo dei processi multisensoriali nell'adolescenza, è stata identificata una finestra temporale di *binding*, ossia un costrutto probabilistico che rappresenta l'intervallo di tempo fra lo stimolo acustico da quello visivo, entro il quale il *binding* è altamente probabile.

Gli studi hanno evidenziato che tale finestra non matura fino all'adolescenza, inoltre tende a maturare a velocità diverse a seconda dei diversi stimoli. Si è notato che matura lentamente quando l'accoppiamento degli stimoli audio-visivi è di tipo semplice (*flash* e *beep*), mentre matura velocemente quando gli stimoli audio-visivi riguardano il linguaggio, i quali hanno maggiore rilevanza per il comportamento e la percezione. Questo potrebbe essere interpretato come una conseguenza del restringimento percettivo multisensoriale (RPM).

1.3 Plasticità cerebrale

Il cervello è in grado di immagazzinare una enorme mole di informazione proveniente dal mondo esterno creando nuovi collegamenti o modificando i collegamenti esistenti fra i neuroni, sulla base dell'esperienza. È questo il fenomeno della neuroplasticità o plasticità cerebrale.

Recenti studi hanno dimostrato la stretta correlazione tra la neuroplasticità e i sistemi multisensoriali nell'uomo adulto. Tali ricerche, mostrano come l'addestramento su un compito semplice, dove bisogna dare un giudizio sulla tempistica dell'accoppiamento di stimoli audio-visivi, apporta un evidente miglioramento per l'acuità temporale multisensoriale. Inoltre dopo 5 giorni di addestramento, durante i quali ai soggetti veniva chiesto di adempiere al loro compito attraverso dei giudizi positivi o negativi, l'acuità temporale multisensoriale subiva un netto miglioramento pari al 40%.

Grazie all'addestramento, i soggetti avevano acquisito nuove capacità, le quali si manifestavano rapidamente (dopo solo 1 giorno di formazione), erano durature (permanevano anche oltre una settimana dalla fine dell'addestramento) e, cosa più importante, riuscivano a generalizzare i concetti riuscendo ad adempiere a compiti che andavano al di là di quelli eseguiti durante la formazione.

Successivi studi, si sono poi focalizzati sull'identificazione delle basi neurali per questa plasticità percettiva multisensoriale, scoprendo che una rete presente nella parte posteriore del solco temporale superiore, la quale coinvolge le regioni delle corteccie acustiche e visive, mostrava sostanziali cambiamenti nell'attivazione a seguito dell'addestramento.

1.4 Apprendimento e Memoria in un contesto multisensoriale

Riuscire a memorizzare elementi che si presentano frequentemente, di tipo uditivo e visivo, risulta agevolato nel momento in cui si presentano in un contesto multisensoriale semanticamente congruente, rispetto a quando si presentano in un contesto puramente unisensoriale.

Se invece si considera un contesto multisensoriale semanticamente incongruente, o si coinvolge uno stimolo non significativo (*task-irrelevant*), la *performance* di memorizzazione risultata compromessa o inalterata.

Questi dati evidenziano che le informazioni multisensoriali hanno un effetto relativamente duraturo sulla successiva elaborazione e recupero delle informazioni unisensoriali. Le informazioni in arrivo sono diversificate in base alle esperienze pregresse, tali effetti si verificano nelle fasi iniziali dell'elaborazione dello stimolo, nei livelli corticali più bassi.

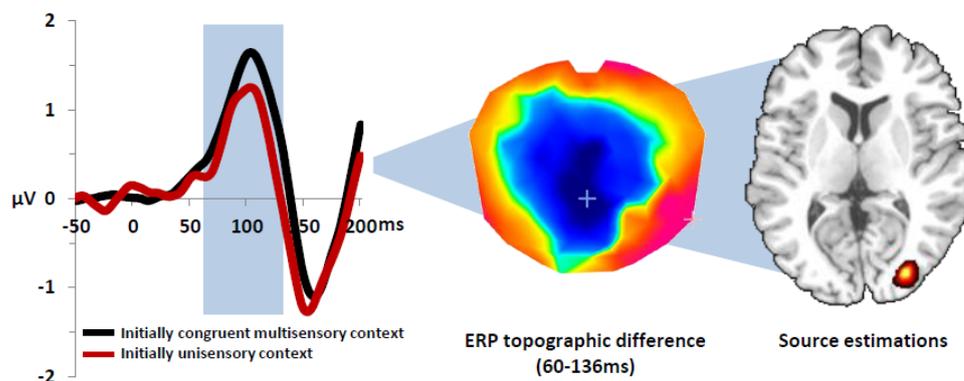


Figura 1.2: Le risposte cerebrali ai ripetuti stimoli variano a seconda che il contesto iniziale sia stato multisensoriale o unisensoriale.

Inoltre il modo in cui un essere umano integra le informazioni acustiche e visive dipende dalle precedenti/nuove esperienze di un compito. Riuscire a comprendere se durante il riconoscimento di un oggetto l'individuo mostra un miglioramento o un peggioramento, potrebbe dipendere dalla forza delle risposte del cervello del soggetto a seguito di una stimolazione multisensoriale, e non unisensoriale. Ciò che viene evidenziato è che il ricordo di oggetti che incontriamo in un contesto multisensoriale è più forte rispetto a quando il ricordo viene associato ad un contesto unisensoriale, ossia solo acustico o solo visivo.

Ciò che emerge è il vantaggio intuitivo di lavorare in un contesto multisensoriale, così da non limitare l'elaborazione dei processi dei soli stimoli correnti.

1.5 Parametri per l'identificazione dello stimolo

Dal punto di vista comportamentale e percettivo, attraverso studi condotti sia su animali che su esseri umani, sono stati identificati dei parametri che vengono utilizzati per quantificare le differenti risposte agli stimoli multisensoriali. Essi sono la percezione e l'accuratezza.

In un tipico modello psicofisiologico, utilizzato per mostrare le caratteristiche di percezione e accuratezza, viene presentato ripetutamente uno stimolo multisensoriale, sotto forma di combinazione audiovisiva, ogni tanto intervallato da stimoli unisensoriali. I partecipanti devono dare informazioni riguardo la capacità di percepire, la localizzazione, l'identificazione o dare dei giudizi riguardo gli/lo stimoli/o per ogni prova.

Un'integrazione multisensoriale veniva identificata quando la risposta allo stimolo multisensoriale si discostava dalla risposta ottenuta attraverso stimolo unisensoriale. Quindi se il tasso di rilevamento/accuratezza era superiore a quello visto per entrambi gli stimoli unisensoriali, allora veniva riconosciuta l'integrazione multisensoriale. Ossia:

$$\hat{p}(AV) > \max[p(A), p(V)] \quad (1.1)$$

Sebbene si verifichi un relativo incremento della risposta rispetto a quella massima unisensoriale, indicando che il partecipante ha utilizzato l'informazione da entrambi i sensi, non è detto che sia stata integrata l'informazione o che si stia verificando un fenomeno di facilitazione statistica. La formula si modifica in tal modo se tengo conto delle facilitazioni:

$$\hat{p}(AV) > p(A) + p(V) - p(A) \times p(V) \quad (1.2)$$

L'equazione indica che la risposta attesa per uno stimolo audiovisivo(AV)

è pari alla somma della probabilità di individuazione di ogni presentazione unisensoriale, a cui viene sottratta la probabilità che entrambi gli stimoli unisensoriali (uditivo e visivo) siano individuati nella stessa prova. Per ogni risultato che supera il valore atteso, l'informazione risulta integrata.

1.6 Integrazione multisensoriale nei singoli neuroni

I primi studi neuronali sull'integrazione multisensoriale vennero condotti sulla struttura del mesencefalo dei mammiferi, il collicolo superiore. Grazie alla elevata incidenza di neuroni multisensoriali, tale organizzazione forniva una struttura funzionale in grado di descrivere le proprietà di base della risposta, le caratteristiche integrative e i meccanismi attraverso i quali i neuroni multisensoriali integrano gli stimoli attraverso le modalità sensoriali.

L'integrazione multisensoriale veniva definita sulla base della frequenza di potenziali d'azione che venivano generati in risposta a un dato stimolo. Il picco che si generava veniva utilizzato come criterio per distinguere se si era verificata un'integrazione dell'informazione a seguito di un ingresso combinato (audiovisivo AV) o se rappresentava la risposta a un *input* unisensoriale (solo acustico A o solo visivo V).

Per quantificare l'integrazione multisensoriale veniva utilizzato un indice detto *Interactive Index* (*ii*) [2],[3].

$$ii = \frac{AV - \max(A, V)}{\max(A, V)} \times 100 \quad (1.3)$$

Dove le variabili AV, A e V rappresentano la frequenza dei potenziali misurati per ogni stimolo dopo che l'attività spontanea è stata rimossa.

Ogni valore diverso da zero di ii rappresenta un'integrazione multisensoriale, e può essere utilizzato per identificare il guadagno o la potenza dell'integrazione. Se si verifica un incremento della velocità di scarica (*firing rate*) rispetto allo stimolo unisensoriale più consistente esso si traduce in valori positivi dell'indice ii e indica un miglioramento della risposta; mentre un decremento della velocità di scarica si traduce in valori negativi dell'indice e ciò indica invece un'involuzione della risposta. L'*Interactive Index(ii)* non è l'unico parametro utilizzato per identificare e quantificare l'integrazione multisensoriale. Infatti, la determinazione dei potenziali d'azione può essere utilizzata per comparare le risposte multisensoriali rispetto a quelle unisensoriali utilizzando un modello additivo, utilizzando la sommatoria lineare di entrambe le risposte unisensoriali. L'indice finale utilizzato per quantificare l'integrazione, prende il nome di *mean statistical contrast (msc)*:

$$msc = \frac{\sum_{i=1}^n [AV_i - (A_i + V_i)]}{n} \quad (1.4)$$

Dove n è il numero totale delle prove effettuate per ogni modalità. Tale modello, inoltre, assume che gli *input* per ogni modalità sono indipendenti tra loro, e distingue tre valori di msc : risposta additiva ($msc=0$), risposta subadditiva ($msc<0$) e risposta superadditiva ($msc>0$).

Attraverso tale metodo si trae il vantaggio di incorporare entrambe le risposte unisensoriali per poter valutare gli effetti dell'integrazione, potendosi basare su un *set* completo di ingressi per quantificare il guadagno o la perdita di integrazione multisensoriale.

È possibile notare che l' ii fornisce, una soglia per riconoscere se l'integrazione multisensoriale è avvenuta o meno, e inoltre fornisce il guadagno della risposta quando si utilizza una seconda modalità sensoriale. Al contrario, l' msc permette di avere molte più informazioni attraverso l'utilizzo di entrambe le modalità sensoriali. Inoltre se ii assume valori positivi,

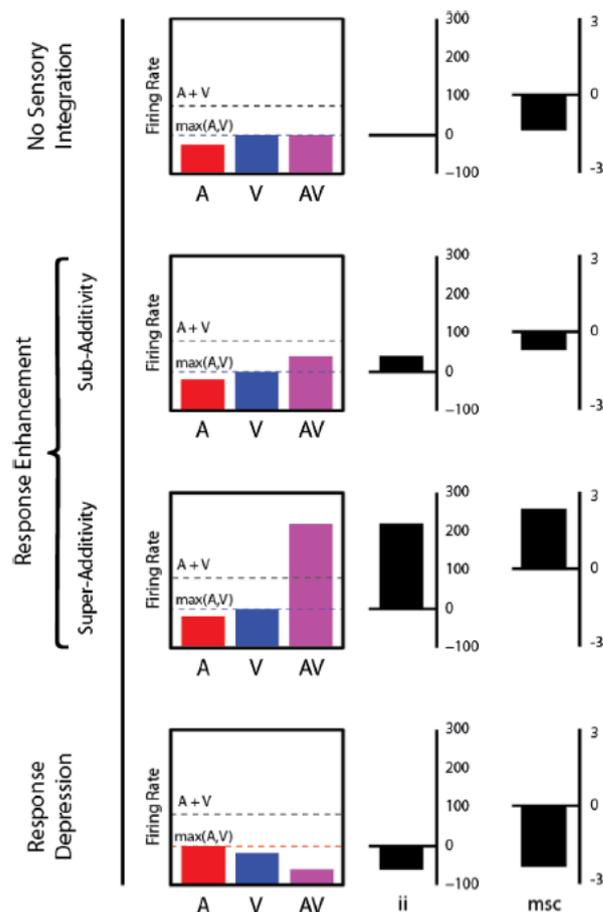


Figura 1.3: Singoli neuroni che rispondono a entrambe le stimolazioni uditive e visive, possono essere caratterizzati da due differenti quantificazioni, l'emphInteractive Index *ii* o *Mean Statistical Contrast msc*

che rappresentano un miglioramento della risposta, essa si può manifestare sotto forma di additività, superadditività o subadditività; mentre invece i valori negativi di *ii* che indicano una involuzione della risposta, può manifestarsi solo come subadditività.

Tali manifestazioni dell'integrazione multisensoriale nei singoli neuroni, non assumono valori prefissati, ma dipendono per lo più da una vasta gamma di fattori. Infatti, le caratteristiche del campo recettivo dei neuroni multisensoriali sono fortemente influenzate dalla natura e dalla potenza dell'integrazione multisensoriale.

Tali cambiamenti si ottengono, facendo variare le seguenti caratteristiche:

- la natura spaziotemporale del campo recettivo
- le relazioni spaziali e temporali tra gli stimoli multisensoriali
- l'efficacia della risposta all'interno dell'eterogeneo campo recettivo.

Ad esempio, se uno stimolo multisensoriale (audiovisivo AV) è presentato in maniera tale che gli stimoli visivi e acustici rientrano nei rispettivi campi recettivi, è molto probabile che si avrà un incremento della risposta, anche quando gli stimoli non si trovano nell'esatta locazione spaziale. Se invece succede che uno dei due stimoli si trova nel proprio campo recettivo e l'altro no, è altamente probabile che si verifichi una depressione della risposta.

Per quanto riguarda l'elaborazione temporale dell'informazione multisensoriale, è stata manipolata la relazione tra gli stimoli multisensoriali accoppiati ed è stato esaminato il conseguente impatto sulle frequenze di scarica dei neuroni multisensoriali. Il più grande guadagno in termini di risposta, si otteneva nel momento in cui i periodi dei picchi di scarica delle singole risposte sensoriali venivano sovrapposte. Se invece gli stimoli venivano presentati in maniera asincrona, si aveva un decremento della risposta. Di fatto, se la disparità temporale tra gli stimoli era sufficientemente grande, si otteneva una depressione.

I fattori temporali e spaziali giocano un ruolo assai importante nell'integrazione multisensoriale. Di fatto, gli stimoli multisensoriali di un oggetto o di un evento, nel mondo reale, si presentano vicini sia spazialmente che temporalmente. Pertanto, le statistiche spaziotemporali di un complesso stimolo multisensoriale e il rapporto di questi stimoli, con l'organizzazione del campo recettivo dei neuroni multisensoriali, assumono un ruolo preponderante per la loro integrazione o *binding* (fusione).

Infine, il terzo fattore che influenza l'integrazione multisensoriale è l'efficacia della componente dello stimolo all'interno del campo recettivo

neuronal. In tale contesto, più sarà debole la risposta a un dato stimolo unisensoriale, più grande sarà il beneficio offerto dalla stimolazione multisensoriale traducendosi in una integrazione. Al contrario, più la risposta unisensoriale aumenta, più l'incremento offerto dalla stimolazione multisensoriale diminuirà. Quindi, le combinazioni di stimoli acustici e visivi deboli produrranno un significativo aumento della risposta, più gli stimoli saranno forti, più tale aumento diminuirà. Tale principio prende il nome di *Efficacia Inversa*.

1.7 Integrazione multisensoriale nelle popolazioni neurali

Anche se l'attività a livello dei singoli neuroni ci fornisce importanti informazioni riguardo l'integrazione multisensoriale, purtroppo fallisce quando si vuole analizzare i processi sotto-soglia a livello sinaptico. Per questo motivo bisogna introdurre lo studio sui *Local Field Potentials (LFPs)*, il quale fornisce una misura dell'elaborazione sinaptica e sull'architettura di *input* locale di una specifica area cerebrale.

Il segnale LFP fornisce una misura delle variazioni di tensione attorno alla punta dell'elettrodo, in riferimento al segnale di *BOLD (Blood Oxygenation Level Dependent)* misurato dall'*fMRI (functional Magnetic Resonance Imaging)*. È costituito da due componenti principali:

- componente ad alta frequenza (0,5-3 kHz), che riflette in gran parte l'attività dei potenziali locali entro una gamma spaziale di 150-300 μ m
- componente a bassa frequenza (<200 Hz), che deriva in gran parte dai cambiamenti di potenziale post-sinaptico attorno alla punta dell'elettrodo.

Può essere inoltre scomposto nelle componenti di bande di frequenza come le onde dell'*EEG (Electroencephalography)*: *delta* (1-4 Hz), *theta*

(4-8 Hz), *alpha* (8-12 Hz), *beta* (12-30 Hz) e *gamma* (>30 Hz). L'energia di queste differenti bande di frequenza è correlata ai differenti aspetti dei processi sensoriali che includono: selettività dello stimolo, allocazione dell'attenzione e selettività della categoria-oggetto.

Il segnale LPF è diventato un importante strumento per lo studio dei processi multisensoriali. Tanto è vero che ciò è stato confermato da recenti osservazioni sull'influenza sotto-soglia di stimoli provenienti dalle modalità *non-driving* o *non-dominant* sui processi sensoriali in quelle aree del cervello tradizionalmente considerate *sensory-specific*, quali la corteccia uditiva e quella visiva.

Infatti, utilizzando il LPF per gli studi della corteccia uditiva dei primati, si è visto che avveniva una buona integrazione di volti e voci, indice di integrazione multisensoriale.

Successivamente, unendo i dati scaturiti dal segnale LPF con un approccio teorico, si è notato che nella corteccia visiva i tempi di *spike* relativi alle fasi di oscillazione di LPF nelle bande delta, portavano informazioni sul contesto naturalistico.

Mentre nella corteccia uditiva, i tempi di *spike* relativi alle fasi di oscillazioni di LPF nelle bande *theta*, portavano informazioni riguardo i tipi di suoni ascoltati.

Purtroppo la natura intrinseca del segnale, solleva importanti dubbi a causa della sua elevata diffusione spaziale, stimata intorno ai 500-800 μm , poiché le sue registrazioni potrebbero includere sia eventi locali che eventi elettrofisiologici lontani.

Per ovviare a tale problema, vengono utilizzati degli elettrodi multi-contatto che ricoprono tutto lo spessore corticale, per poter ottenere un'analisi laminare del segnale LPF.

Utilizzando poi la derivata spaziale del segnale si riduce la diffusione spaziale, così da includere solo gli eventi locali.

Per poter valutare i cambiamenti dell'attività multisensoriale a livello delle popolazioni neurali, Stevenson e colleghi [4], si sono concentrati

sull'utilizzo della codifica neurale. Uno degli strumenti più utilizzati è il CCG (*cross-correlelogram*), il quale mette in correlazione l'attività tra coppie di neuroni e descrive i cambiamenti della probabilità di scarica di un neurone bersaglio relativo al tempo di scarica di un neurone di riferimento. Quando una coppia di neuroni si attiva in maniera sincrona, nel CCG appaiono picchi e avvallamenti che stanno ad indicare rispettivamente interazioni eccitatorie e interazioni inibitorie.

Per poter analizzare e quantificare l'integrazione multisensoriale nelle popolazioni neurali degli esseri umani, viene utilizzato l'elettroencefalogramma (EEG) per la sua non-invasività, il quale spesso utilizza un segnale che prende il nome di *potenziale event-related (ERP)*.

Utilizzando questo metodo per registrare i potenziali dei neuroni, vi sono delle problematiche legate alla risoluzione spaziale che risulta grossolana: a causa della grandezza dell'area è impossibile poter localizzare quali neuroni hanno generato i potenziali. Milioni di neuroni risiedono in quel volume, ed è inoltre difficile poterli distinguere a causa della loro eterogeneità.

Nell'analisi sulle singole cellule, per poter identificare l'integrazione multisensoriale, veniva utilizzata come criterio la massima risposta unisensoriale [2][3]. Ma nel caso delle popolazioni neurali, tale criterio non è ottimale, in quanto, si può facilmente superare la massima risposta unisensoriale, poiché la popolazione neurale potrebbe rispondere a due o più modalità sensoriali.

Un altro metodo utilizzato per indagare all'interno delle popolazioni neurali, è l'*fMRI (functional Magnetic Resonance Imaging)*. A differenza dell'EEG che ha un'ottima risoluzione temporale, ma una scarsa risoluzione spaziale, l'*fMRI* ha un'ottima risoluzione spaziale, ma una scarsa risoluzione temporale. La scarsa risoluzione spaziale è dovuta al fatto che in un singolo *voxel (pixel volumetrico)* sono contenuti decine se non centinaia di migliaia di neuroni. Per ovviare a tali differenze, si è raggiunto un compromesso che consiste nell'utilizzare l'EEG come strumento per

studiare quando avviene l'integrazione, mentre si utilizza l'fMRI per studiare dove avviene l'integrazione.

Nel caso dell'fMRI, le attivazioni del BOLD di un dato *voxel* vengono sommate in maniera lineare e tempo-invariante. Inoltre, un'ipotesi NULL conservativa equivale a sommare le risposte BOLD per gli stimoli unisensoriali, ciò corrisponde all'*additive criterion*. Uno dei limiti dell'fMRI è l'impossibilità di rilevare la superadditività, che può essere dovuto a diversi fattori.

In primo luogo, ciò può essere dovuto al fatto che i neuroni multisensoriali sono spesso alternati da consistenti popolazioni di neuroni unisensoriali. In secondo luogo, le proprietà delle risposte dei neuroni multisensoriali sono assai eterogenee, e includono neuroni che possono mostrare un miglioramento della risposta a seguito di una complessa stimolazione, altri possono mostrare una depressione della risposta, infine, altri possono invece mostrare nessuna interazione. A causa di ciò, la probabilità che un singolo *voxel* mostri una risposta superadditiva a uno stimolo multisensoriale è assai bassa.

Per questo motivo, quando viene utilizzato l'fMRI si fa riferimento al *maximum criterion*.

1.8 Comprendere l'attività neurale attraverso comportamento e percezione

Tutte le metriche illustrate finora per quantificare l'integrazione multisensoriale risultano essere affette da qualche errore, ma in ogni caso hanno fornito importanti informazioni per poter comprendere come avviene l'integrazione nel nostro cervello.

In ogni caso, se tali metodi vengono utilizzati singolarmente, essi forniscono solo una parziale comprensione di come avvengono i processi sensoriali. Per poterli comprendere appieno, bisogna integrare tali studi sull'attività neurale con le analisi percettive e comportamentali.

Le prime analisi neurometriche/psicometriche condotte sui modelli animali per registrare una prova comportamentale, venivano applicate solo per lo studio dei singoli sistemi sensoriali, con una lenta introduzione verso lo studio dei sistemi multisensoriali.

Di fatto, tali analisi possono mostrare la forte inferenza tra l'attività dei singoli neuroni (e delle popolazioni neurali) e i processi comportamentali [4].

Diversi studi hanno tentato di collegare la neurofisiologia al comportamento durante l'esecuzione di un compito multisensoriale. Uno di questi studi è stato condotto durante l'esecuzione di un compito multisensoriale, dove si voleva correlare l'attività dei neuroni del collicolo superiore (SC) con la misurazione del tempo dei movimenti dell'occhio saccadico rispetto all'obbiettivo.

Per poter comprendere meglio come esistano tali connessioni tra l'attività neurale e il comportamento, molte ricerche si sono concentrate sulla capacità di integrare i segnali.

Quando segnali provenienti da diverse modalità sono disponibili simultaneamente, la precisione comportamentale aumenta come affermato dai modelli, ottimizzati statisticamente, sull'integrazione dei segnali (modelli Bayesiani). Tali modelli sono utili per comprendere come i neuroni multisensoriali integrano i segnali nelle diverse regioni cerebrali, e inoltre sono importanti poiché possono colmare il *gap* tra le teorie sull'integrazione dei segnali e la fisiologia delle popolazioni dei neuroni multisensoriali.

In un sistema altamente semplificato, come ipotizzato da Fetsch [5], esistono due popolazioni di neuroni sensoriali primari, ognuna delle quali riceve informazioni unisensoriali da diverse modalità, che vengono poi trasmesse a una popolazione neurale multisensoriale, la quale genera in risposta una particolare scelta percettiva o comportamentale.

L'*output* di queste due aree unisensoriali converge sull'area multisensoriale con dei pesi sinaptici diversi. Tali pesi sinaptici sono dipendenti

dall'efficacia e/o dal numero di connessioni sinaptiche esistenti tra le singole aree unisensoriali e l'area multisensoriale.

Suddetta attività sinaptica può essere misurata attraverso le registrazioni del segnale LPF, il quale fornisce una diretta misura dei cambiamenti sinaptici sotto-soglia.

Le registrazioni dell'attività sinaptica extracellulare, riflettente l'attività sinaptica, fornisce l'*output* finale del calcolo della rete, con le velocità di scarica (*firing rate*) che variano quando vengono riprodotti o stimoli unisensoriali oppure stimoli multisensoriali combinati.

Da ciò, si può estrapolare la velocità di scarica associata alla condizione multisensoriali dalle velocità di scarica delle condizioni unisensoriali (ad esempio, attraverso una somma ponderata con differenti pesi neurali per le distinte condizioni unisensoriali).

In aggiunta, è possibile ricavare l'attività delle popolazioni neurali multisensoriali dai circuiti a valle, i quali generano una scelta percettiva o una risposta comportamentale. Il sistema ipotizzato da Fetsch, inoltre, mostra un'analogia con il modello che verrà descritto successivamente nel terzo capitolo.

Capitolo 2

Analisi dei processi sensoriali nei soggetti autistici

2.1 L'autismo attraverso l'effetto McGurk

L'autismo è un grave disturbo dello sviluppo neurologico che si manifesta entro i primi 30 mesi di vita del bambino. Solitamente i bambini, e quindi gli adulti, affetti da tale patologia, presentano delle caratteristiche comuni:

- compromissione della qualità dell'interazione sociale
- alterazione della qualità della comunicazione
- comportamento e interessi limitati, stereotipati e ripetitivi

A causa della varietà e della complessità dei disturbi, e a causa dell'impossibilità di fornire una diagnosi clinica unitaria, negli ultimi anni è stato introdotto un nuovo termine che accoglie tutti questi fattori.

Oggi, quindi, si parla di Disturbo dello Spettro Autistico(DSA).

Gli individui affetti da tale patologia, come anticipato prima, mostrano una grave alterazione delle capacità di comunicazione. La percezione del discorso è un aspetto particolare della comunicazione che può risultare

alterato nei soggetti autistici.

Per esempio, quando si vuole comprendere il discorso di una persona bisogna integrare una grande quantità di informazioni, quali ad esempio, la voce, i movimenti delle labbra e del viso e, a volte risultano utili, anche i movimenti del corpo. Per un soggetto autistico la capacità di integrare queste informazioni audiovisive è assai compromessa. Tali attitudini, diventano ancora più importanti quando riguardano i bambini autistici. Infatti, uno sviluppo appropriato di tali proprietà integrative può fortemente influenzare le capacità del bambino di comprendere il linguaggio. Di fatto, riuscire a capire il discorso della persona non riguarda non solo la comprensione delle parole, o dei movimenti articolari del viso, ma anche quelle informazioni che riguardano, l'intonazione, lo stato di eccitabilità, lo *stress*, l'emotività. Anche questo tipo di informazioni, che risultano alterate nei soggetti autistici, sono intuibili attraverso le due principali modalità, quella visiva e quella acustica.

Alcune ricerche hanno evidenziato che i soggetti con DSA hanno mostrato dei deficit anche nell'elaborazione dei processi unisensoriali, portando i ricercatori a chiedersi se tali deficit comunicativi possano essere causati da un difetto ancor più sostanziale dei processi di elaborazione multisensoriale [6].

Gli studi che vengono svolti per poter comprendere come avvengono tali processi multisensoriali e, più nel dettaglio, analizzare la percezione del linguaggio multisensoriale nei soggetti autistici, utilizzano un metodo di illusione audiovisiva, conosciuto con il nome di Effetto McGurk.

Tale metodo consiste in un *mismatching* degli stimoli acustici e visivi presentati insieme. Per meglio comprendere, l'illusione si basa sulla registrazione di un determinato fonema, ad esempio /ba/, che viene sovrapposto al video di una persona che con il labiale sta articolando un fonema incongruente con quello acustico, ad esempio /ga/. Ciò che viene però percepito è un fonema differente sia dal fonema acustico che dal visivo (fonema visivo), ossia una fusione di entrambi, ad esempio /da/.

Questa fusione multisensoriale fornisce un importante spunto per poter indagare su come vengono integrati tali tipi di segnali, mostrando come la modalità visiva influenza fortemente la modalità acustica.

Di fatto, l'effetto McGurk è altamente efficace nei bambini normotipici, ossia quegli individui le cui capacità neurologiche risultano intatte, mentre i bambini con DSA esibiscono una minore sensibilità a tale illusione.

2.2 Sensibilità al McGurk effect nei soggetti con DSA(studio di Taylor)

Lo studio di Taylor [7], effettuato nel 2010, si pone l'obiettivo di creare un collegamento tra l'età dei bambini con DSA e l'incremento della sensibilità all'illusione di McGurk. Egli suppose che i *deficit* dell'integrazione dei bambini autistici fossero dovuti a un ritardo dello sviluppo delle loro capacità, piuttosto che a un *deficit* fisso e permanente nel corso della loro vita.

Il suo studio venne condotto su un campione di 24 partecipanti autistici, con età compresa tra gli 8 e i 16 anni. Il compito da eseguire per i partecipanti, era semplicemente ripetere i fonemi che avevano ascoltato. Utilizzando l'analisi di regressione, veniva valutato se vi era stata una significativa traiettoria di sviluppo per una data capacità.

Quando i bambini ripetevano il fonema acustico, si riscontravano elevati indici di accuratezza in entrambi i gruppi. Le prestazioni dei bambini testati erano alte sia per i bambini piccoli che per gli adolescenti, ma la differenza sostanziale è che il gruppo con DSA mostrava un marginale miglioramento con l'età, mentre i ragazzi normotipici no.

Per quanto riguarda l'accuratezza visiva, ovvero quando veniva richiesto di osservare il labiale di chi stava loro parlando, entrambi i gruppi hanno mostrato un miglioramento con l'età, ma è interessante notare che i

bambini con DSA ottenevano risultati peggiori rispetto ai bambini normotipici.

L'analisi, però, non è stata condotta sulla base di un'interazione gruppo-età, quindi l'indice di sviluppo durante l'infanzia risultava equivalente per entrambi i gruppi.

Al contrario, mentre per i soggetti normotipici non è stato rilevato nessun cambiamento per la sensibilità all'effetto McGurk in funzione dell'età, per i bambini autistici, i quali inizialmente presentavano indici più bassi rispetto ai bambini normotipici, è stato evidenziato una progressione nella sensibilità all'effetto McGurk, in funzione dell'età.

Analizzando i dati, è risultato evidente che i bambini con DSA non raggiungono livelli di sensibilità all'effetto McGurk se non intorno ai 14-16 anni, mentre nei bambini normotipici tale sensibilità è rilevabile già intorno agli 8-10 anni.

Da questo studio è possibile comprendere come la possibilità di recuperare durante l'adolescenza per i bambini autistici rappresenta un elemento sfruttabile a fini terapeutici. Se i miglioramenti dei loro *deficit* multisensoriali avvengono anche senza intervenire, poter delineare una strategia di trattamenti per questi *deficit*, apporterebbe dei miglioramenti alle loro capacità comunicative, già nei primi anni di vita del bambino.

Una delle limitazioni dello studio di Taylor, è dovuta all'assenza di rumore. Né gli stimoli visivi, né quelli acustici, erano stati degradati o sporcati.

Di fatto, gli individui che soffrono di autismo hanno maggiori difficoltà nel comprendere un discorso in condizioni rumorose, e traggono minore beneficio nella lettura del labiale rispetto ai soggetti normotipici.

2.3 Integrazione audiovisiva in condizioni rumorose(studio di Foxe)

Nel 2013, il professor John J. Foxe e i suoi colleghi hanno condotto uno studio su un gruppo di bambini e adolescenti con DSA, al fine di valutare, in presenza di rumore, le loro capacità di integrazione multisensoriale [6]. Lo scopo della ricerca era quello di delineare la traiettoria di sviluppo delle capacità integrative multisensoriali, ovvero valutare se gli individui con DSA potevano mostrare un incremento delle loro abilità di integrare ciò che vedevano e udivano, e quindi riuscire a recuperare, rispetto ai soggetti normotipici, nel corso dell'adolescenza.

2.3.1 Metodo e Analisi

Alla ricerca hanno partecipato 226 ragazzi con età compresa fra i 5 e i 17 anni, di cui 84 ragazzi erano autistici e 142 erano normotipici, o anche detti TD(*typical-development*).

Age	TD					
	M_{age}	n_{TD}	n_{IQ}	VIQ	PIQ	FSIQ
5-6	6	10	6	107.83 (11.3)	104.33 (12.14)	106.17 (13.23)
7-9	8 (0.88)	38	28	114.75 (14.4)	107.86 (12.02)	113.04 (13.64)
10-12	11.37 (0.72)	43	27	111.67 (14.29)	106.93 (12.66)	110.33 (12.84)
13-15	14.14 (0.85)	31	28	111.89 (14.53)	106.93 (15.19)	110.25 (14.57)
16-17	16.68 (0.48)	20	19	108.47 (16.5)	102.1 (15.85)	105.84 (15.9)
\sum_n M		142	108	111.75 (14.56)	106.18 (13.67)	109.99 (14.06)

ASD						
M_{age}	n_{ASD}	n_{IQ}	VIQ	PIQ	FSIQ	
5.67	4	3	93.33	97	95	
8 (0.62)	25	22	97.05 (20.47)**	102.36 (18.05)	99.59 (18.59)**	
10.96 (0.81)	29	24	102.21 (20.96)	107.96 (18.89)	105.37 (19.91)	
13.45 (0.69)	21	11	99.64 (15.63)*	110.82 (9.78)	105.36 (12.94)	
16.75 (0.5)	5	4	86.25 (22.85)	102.5 (10.66)	94.25 (15.67)	
	84	64	98.59 (19.53)**	105.67 (16.78)	102.2 (17.66)**	

Figura 2.1: Suddivisione dei partecipanti

I bambini normotipici venivano esclusi se avevano disturbi educazionali, attenzionali, psichiatrici o se mostravano disturbo dell'iperattività; mentre dal gruppo dei bambini autistici sono stati esclusi quelli con problemi visivi e soggetti a crisi.

Questi due gruppi, vennero poi suddivisi, a loro volta, in 5 gruppi, come mostrato nella Figura 2.1.

Sono state utilizzate 300 registrazioni di parole monosillabiche pronunciate da una voce femminile. Sono stati scelti quei 300 monosillabi sulla base delle seguenti ipotesi: rappresentano monosillabi che appartengono al lessico di un bambino con età compresa nel *range* dell'esperimento, e inoltre sono parole che un bambino ascolta con alta probabilità ogni giorno.

Ad ogni registrazione video, della durata di 1.3 s, è stato sovrapposto il segnale acustico. Lo stimolo acustico si verifica dopo 520 ms dall'inizio del video.

Le parole venivano pronunciate in 7 diverse condizioni: 6 diverse condizioni di rumore rosa a 50, 53, 56, 59, 62 dBA di pressione sonora, e una condizione in assenza di rumore. Il rapporto segnale/rumore(SNR) è stato scelto in modo che il riconoscimento del fonema vari dallo 0% al 100%. Il *monitor* attraverso il quale si proiettavano le registrazioni, veniva posto ad una distanza di 80 cm dagli occhi dell'osservatore. La parte audio invece veniva fatta ascoltare attraverso delle cuffie.

L'esperimento consisteva in 3 condizioni poste in maniera *random*:

- condizione *A-alone*(*auditory-alone*), la parola veniva ascoltata in presenza di un fermoimmagine del viso dello *speaker*
- condizione *AV*(*audiovisual*), la parola veniva ascoltata assieme alla riproduzione video dell'interlocutrice che pronunciava la parola
- condizione *V-alone*(*visual-alone*), veniva mostrata la registrazione dello *speaker* che pronunciava la parola, senza audio.

Ogni parola veniva presentata per ogni condizione.

I partecipanti venivano istruiti a osservare lo schermo e riportare ciò che avevano udito, o visto nel caso *V-alone*. Un operatore aveva il compito di monitorare la situazione e garantire che i partecipanti mantenessero lo sguardo fisso sul *monitor*.

Le percentuali di risposta per ogni condizione, sono state analizzate attraverso un'analisi di varianza (RM-ANOVA), considerando la condizione di stimolo, il livello di SNR, le differenze tra i gruppi in base all'età. È bene precisare, che la condizione *V-alone* venne analizzata separatamente, poiché rappresentata in assenza di rumore.

L'analisi inoltre è stata condotta per 3 sottogruppi di età: 7-9, 10-12 e 13-15.

2.3.2 Effetti del SNR, delle articolazione dello *speaker* e dell'età sul riconoscimento delle parole

Come si può notare dalla Figura 2.2, la *performance* peggiora al diminuire del SNR(*signal noise ratio*), mentre si può notare un miglioramento quando è possibile osservare il labiale dell'interlocutrice. Si osserva una significativa interazione tra il tipo di stimolo e il rapporto segnale/rumore(SNR), constatabile dalla non uniformità dei valori di guadagno in condizione AV in funzione dei diversi valori di SNR.

Inoltre il fattore età è stato determinante, ciò è evidenziato dalla forte dipendenza tra l'età e il SNR, suggerendo una triplice dipendenza tra età, tipo di stimolo e SNR.

2.3.3 Differenze tra normotipici e autistici

Per confrontare meglio i risultati, hanno suddiviso i partecipanti in tre gruppi di età e riportato una media delle *performance* per la condizione *A-alone*, AV, e per i valori di guadagno AV per ogni livello di SNR, come è possibile vedere nelle Figura 2.2 e Figura 2.3.

2.3 Integrazione audiovisiva in condizioni rumorose (studio di Foxe)

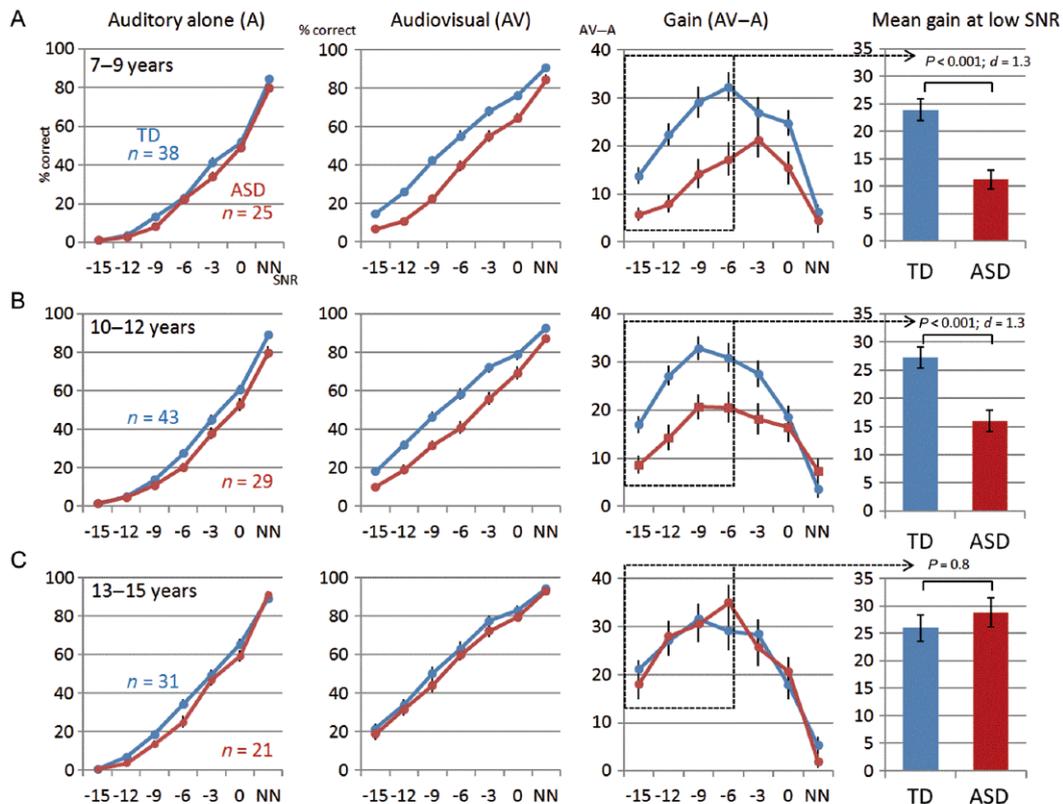


Figura 2.2: Performance di riconoscimento multisensoriali del parlato in funzione della diagnosi e dell'età.

È possibile notare come per entrambi i gruppi, le prestazioni nelle condizioni *A-alone* e AV sono risultate abbastanza alte, quando le parole venivano rappresentate in assenza di rumore e mostravano un miglioramento marginale con il progredire dell'età. Questo sta ad indicare che la difficoltà dei compiti era appropriata per entrambi i gruppi e per tutte le fasce di età.

Per quanto riguarda il gruppo dei bambini autistici, si sono riscontrate livelli di prestazione più bassi per la condizione *A-alone*, in tutti i gruppi d'età, con una differenza stimata del 4.23%.

Nella Figura 2.3, osservando il pannello A, si nota un miglioramento graduale nella prestazione *A-alone* per entrambi i gruppi, con un ritardo stimato di 2-3 anni per i bambini con DSA.

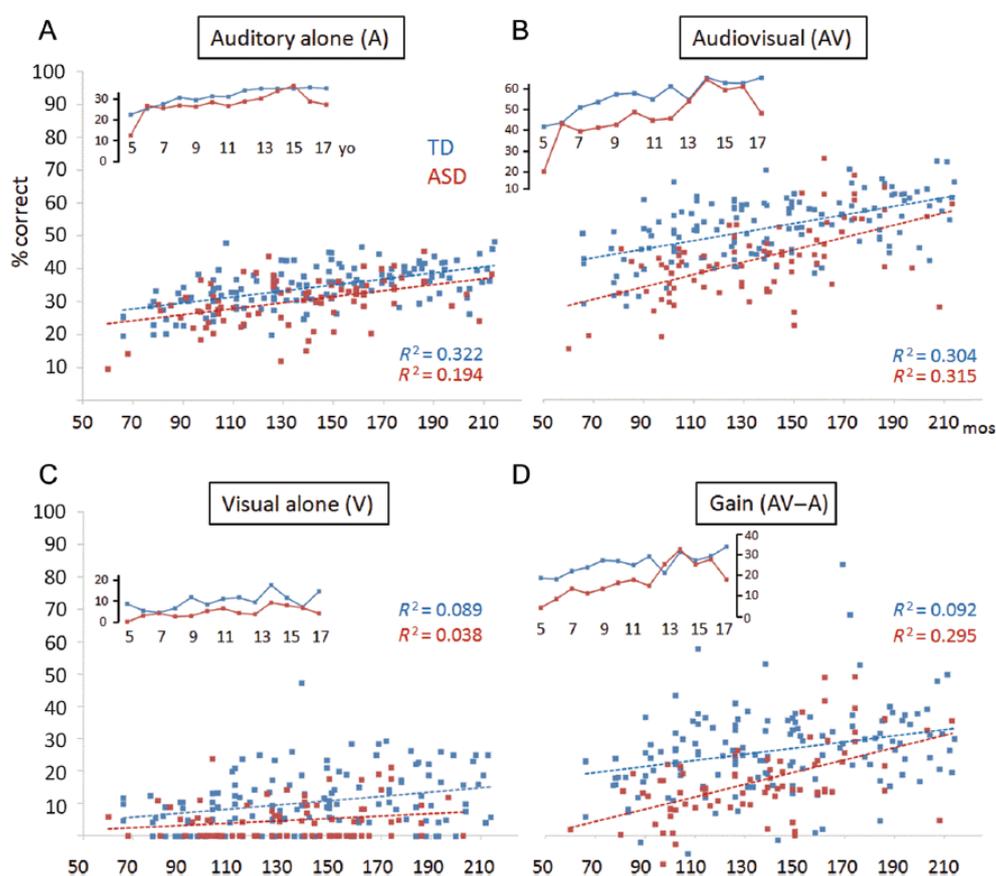


Figura 2.3: Andamento delle performance individuali dei partecipanti nelle condizioni *A-alone*, *V-alone*, AV e il guadagno AV (AV-A), in funzione dell'età espressa in mesi. Il guadagno AV è riferito ai livelli più bassi di SNR.

Inoltre, si vuole evidenziare le differenze tra i gruppi di 7-9 anni e 10-12 anni per quanto riguarda la condizione AV. In accordo a ciò, tali differenze si riscontrano anche nella condizione che delinea il guadagno AV (pannello D), sempre per i gruppi di 7-9 anni e 10-12 anni.

Se ci concentriamo sul pannello B, la *performance* dei bambini normotipici aumenta intensamente tra i 5 e gli 8 anni, e poi lentamente fino all'adolescenza. Al contrario, per i bambini autistici i miglioramenti si osservano fino ai 12 anni e la loro prestazione è simile a un bambino TD di 6-9 anni, ciò indica un ritardo di sviluppo nei bambini autistici.

Infine all'età di 13 anni, le differenze tra i due gruppi sono minori rispetto a prima, suggerendo un recupero, nella condizione AV, dei bambini autistici, riscontrabile anche nel grafico del guadagno AV.

Per comprendere invece i risultati della condizione *V-alone*, ovvero la lettura del labiale, faremo riferimento al pannello C.

È interessante notare che, degli 84 partecipanti con DSA, il 50% non è riuscito ad identificare le parole contro il 19.7% dei partecipanti normotipici, ovvero circa 28 bambini, mostrando l'incapacità dei bambini autistici di integrare gli stimoli visivi a qualsiasi età.

2.3.4 *Eye Tracking*

Durante l'analisi dei dati, gli autori hanno scelto di valutare, in percentuale, quanto i partecipanti mantenessero lo sguardo fisso. Ciò è stato fatto sempre per le tre fasce d'età. Non sono state rilevate differenze significative tra il gruppo con DSA e il gruppi dei partecipanti TD.

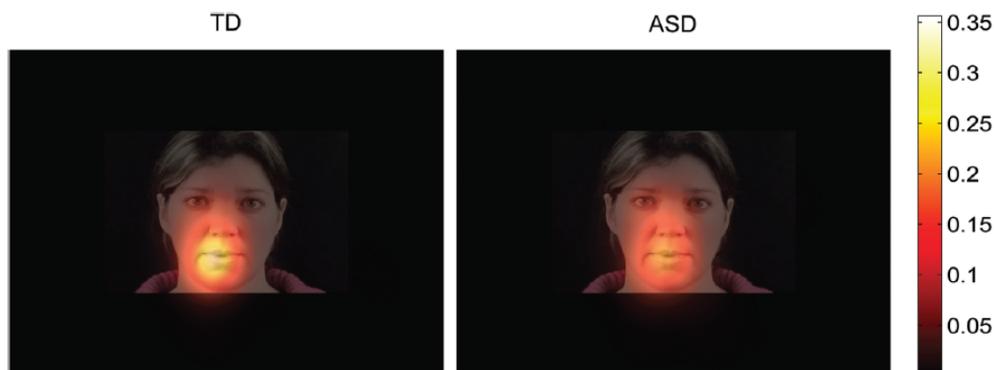


Figura 2.4: Mappa colorimetrica della fissazione dello sguardo durante *test V-alone*, ottenuta combinando i dati di tutte le età. I colori più brillanti indicano una maggiore consistenza della fissazione.

L'unico dato rilevante, riguardava la percentuale di fissazione dello sguardo sulla bocca, specialmente nei partecipanti di giovane età.

Per il gruppo di controllo, non hanno rilevato dati significativi, mentre

per i bambini autistici si è evidenziata una correlazione tra questo dato e il guadagno in SNR.

2.3.5 Discussione finale

Lo studio è stato svolto al fine di comprendere se i *deficit* delle abilità integrative degli stimoli visivi e acustici, in condizioni di rumore, nei bambini con DSA, potessero avere delle implicazioni per le strategie riabilitative.

I dati hanno evidenziato gravi *deficit* nella percezione multisensoriale del linguaggio, nei partecipanti di età scolastica primaria (dai 5 ai 12 anni). Inoltre è emerso che tali *deficit* non sono il semplice risultato di scarse abilità integrative unisensoriali. È possibile supporre che tali *deficit* siano presenti già nelle prime fasi di acquisizione del linguaggio, ancora più gravi nei primi anni dell'infanzia.

In aggiunta, i dati mostrano che tali *deficit* aumentano all'aumentare del rumore.

Tuttavia se si osservano i risultati dei *test* della condizione *A-alone*, i partecipanti con DSA mostrano una *performance* piuttosto simile alla loro controparte normotipica, e non sembrano essere influenzati dal rumore. Oltre a ciò, i risultati sottolineano le capacità di recupero dei soggetti autistici di età superiore ai 12 anni, esibendo delle capacità integrative multisensoriali quasi prive di *deficit*.

Vi sono diverse teorie che spiegano tale recupero delle abilità. Una di esse, riguarda la mielinizzazione delle connessioni sinaptiche, l'altra riguarda la pubertà e la nascita di un potenziale interesse sociale.

In ogni caso, tale studio ha dimostrato come le capacità integrative del linguaggio aumentano con il passare del tempo, mostrando che il sistema continua ad avere plasticità per i ragazzi normotipici, anche nell'adolescenza, facendo ben sperare in una intensificazione delle tardive capacità integrative multisensoriali nei soggetti con DSA utilizzando appropriate

strategie di pratica e motivazione.

2.4 Paradigma del ‘discorso nel rumore’ (studio di Smith e Bennetto)

Una delle sfide percettive quotidiane che un soggetto autistico deve affrontare, è la comprensione del discorso in un ambiente rumoroso.

Invece, gli individui normotipici hanno meno difficoltà a comprendere un discorso in presenza di rumore e, inoltre, utilizzano le informazioni visive per aumentare la comprensione del segnale acustico se vi è del rumore.

Un soggetto autistico, invece, ha difficoltà sia nell’integrare le informazioni acustiche e visive sia nel comprendere un discorso in presenza di rumore, accrescendo così la difficoltà di poter svolgere le normali azioni quotidiane.

Nel 2007, le docenti Elizabeth G. Smith e Loisa Bennetto [8], hanno condotto uno studio atto a comprendere se gli individui con DSA erano capaci di utilizzare le informazioni visive, per migliorare la comprensione dei segnali acustici, a cui era aggiunto del rumore.

2.4.1 Metodo e Analisi

Alla ricerca hanno partecipato 18 adolescenti con DSA e 19 adolescenti TD(*typical development*), quest’ultimi costituivano il gruppo campione. I partecipanti sono stati scelti in base all’età, al quoziente intellettivo, al genere, e in base a un test linguistico.

Sono stati esclusi dalla ricerca, soggetti autistici che presentavano traumi infantili dello sviluppo e con diagnosi di sindrome genetica. Per quanto riguarda i soggetti normotipici, sono stati esclusi tutti coloro che presentavano traumi infantili, disturbi psichiatrici, ritardi di linguaggio o ritardi mentali.

	Autism Group <i>M (SD)</i> [range]	Comparison Group <i>M (SD)</i> [range]
<i>n</i>	18	19
Age	15.84 (2.17) [12.42–19.50]	16.08 (2.04) [12.00–19.17]
CELF-4 RLI ^a	104.05 (8.72) [88–119]	104.26 (5.48) [96–117]
FSIQ ^b	108.1 (14.23) [77–129]	112.37 (8.81) [94–124]
Gender (M:F)	13:5	15:4

^aClinical Evaluation of Language Fundamentals, 4th Ed. Receptive Language Index.

^bFSIQ was measured with the WISC-IV or WAIS-III.

Figura 2.5: Descrizione delle caratteristiche dei gruppi dei soggetti autistici e dei soggetti campione.

Gli stimoli a cui venivano sottoposti i partecipanti, consistevano in corte frasi, solitamente di 5-7 parole, che contenevano 3 parole chiave.

Vennero utilizzate 48 frasi, pronunciate da cinque ragazze di età compresa fra i 23 e i 28 anni. Ad ogni ragazza venivano assegnate 9 o 10 frasi, che venivano registrate in una stanza insonorizzata con una videocamera professionale, dotata di microfono unidirezionale. Inoltre le ragazze avevano il collo e la gola coperte, poiché i movimenti di queste zone potevano favorire le percezioni audiovisive.

Il rumore è stato creato sovrapponendo le registrazioni di quattro ragazze che leggevano degli estratti da dei libri per bambini, filtrando i suoni articolatori, per rendere uniforme il grado di difficoltà.

I partecipanti venivano posti a 80 cm dal *monitor*. Le frasi venivano presentate per mezzo del *software DirectRT*, in tre diverse condizioni, con un ordine stabilito:

- condizione solo acustica
- condizione audiovisiva
- condizione solo visiva, ovvero lettura del labiale

Per le prime due condizioni hanno utilizzato il metodo di McLeod e Summerfield (1990) per ottenere la soglia di riconoscimento del discorso

(*speech reception threshold*, SRT), utile per stimare il rapporto segnale/rumore (*signal noise ratio*, SNR) per ogni condizione, di modo che un individuo era capace di individuare il 50% delle volte le 3 parole chiave.

2.4.2 Risultati a confronto

Condizioni solo Acustica e Audiovisiva

Tali analisi si basava sull’ipotesi che i soggetti autistici non avrebbero beneficiato dell’aggiunta dell’informazione visiva. Utilizzando un modello di analisi di multivarianza (MANOVA), gli autori hanno analizzato i valori di SRT dei due gruppi in funzione delle due condizioni, solo acustica e audiovisiva.

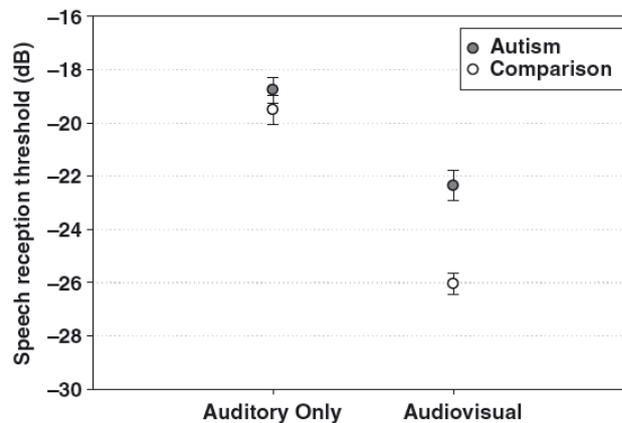


Figura 2.6: Soglia di ricezione media del discorso nelle condizioni solo Acustica e Audiovisiva. Valori più negativi di SRT indicano migliori prestazioni.

Dalla Figura 2.6, possiamo vedere che nel caso del solo stimolo acustico, sia i soggetti autistici che i normotipici presentano un valore di SRT intorno ai -19 dB. Quando al segnale acustico viene aggiunto quello visivo, quindi condizione Audiovisiva, si può notare come i valori di SRT per i soggetti TD cala a -26 dB, mentre per i soggetti con DSA la soglia scende di poco, intorno ai -22 dB.

Anche se il valore della soglia non è migliorato di molto come per i soggetti normotipici, rispetto alla condizione solo acustica, lo spostamento della soglia mostra comunque la capacità dei soggetti autistici di comprendere meglio le parole chiave quando possono osservare i movimenti delle labbra rispetto a quando possono solo ascoltare.

Letture del labiale

Per quanto riguarda la condizione solo visiva, ovvero la lettura del labiale, i risultati mostrano come i soggetti con DSA riescono a riconoscere molte meno parole rispetto alla loro controparte.

Infatti i soggetti autistici sono stati in grado di riconoscere solo il 14% delle parole attraverso la lettura del labiale con il 39% raggiunto dai soggetti normotipici.

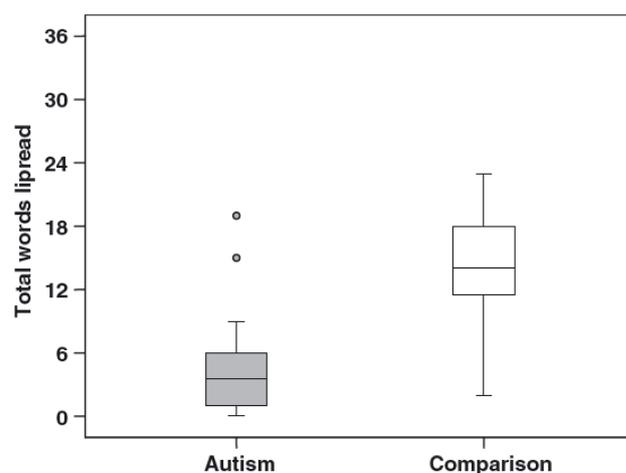


Figura 2.7: Soglia di ricezione media del discorso nelle condizioni solo Acustica e Audiovisiva. Valori più negativi di SRT indicano migliori prestazioni.

2.4.3 Discussione finale

I risultati di tale ricerca hanno dimostrato la presenza di *deficit* di integrazione audiovisiva in soggetti autistici. Aggiungendo le informazioni

visive a quelle acustiche, si è notato un miglioramento nella comprensione del linguaggio in presenza di rumore, anche se non è stato così elevato come per i soggetti normotipici. Inoltre si è scoperto che i soggetti autistici non sono abili a svolgere compiti che riguardano la lettura del labiale strettamente correlati con il paradigma audiovisivo.

Anche se si considera la varianza come causa unisensoriale, le differenze tra i due gruppi restano.

2.5 Risultati discordanti degli studi condotti in presenza di rumore

I due studi che analizzano le capacità di integrazione audiovisiva, in condizioni rumorose, di soggetti con DSA, sono quello di J.J. Foxe e quello di Smith e Bennetto.

I risultati di queste due ricerche sono però abbastanza discordanti. Se Foxe ha basato il proprio studio sull'ipotesi che le capacità integrative multisensoriali, fossero dipendenti dall'età, lo studio di Smith e Bennetto sembra contraddire questa ipotesi.

Lo studio di Foxe mostrava come le capacità dei soggetti autistici tendessero a migliorare, non appena si analizzavano gli adolescenti, con età superiore ai 12 anni, quasi normalizzandosi con le capacità dei soggetti normotipici.

La ricerca di Smith e Bennetto analizzò proprio ragazzi con età compresa fra i 12.4 e 19.5 anni.

Questo è proprio il *range* di età che, secondo Foxe, doveva mostrare dei miglioramenti, ma, al contrario, gli altri due ricercatori hanno dimostrato che, i soggetti con DSA, continuavano a mostrare dei *deficit* di integrazione audiovisiva.

Una delle possibili fonti di discrepanza tra i due studi, potrebbe riguardare l'analisi della fissazione dello sguardo dei partecipanti.

Lo studio di Foxe ha tenuto traccia dello sguardo dei partecipanti, evidenziando che i soggetti con DSA, generalmente, mostravano buone capacità di mantenimento dello sguardo durante l'esperimento. È inoltre utile far notare, che lo studio di Foxe è stato condotto utilizzando dei metodi di calibrazione e con l'uso di una telecamera per l'*eye-tracking* posta sul *monitor* di fronte ai partecipanti, che potrebbero averli influenzati.

Inoltre vicino ai soggetti che eseguivano l'esperimento, vi era un operatore che controllava se i partecipanti si distraessero.

Un altro aspetto da considerare, sempre nell'ambito della fissazione, riguarda il compito che i ragazzi dovevano svolgere. Nell'esperimento di Smith e Bennetto, i ragazzi dovevano cercare di individuare tre parole chiave all'interno di una frase di circa 5-7 parole, mentre il compito da svolgere, nello studio di Foxe, era quello di riconoscere delle parole monosillabiche. Per i partecipanti il tempo da dedicare attentamente al compito era molto più elevato nello studio della ricercatrice, per questo motivo potevano facilmente distrarsi.

La seconda differenza che bisogna considerare riguarda gli stimoli. Gli stimoli che venivano presentati nello studio di Smith e Bennetto potevano contenere informazioni contestuali, che potevano fornire un aiuto una volta che si era compresa una parola, rendendo quindi più facile l'individuazione delle altre parole chiave. Ciò che si può ipotizzare, è che i soggetti normotipici siano stati avvantaggiati da tali informazioni contestuali, e siano riusciti facilmente ad individuare le tre parole chiave rispetto ai soggetti autistici.

Infine, l'ultima differenza che possiamo analizzare riguarda il rumore.

Nello studio di Foxe viene utilizzato del rumore rosa, mentre nello studio di Smith e Bennetto il rumore che viene aggiunto, è stato creato registrando quattro ragazze che leggevano degli estratti da dei libri per bambini. Queste registrazioni venivano poi filtrate attraverso un filtro passa-basso per rimuovere i suoni articolatori. Si potrebbe sostenere che, anche se le registrazioni sono state filtrate, in ogni caso aggiungere

tale tipo di rumore potrebbe rappresentare una sfida più ardua per i soggetti con DSA.

In ogni caso, si può facilmente notare che i risultati sono discordanti tra loro.

Capitolo 3

Modello Matematico della Rete Neurale

3.1 Introduzione al modello

In base a quanto esposto nei capitoli precedenti, uno dei maggiori *deficit* mostrati dai bambini affetti da disturbo dello spettro autistico(DSA) riguarda la capacità di comprendere e comunicare con le persone che li circondano, soprattutto se i bambini si trovano in un ambiente rumoroso. Una delle ipotesi maggiormente sostenute, riguarda il ritardo di sviluppo dei meccanismi sottostanti l'integrazione multisensoriale per quanto riguarda i bambini autistici. Per poter analizzare tali tipi di problematiche, è stato sviluppato un modello computazionale [9] basato su una rete neurale. Nel modello vengono descritte le connessioni tra le due regioni unisensoriali, dette sinapsi cross-modali, e anche le connessioni, dette sinapsi *feedforward*, che si stabiliscono, a loro volta, tra le aree unimodali e l'area multisensoriale.

Attraverso questo modello viene simulata l'integrazione e l'elaborazione sensoriale, sia a livello unisensoriale che a livello multisensoriale.

A differenza del progetto precedente, nella seguente Tesi non è stato effettuato l'addestramento della rete, ipotizzando delle sinapsi già mature,

pronte a simulare le risposte dei soggetti normotipici e autistici.

In particolare, la rete è stata capace di riprodurre fin da subito l'effetto McGurk, nel caso normotipico.

In base agli studi precedenti [6], l'area neurale di particolare interesse per lo studio dell'elaborazione multisensoriale del linguaggio è il solco/giro temporale superiore (STG/S, *superior temporal gyrus/sulcus*), l'area corticale associativa adibita alla percezione del linguaggio. Inoltre, è importante sottolineare che l'integrazione multisensoriale avviene non solo nell'area STG/S, ma anche nelle prime fasi di elaborazione corticali dei segnali sensoriali, da cui si deduce che una certa multisensorialità è presente anche tra le regioni unisensoriali, visiva e uditiva, prima ancora che i segnali convergano sui neuroni dei centri di elaborazione dell'area multisensoriale.

Di fatto, l'integrazione multisensoriale relativa al linguaggio coinvolge i processi delle aree unimodali attraverso le sinapsi cross-modali e poi giunge nella regione multisensoriale attraverso le connessioni *feedforward*.

Attraverso questo modello, si vuole testare la relazione tra l'intensità delle sinapsi e la capacità di effettuare l'integrazione dei segnali, sia a livello unisensoriale che a livello multisensoriale, nei soggetti affetti da disturbi dello spettro autistico. La capacità integrativa viene testata attraverso l'illusione audiovisiva, meglio conosciuta come effetto McGurk.

Inoltre, con tale modello si vuole evidenziare come i *deficit* presenti nei bambini con età compresa fra gli 8-11 anni, e si riflettono nell'elaborazione multisensoriale senza avere ripercussioni nell'elaborazione unisensoriale, come verrà infatti mostrato nel capitolo inerente i risultati delle simulazioni.

3.2 Descrizione qualitativa del modello

Il modello consiste una regione multisensoriale, l'area STGS, che riceve proiezioni eccitatorie da due *array* di neuroni unisensoriali: il primo, è

dedicato alla rappresentazione dei segnali acustici (fonemi), il secondo è dedicato alla rappresentazioni dei segnali visivi (visemi) rappresentativi del linguaggio.

Le tre regioni sono costituite da N neuroni acustici, N visivi e N multisensoriali, dove N è impostato a 100 nelle simulazioni effettuate. Per comprendere meglio l'architettura della rete neurale è possibile osservare la Figura 3.1.

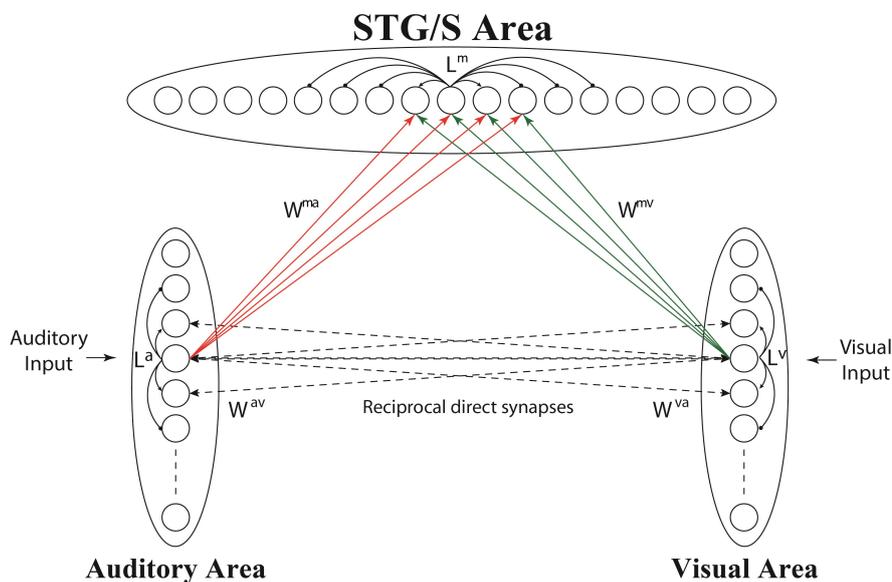


Figura 3.1: Architettura della rete neurale

Le risposte neurali, a qualsiasi ingresso, sono descritte da un'equazione differenziale del primo ordine, che simula le proprietà integrative della membrana cellulare, e una relazione sigmoide stazionaria, dalla quale si ottiene la soglia inferiore e quella di saturazione, che è impostata a 1. Le due unità visiva e acustica, presenti nel modello, rappresentano un insieme di neuroni che quando attivati insieme codificano per il suono e movimento facciale di un dato fonema, rispettivamente per ingressi acustici e visivi. I neuroni delle regioni unimodali sono organizzati topologicamente sulla base del principio di somiglianza, ovvero due suoni o due movimenti facciali simili attivano gruppi di neuroni prossimi in

queste aree.

L'organizzazione topologica dei neuroni può essere descritta utilizzando la disposizione del 'Cappello Messicano', come mostrato nella Figura 3.2: ogni neurone è connesso con gli altri della stessa area attraverso connessioni laterali eccitatorie e inibitorie, le unità prossimali si eccitano reciprocamente, mentre inibiscono quelle più lontane. Suddette connessioni intra-area, nella Figura 3.1, vengono identificate con i simboli L^a e L^v .

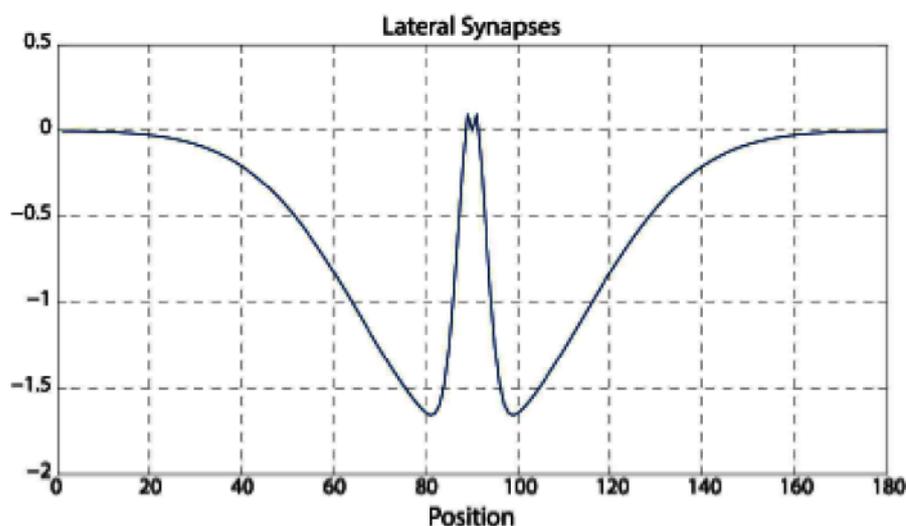


Figura 3.2: Sinapsi laterale secondo la disposizione del 'Cappello Messicano'. Le zone prossimali vengono eccitate, mentre quelle distali vengono inibite

Tale tipo di disposizione produce una 'bolla di attivazione' in risposta a uno specifico ingresso: si attiva non solo il neurone adibito per quella specifica caratteristica, ma anche quelli prossimali connessi attraverso una connessione sufficientemente eccitatoria.

I neuroni delle aree unisensoriali ricevono *input* esterni descritti attraverso una funzione Gaussiana, dove il punto centrale rappresenta lo stimolo specifico con la sua ampiezza, mentre la *deviazione standard* rappresenta l'incertezza dello stimolo. Per simulare l'esperimento in condizioni rumorose, questi *input* esterni sono accompagnati da una componente di rumore, ottenuto da una distribuzione uniforme.

Inoltre, come mostrato nella Figura 3.1, è stata considerata l'esistenza di interazioni cross-modali tra le aree unisensoriali. L'ingresso cross-modale viene calcolato assumendo che i neuroni delle due regioni unisensoriali sono reciprocamente connessi attraverso connessioni eccitatorie a lungo raggio, W^{va} e W^{av} .

La terza area neuronale presente nel modello, rappresenta i neuroni multisensoriali della regione corticale STG/S, coinvolta nel riconoscimento dei fonemi e nell'integrazione multisensoriale. I neuroni ricevono proiezioni eccitatorie dai neuroni delle regioni unisensoriali, che codificano per stimoli visivi e acustici simili e sono reciprocamente connessi attraverso connessioni laterali secondo la disposizione del 'Cappello Messicano', organizzati in base al principio di somiglianza. Suddette connessioni laterali vengono identificate con il simbolo L^m .

L'*output* dei neuroni dell'area STG/S viene comparato con una soglia fissata al 30% del massimo dell'attività dei neuroni; viene quindi calcolato il baricentro dell'attività soprasoglia, e valutato il fonema più vicino al baricentro stesso. Tale valore viene assunto uguale al fonema riconosciuto dalla rete in risposta a quella particolare stimolazione audio-visiva. Diremo che la rete effettua un corretto riconoscimento se il fonema che produce in uscita è uguale a quello fornito in ingresso (e corretto da rumore).

3.3 Descrizione matematica della rete neurale

Ogni neurone viene identificato con un apice n riferito alla specifica area corticale ($n=a$ per l'area acustica, v per l'area visiva e m per l'area multisensoriale), e con un pedice, j o k , il quale indica la posizione all'interno della regione ($j, k=1, 2, \dots, N$, con N pari a 100). I simboli $u(t)$ e $y(t)$ rappresentano, rispettivamente, l'ingresso e l'uscita di un dato neurone all'istante di tempo t . Quindi, $y_j^n(t)$ rappresenta l'uscita del neurone di posizione j , appartenente all'area n , descritta dalla seguente equazione

differenziale:

$$\tau \frac{dy_j^n(t)}{dt} = -y_j^n(t) + F(u_j^n(t)) \quad (3.1)$$

Dove τ è la costante di tempo del neurone e $F(u)$ rappresenta la relazione sigmoideale:

$$F(u_j^n) = \frac{1}{1 + e^{-s(u_j^n - \theta)}} \quad (3.2)$$

s e θ sono i parametri che stabiliscono la pendenza e la posizione centrale della relazione sigmoideale.

L'ingresso della rete, $u_j^n(t)$, che raggiunge un neurone, è la somma di due componenti: una componente interna alla rete ($l_j^n(t)$) dovuta al contributo delle connessioni laterali proveniente da altri neuroni della stessa area, e una componente proveniente da una sorgente esterna ($o_j^n(t)$).

Quindi, si ha:

$$u_j^n(t) = l_j^n(t) + o_j^n(t) \quad (3.3)$$

Per simulare l'ingresso laterale, $l_j^n(t)$, i neuroni di ogni area interagiscono attraverso le connessioni laterali eccitatorie e inibitorie: ogni neurone eccita, ed è eccitato a sua volta, dai neuroni prossimali, e inibisce, e viene inibito a sua volta, dai neuroni più distanti. L'input che un neurone unisensoriale riceve dagli altri neuroni della stessa area è definito così:

$$l_j^n(t) = \sum_k L_{jk}^n \cdot y_{jk}^n(t) \quad (3.4)$$

Dove L_{jk}^n è la forza della connessione laterale dal neurone presinaptico alla

posizione k al neurone postsinaptico alla posizione j entrambi appartenenti alla stessa regione n , e $y_k^n(t)$ è l'attività del neurone presinaptico alla posizione k . Queste connessioni sono simmetriche e sono disposte secondo la distribuzione del 'Cappello Messicano', una zona centrale eccitatoria circondata da un anello inibitorio:

$$L_{jk}^n = \begin{cases} L_{ex0} \cdot e^{-\frac{(d_{jk})^2}{2\sigma_{ex}^2}} - L_{in0} \cdot e^{-\frac{(d_{jk})^2}{2\sigma_{in}^2}} & \text{if } d_{jk} \neq 0 \\ 0 & \text{if } d_{jk} = 0 \end{cases} \quad (3.5)$$

Nell'equazione (3.5), L_{ex0} e σ_{ex} definiscono la funzione Gaussiana eccitatoria, mentre L_{in0} e σ_{in} descrivono quella inibitoria, e d_{jk} rappresenta la distanza tra il neurone presinaptico e quello postsinaptico della stessa regione. Per evitare gli effetti di bordo, le connessioni sono state realizzate in modo da ottenere una struttura circolare così che, ogni neurone di ogni regione, riceva lo stesso numero di connessioni laterali, secondo la seguente relazione:

$$d_{jk} = \begin{cases} |j - k| & \text{if } |j - k| \leq N/2 \\ N - |j - k| & \text{if } |j - k| > N/2 \end{cases} \quad (3.6)$$

Entrambe le regioni unisensoriali hanno lo stesso schema per la realizzazione delle connessioni laterali. Le connessioni laterali della regione multisensoriale sono state impostare per fornire il baricentro dell'attività evocata.

La componente esterna dell'ingresso ($o_j^n(t)$), rappresenta tutti i contributi provenienti da posizioni lontane. Tali input hanno una differente

espressione per le regioni unisensoriali ($n=a, v$) e per quella multisensoriale ($n=m$).

L'input di ogni area unisensoriale comprende uno stimolo sensoriale proveniente dal mondo esterno ($e_j^n(t)$), una componente cross-modale proveniente dall'altra regione unisensoriale ($c_j^n(t)$) e una componente di rumore (r) realizzato da una distribuzione standard uniforme con intervallo $[-r_{max} + r_{max}]$, dove r_{max} è uguale al 30% dello stimolo esterno più forte usato durante le simulazioni.

$$o_j^n(t) = e_j^n(t) + c_j^n(t) + r \quad n = a, v \quad (3.7)$$

Lo stimolo sensoriale esterno viene simulato utilizzando una funzione spaziale Gaussiana, per riprodurre l'incertezza degli stimoli esterni. Assumendo uno stimolo codificato in modalità n ($n=a$ o v) dal neurone alla posizione p^n , il conseguente ingresso della rete può essere così descritto

$$e_j^n(t) = E_0^n e^{-\frac{(d_j^n)^2}{2(\sigma_n)^2}} \quad (3.8)$$

Dove E_0^n rappresenta la forza dello stimolo, d_j^n è la distanza tra il neurone alla posizione j e lo stimolo alla posizione p^n , e σ_n descrive il grado di incertezza nella localizzazione sensoriale. E_0^n può variare da simulazione a simulazione fino a circa il 40% del suo valore, questo per simulare le differenze tra i soggetti, da esperimento a esperimento.

Per evitare anche qui gli effetti di bordo, gli ingressi esterno sono stati implementati secondo una struttura circolare; la distanza d_j è definita così:

$$d_j^n = \begin{cases} |j - p^n| & \text{if } |j - p^n| \leq N/2 \\ N - |j - p^n| & \text{if } |j - p^n| > N/2 \end{cases} \quad (3.9)$$

L'input cross-modale, $c_j^n(t)$, è stato definito in modo tale che ogni neurone riceve un'eccitazione dai neuroni delle altre modalità in tal modo:

$$c_j^a(t) = \sum_{k=1}^N W_{jk}^{av} \cdot y_{jk}^v(t - \Delta t^c) \quad (3.10a)$$

$$c_j^v(t) = \sum_{k=1}^N W_{jk}^{va} \cdot y_{jk}^a(t - \Delta t^c) \quad (3.10b)$$

Il parametro Δt^c rappresenta il ritardo con il quale gli ingressi cross-modali vengono scambiati tra le due regioni.

L'input eccitatorio esterno del neurone multisensoriale è dovuto solo alle connessioni *feedforward* delle due regioni unisensoriali. Ovvero

$$i_j^m(t) = \sum_{k=1}^N W_{jk}^{ma} \cdot y_{jk}^a(t - \Delta t^m) + \sum_{k=1}^N W_{jk}^{mv} \cdot y_{jk}^v(t - \Delta t^m) \quad (3.11)$$

Dove W_{jk}^{ma} e W_{jk}^{mv} , identificano le connessioni che collegano il neurone presinaptico alla posizione k nell'area unisensoriale, acustica e visiva, al j -esimo neurone nell'area multisensoriale. In tale equazione, il parametro Δt^m è il ritardo con il quale l'area multisensoriale riceve gli *input* dalle regioni unisensoriali.

Capitolo 4

Risultati

Sulla base degli studi presentati nei capitoli precedenti [7][6][8], abbiamo visto che, sia i bambini normotipici che quelli autistici, hanno buone capacità nel comprendere gli stimoli acustici quando presentati da soli, anche in condizioni rumorose. Le differenze tra i due soggetti si intensificano nel momento in cui lo stimolo acustico è accompagnato da uno stimolo visivo. Per i bambini normotipici lo stimolo visivo tende a migliorare la percezione delle informazioni quando è accoppiato ad un segnale acustico; per un bambino autistico, i due segnali accoppiati non apportano grandi miglioramenti nella comprensione delle informazioni.

In questo capitolo, vengono mostrati i risultati ottenuti facendo simulare, alla nostra rete neurale, il comportamento sia di un bambino normotipico, sia di un bambino autistico, quando ad essi vengono presentati degli stimoli in condizioni rumorose, simulando una delle illusioni audiovisive più utilizzate per comprendere l'integrazione multisensoriale. Tale illusione prende il nome di effetto McGurk, già introdotto nel secondo capitolo.

Gli stimoli presentati in ingresso alla rete, sono stati forniti in tre condizioni: solo Acustica(A), solo Visiva(V) e Audiovisiva(AV). Inoltre, è importante sottolineare che lo stimolo riprodotto nelle prime due condizioni non rappresentava lo stesso, fonema/visema: infatti nella terza condizione gli stimoli venivano presentati in maniera disallineata, con

una distanza di 4 posizioni, proprio per vedere come la rete risponde all'effetto McGurk.

Per ognuna di queste condizioni, è stata valutata la risposta dell'area multisensoriale, come possiamo vedere nella Figura 4.1.

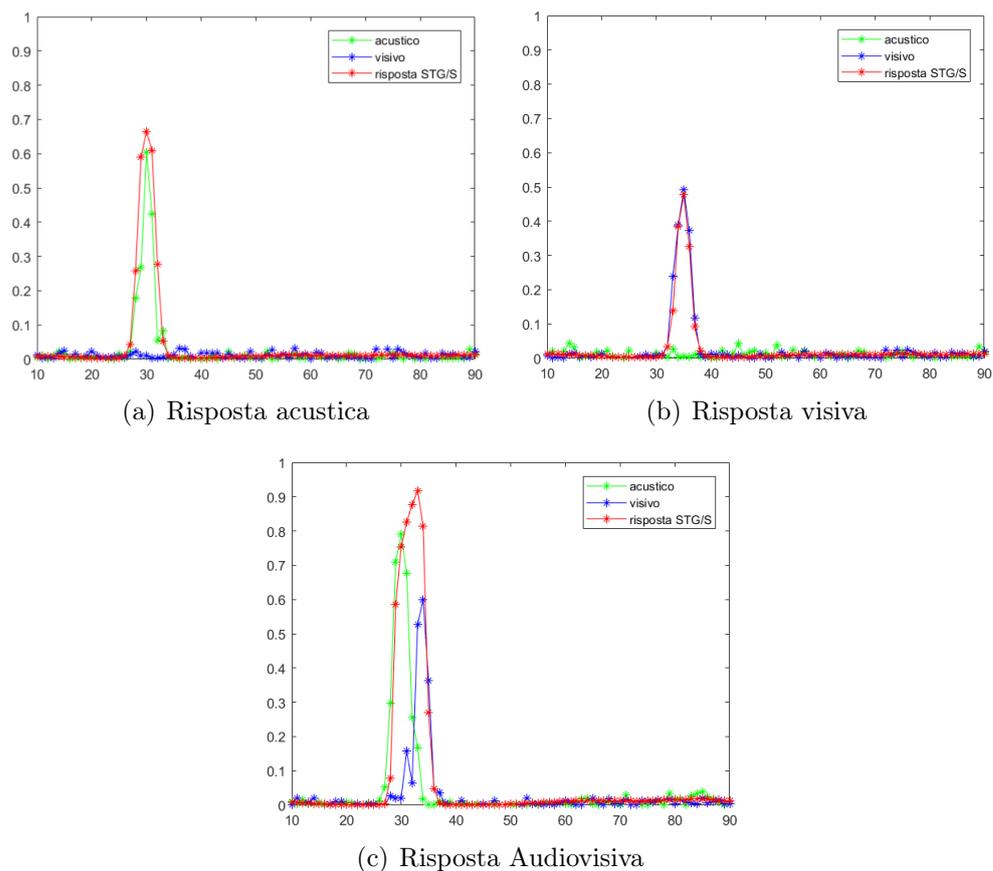


Figura 4.1: Risposta dell'area multisensoriale STG/S per ognuna delle tre condizioni

4.1 Scelta del caso d'interesse

Per poter analizzare come le prestazioni di un bambino autistico, sottoposto all'effetto McGurk, cambiano a seconda dell'intensità delle connessioni sinaptiche, abbiamo cercato di riprodurre attraverso la rete neurale,

il caso di un bambino normotipico sottoposto a diversi stimoli, cercando di identificare anche un buon valore di rumore.

Per prima cosa, abbiamo scelto tre possibili valori del rapporto segnale/rumore (*signal noise ratio*, SNR): 10%, 20%, 30%.

Per ognuna di queste simulazioni, abbiamo assegnato un diverso valore per l'intensità dei due stimoli in ingresso e successivamente, effettuato ulteriori simulazioni, assegnando lo stesso valore all'intensità degli *input*, registrando la risposta dell'area multisensoriale nelle tre condizioni.

Per quanto riguarda gli stimoli, su una base di 100 fonemi, abbiamo identificato una finestra di 60 fonemi da far riprodurre alla rete.

Successivamente, i dati raccolti per ogni caso, venivano riportati su istogramma per poter valutare la risposta multisensoriale rispetto al segnale acustico. Di fatto, per poter comprendere se la rete era sensibile o meno all'effetto McGurk, bisogna vedere di quanto la risposta multisensoriale si discosta dal fonema acustico.

In ascissa sono presenti i valori 0, 1, 2, 3 e 4, che rappresentano la distanza fra il fonema acustico dato in *input* e il fonema riconosciuto dallo strato multisensoriale. Si ricordi che 4 rappresenta la posizione dello stimolo visivo. In ordinata vi è la scala percentuale dei riconoscimenti.

Abbiamo considerato presente l'effetto McGurk nel momento in cui la risposta multisensoriale si discostava di due o più posizioni dal segnale acustico.

Le prime simulazioni sono state condotte con un valore di SNR pari al 10%. Si è subito notato che la rete rispondeva in maniera troppo precisa, identificando spesso il fonema sia acustico, che visivo.

Quindi si è deciso di scartare tale caso d'indagine poiché il rumore risultava quasi ininfluenza.

Successivamente l'attenzione si è spostata verso gli altri due casi contemplati.

Le simulazioni condotte riguardano i seguenti casi:

- A. Caso SNR=20%, intensità stimolo acustico pari a 23, intensità stimolo visivo pari a 18
- B. Caso SNR=20%, intensità dello stimolo acustico uguale a quello visivo, con un valore pari a 21
- C. Caso SNR=30%, intensità stimolo acustico pari a 23, intensità stimolo visivo pari a 19
- D. Caso SNR=30%, intensità dello stimolo acustico uguale a quello visivo, con un valore pari a 21

Nel caso **A**, la rete per il 59.1% è stata in grado di riconoscere correttamente il fonema visivo, quando questo veniva fornito da solo, mentre il 100% delle volte ha identificato il fonema acustico.

Quando alla rete veniva presentato lo stimolo multisensoriale, la rete multisensoriale rispondeva nel 9.8% dei casi con un effetto McGurk.

Nel caso **B**, la rete ha identificato il fonema visivo per il 96.72% delle volte, e quello acustico il 98.4% delle volte. Nel momento in cui venivano forniti in ingresso entrambi gli stimoli, sempre disallineati, ossia non identificanti lo stesso fonema, la rete produceva una alta percentuale di effetto McGurk, circa il 77.05%.

In seguito si è passati alle simulazioni con il rapporto segnale rumore pari al 30%.

Nel caso **C** si è scelto di aumentare di un punto il valore dell'intensità dello stimolo visivo rispetto al caso dove il rapporto segnale/rumore era del 20%.

Per il 68.9% la rete è riuscita ad identificare il segnale visivo nella condizione solo visiva, e per l'88.5% è riuscita ad identificare quello acustico. Quando poi l'ingresso da unisensoriale diventava multisensoriale, la rete ha mostrato una sensibilità all'effetto McGurk del 36.07%.

Infine nel caso **D**, dove l'intensità degli stimoli è uguale, la rete è riuscita ad identificare circa l'80% delle volte sia il segnale visivo che quello

acustico, e per la sensibilità all'effetto McGurk, la percentuale è assai elevata, circa del 72%.

Argomentazione dei risultati

In base ai risultati pervenuti a seguito di queste prove, ci siamo chiesti quale fosse il caso che potesse meglio simulare le risposte di un bambino normotipico, quindi quale dei 4 casi simili al meglio una condizione, quanto più possibile, simile alla realtà. Per poter scegliere quale caso fosse il più performante, abbiamo osservato tre parametri: percentuale di riconoscimento del segnale acustico nella condizione solo acustica, percentuale di riconoscimento del segnale visivo nella condizione solo visiva e infine percentuale di sensibilità all'effetto McGurk nella condizione Audiovisiva.

Per poter scegliere il caso migliore, ci siamo basati anche sull'ipotesi che il riconoscimento visivo sia più debole nei bambini autistici, rispetto a quello acustico, mentre nei bambini normotipici il segnale visivo se fornito in accoppiata a quello acustico, aumenta la prestazione.

Se confrontiamo il caso **B** con il caso **D**, dove i segnali acustico e visivo hanno pari intensità, possiamo vedere che le percentuali di riconoscimento dei singoli fonemi sono assai elevate.

Inoltre, essendo in condizioni rumorose, la capacità di percepire correttamente il fonema visivo dovrebbe risultare più bassa, quindi si è deciso di scartare tali due casi.

In seguito si è passato ad analizzare i casi **A** e **C**. Questi due casi potrebbero sembrare entrambi ottimali, ma vi sono delle differenze.

Nel caso **A** la percentuale di riconoscimento del segnale acustico è troppo elevata, poco simile alla realtà.

Di fatto, quando siamo in un ambiente rumoroso, riuscire a comprendere ciò che ci viene detto risulta più difficoltoso, e una percentuale pari al 100% non sembra rappresentare una condizione reale.

Questo ci porta verso caso **C**, dove l'intensità degli stimoli è diversa e

il rumore si aggira intorno al 30%. Infatti in questo caso, la rete ha identificato correttamente solo l'88.5% dei fonemi acustici, e il 68.9% dei visemi(fonemi visivi).

Per poter confermare tale scelta possiamo basarci sul terzo parametro, ovvero la sensibilità all'effetto McGurk. Nella Figura 4.2 sono presenti i 4 istogrammi corrispondenti ai 4 casi simulati.

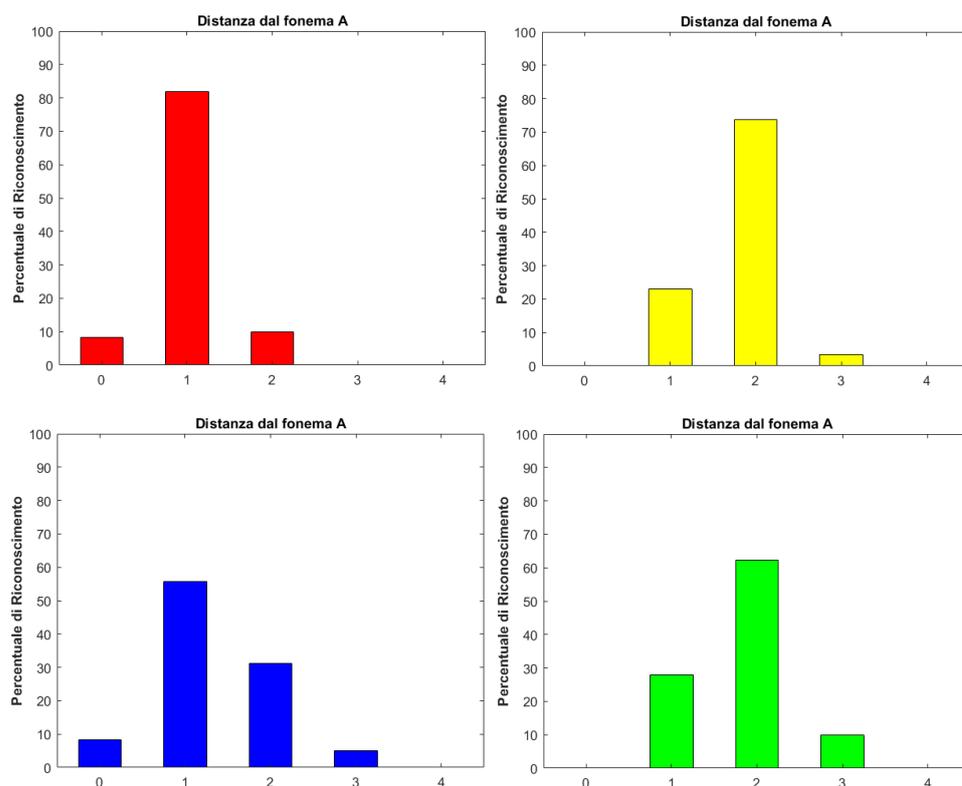


Figura 4.2: Istogrammi raffiguranti la risposta multisensoriale STG/S rispetto al segnale acustico, in percentuale, nei 4 casi(A rosso, B giallo, C blu, D verde)

Gli istogrammi mostrano la risposta multisensoriale dell'area STG/S (*superior temporal gyrus/sulcus*), che è l'area corticale associativa coinvolta nella percezione del linguaggio, valutata rispetto al segnale acustico, in percentuale, nei 4 casi sopra nominati. Quando la risposta multisensoriale si discosta di due posizioni rispetto allo stimolo acustico, viene

riconosciuto l'effetto McGurk, quando invece la risposta riproduce esattamente lo stimolo acustico, o si discosta di una sola posizione, non si è in presenza del verificarsi dell'illusione audiovisiva.

Nei casi **B** e **D**, corrispondenti agli istogrammi sulla destra, possiamo vedere una elevata sensibilità all'effetto McGurk. Tali casi sono stati scartati anche in base a tale condizione, poiché in condizioni rumorose la sensibilità all'effetto McGurk dovrebbe risultare minore anche per un bambino normotipico.

Per i casi **A** e **C**, il fattore discriminante, è stato proprio questo. Infatti come si può vedere dal pannello in basso a sinistra, raffigurante il caso **C**, la percentuale di sensibilità all'effetto McGurk si aggira intorno al 36%. Tale dato è infatti concorde con le ipotesi di riprodurre le risposte di un bambino normotipico, quando sottoposto a una stimolazione multisensoriale in presenza di rumore.

Quindi è stato scelto proprio tale caso, per poter poi confrontare e comprendere come varia la risposta di un bambino autistico a seconda dell'intensità delle connessioni sinaptiche, rispetto ad un bambino normotipico.

4.2 Verso la simulazione del soggetto con DSA

Dopo aver simulato la risposta di un bambino normotipico, abbiamo cercato di riprodurre la risposta di un bambino autistico, sempre attraverso l'analisi della sensibilità all'effetto McGurk.

L'ipotesi di partenza è la seguente: un bambino normotipico, con età compresa fra i 9-11 anni, risponde all'illusione audiovisiva con una percentuale che si aggira intorno a quella mostrata dal caso d'interesse simulato attraverso la rete neurale. Per un bambino autistico, tale percentuale tende a diminuire rispetto alla sua controparte normotipica, a causa di diversi fattori, i quali potrebbero essere dovuti ad una scarsa attenzione del bambino, in questo caso si parla di *attentional bias*, o all'indebolimento delle connessioni sinaptiche, denominato *architectural bias*.

In questo progetto di Tesi, l'ipotesi principale dei ritardi relativi all'integrazione multisensoriale si basano sul *bias architectural*, e per poter riprodurre la risposta di un soggetto con DSA, abbiamo considerato due diverse ipotesi.

Nella prima, abbiamo danneggiato le connessioni sinaptiche dette cross-modali, che nel modello sono state identificate con i simboli W^{va} e W^{av} , ovvero quelle connessioni che si stabiliscono tra le aree unisensoriali acustica e visiva, lasciando invariate le connessioni *feedforward*.

Come seconda ipotesi, abbiamo invece danneggiato le connessioni che si stabiliscono tra le aree unisensoriali e l'area multisensoriale, le sopraccitate sinapsi *feedforward*. I simboli utilizzati per identificare tali connessioni sono W^{ma} e W^{mv} . È utile sottolineare che in questa seconda ipotesi, le connessioni cross-modali sono state lasciate intatte.

In entrambi i casi, il processo di danneggiamento è stato eseguito indebolendo le sinapsi attraverso un fattore moltiplicativo, come è stato descritto nel capitolo 3.

4.2.1 Danneggiamento connessioni cross-modali

Utilizzando il caso d'interesse che simula le risposte di un bambino normotipico, abbiamo modificato per prima cosa le connessioni cross-modali, riducendolo una volta al 70% e una volta all'80%.

Ora per poter comprendere come la rete simuli la risposta all'effetto McGurk a seguito del danneggiamento delle sinapsi, abbiamo anche in questo caso riportato su un istogramma le risposte dell'area multisensoriale a uno stimolo audio-visivo, rispetto al fonema acustico, come mostrato in Figura 4.3, dove per chiarezza è riportata anche il del caso normotipico.

Abbiamo valutato anche in queste simulazioni le percentuali di riconoscimento dei fonemi, interrogando la rete nelle tre condizioni: solo acustica, solo visiva e audiovisiva.

Nella simulazione con le sinapsi cross-modali al 70% della loro intensità,

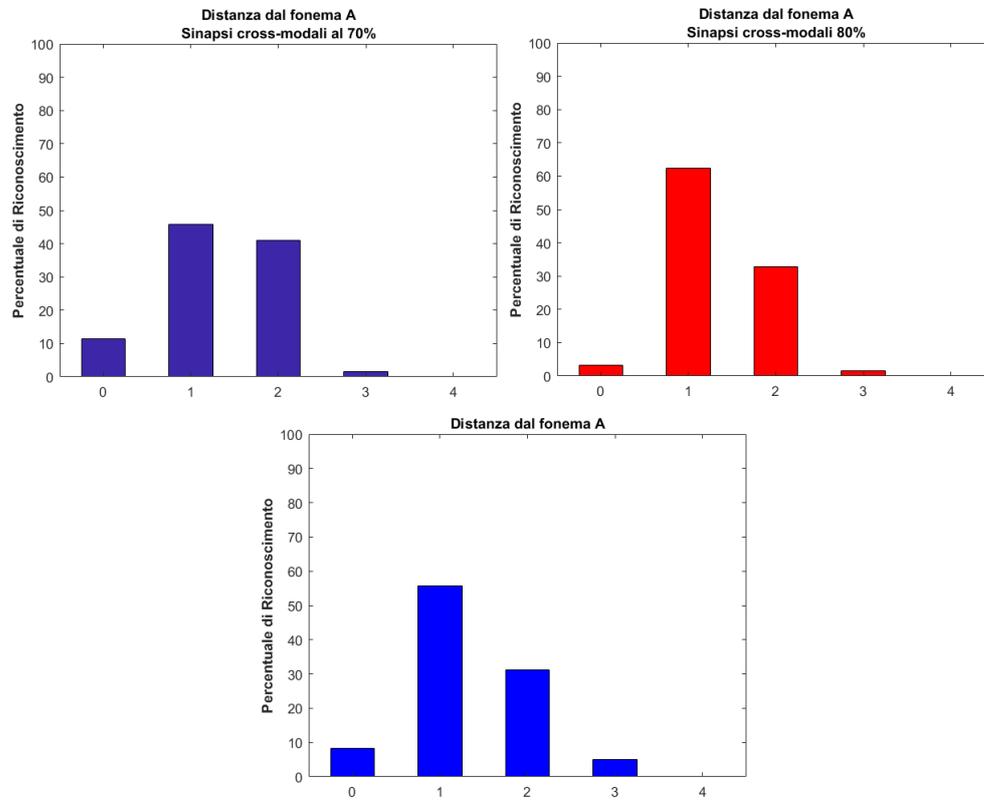


Figura 4.3: Risposta dell'area multisensoriale, in percentuale, valutata rispetto al fonema acustico. A sinistra le sinapsi cross-modali sono danneggiate del 30%, a destra del 20%. Il terzo grafico mostra il caso normotipico

abbiamo osservato come la rete rispondesse anche nelle condizioni unisensoriali.

La rete è stata in grado di riconoscere correttamente per l'86.88% delle volte il fonema acustico nella condizione solo acustica. Per quanto riguarda invece il fonema visivo nella condizione solo visiva, la rete ha riconosciuto correttamente il visema nel 70.5% dei casi.

Per quanto riguarda invece il caso in cui le sinapsi sono danneggiate del 20%, il fonema acustico è stato riconosciuto correttamente il 90% delle volte, mentre quello visivo per il 68.9% delle volte.

Nel sottoparagrafo 4.2.3 viene valutata la risposta multisensoriale. La Figura 4.5 mostra come si modifichi la sensibilità all'effetto McGurk a

seguito del danneggiamento delle sinapsi cross-modali. In ascissa sono riportati i valori attribuiti alle sinapsi cross-modali, dove il valore 1 rappresenta il caso del bambino normotipico.

4.2.2 Danneggiamento sinapsi *feedforward*

Le sinapsi *feedforward*, o anche dette connessioni in avanti, rappresentano le connessioni eccitatorie a lungo raggio che si stabiliscono fra ognuna delle aree unisensoriali, acustica e visiva, e l'area multisensoriale.

Anche in questo caso abbiamo danneggiato tali connessioni sinaptiche per poter comprendere quale fosse il caso migliore per poter estrapolare in fine la risposta di un bambino autistico.

Come per il caso delle sinapsi cross-modali, anche per le sinapsi in avanti abbiamo indebolito le sinapsi del 20% e del 30%.

Abbiamo interrogato la rete, fornendo gli stimoli nelle tre condizioni: solo acustica, solo visiva e audiovisiva.

Valutando le risposte della rete agli ingressi unisensoriali, possiamo vedere che nel caso in cui le sinapsi in avanti sono state indebolite del 30%, la percentuale di riconoscimento del fonema acustico è dell'86.88%, mentre la percentuale di riconoscimento del fonema visiva nella condizione solo visiva è del 26.23%.

Per quanto riguarda la risposta della rete all'ingresso di uno stimolo multisensoriale, abbiamo graficato la risposta attraverso l'ausilio dell'istogramma, come possiamo vedere nella Figura 4.4, in cui è illustrato l'istogramma della simulazione con le sinapsi *feedforward* all'80% delle loro potenzialità e anche il caso normotipico.

Per la simulazione con le sinapsi *feedforward* danneggiate al 20%, le percentuali di riconoscimento dei fonemi acustici e visivi, valutati nelle condizioni unisensoriali sono rispettivamente 78.7% e 50.82%.

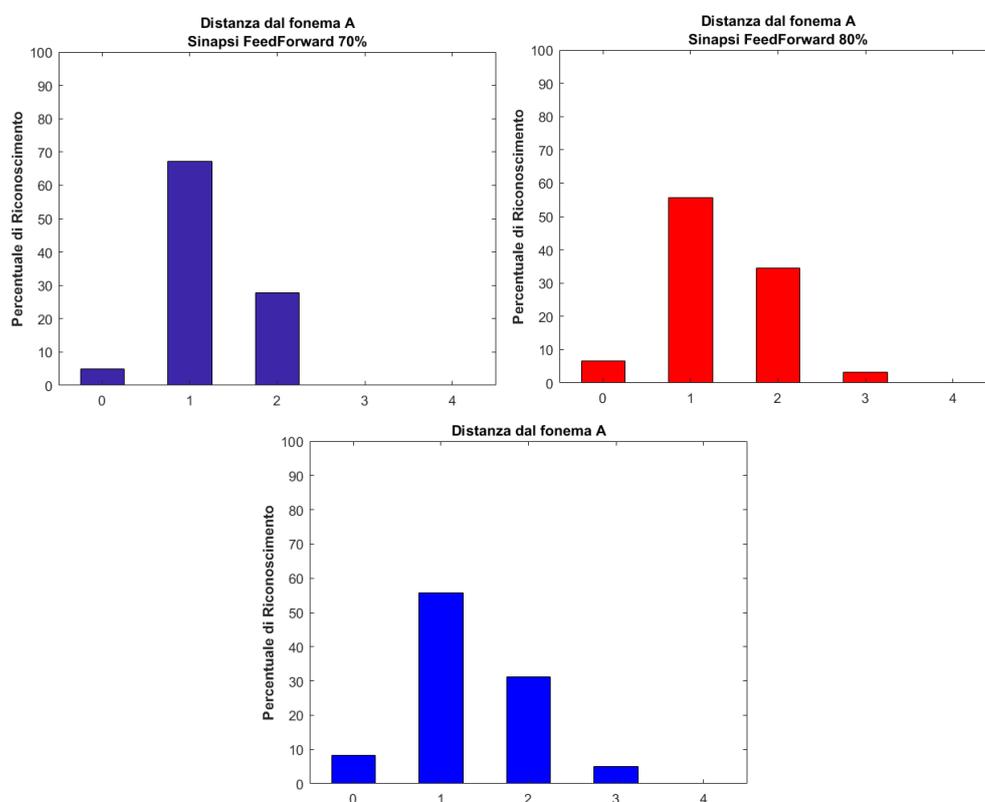


Figura 4.4: Risposta dell'area multisensoriale, in percentuale, valutata rispetto al fonema acustico. A sinistra le sinapsi *feedforward* sono danneggiate del 30%, a destra del 20%. Il terzo grafico mostra il caso normotipico

4.2.3 Confronto fra le due ipotesi

Bisogna notare che, per quanto riguarda il riconoscimento dei singoli fonemi, se osserviamo i risultati ottenuti con il danneggiamento delle sinapsi cross-modali vediamo che le percentuali di riconoscimento sono molto vicine a quelle ottenute con il caso del bambino normotipico. Già con questi risultati si può notare come l'ipotesi di danneggiare solo le aree unisensoriale non mostra un'enorme differenza con il caso del bambino normotipico. Potremmo dedurre, già con questi solo risultati, che la differenza tra i due soggetti risieda proprio nelle connessioni in avanti, ossia quelle connessioni eccitatorie che si stabiliscono tra le aree unisensoriali e l'area associativa multisensoriale. Di fatto, il bambino autistico non

mostra profondi *deficit* quando viene interrogato per la percezione delle informazioni unisensoriali, ma i *deficit* si accentuano nel momento in cui i compiti diventano più complessi, in particolar modo quando il compito che gli viene richiesto è posto in una condizione rumorosa.

Se invece osserviamo le percentuali di riconoscimento ottenute indebolendo le connessioni *feedforward*, possiamo notare la stessa correlazione tra il danneggiamento della sinapsi e il corretto riconoscimento dei fonemi. Infatti, concentrandoci sulle percentuali di riconoscimento del visema, è importante sottolineare che più le connessioni sono danneggiate, più la rete ha mostrato difficoltà nel riconoscere il fonema visivo. Tanto è vero che, nel caso normotipico, la percentuale di riconoscimento del visema è del 68.9%, mentre invece quando la sinapsi è danneggiata al 20% vediamo che tale percentuale cala al 50.82%, e ancor di più si abbassa nel momento in cui le sinapsi sono state danneggiate del 30%.

La percentuale infatti è scesa al 26.23%. Infine, per poter scegliere de-

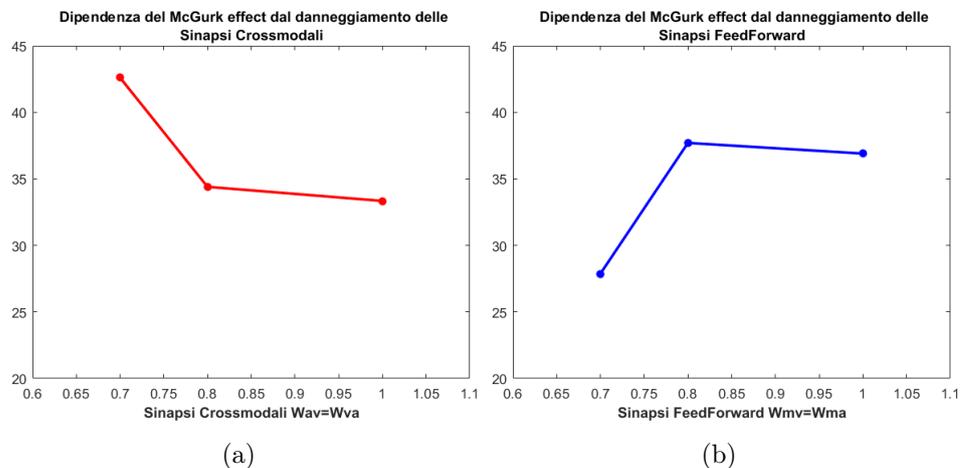


Figura 4.5: Andamento della sensibilità all'effetto McGurk a seguito del danneggiamento delle sinapsi cross-modali, a sinistra, e delle sinapsi *feedforward*, a destra

finitivamente quale delle due ipotesi si avvicina il più possibile alla simulazione della risposta di un bambino affetto da disturbi dello spettro autistico, abbiamo costruito un grafico dove viene mostrata la dipendenza

tra la sensibilità all'effetto McGurk e la risposta dell'area multisensoriale all'ingresso multisensoriale.

Come anticipato, tale grafico è mostrato nella Figura 4.5. Esso è stato realizzato per entrambe le ipotesi e possiamo vedere in tale figura come l'andamento sia l'uno l'inverso dell'altro.

Nella Figura 4.5(a), la sensibilità all'effetto McGurk tende ad aumentare al diminuire dell'intensità delle sinapsi cross-modali: questo perché, avendo indebolito le connessioni tra le aree unisensoriali, il segnale acustico attira di meno verso di sé il segnale visivo, il che si traduce in un baricentro posizionato in posizione intermedia fra fonema e visema, e quindi in un aumento della sensibilità all'effetto McGurk.

Però, ciò che noi vogliamo ottenere attraverso questo progetto di Tesi, è la dipendenza tra la risposta di un soggetto autistico e la sensibilità all'effetto McGurk.

La sensibilità all'illusione audiovisiva tende a diminuire nel soggetto autistico, rispetto ad un soggetto normotipico, a causa della ridotta capacità di riuscire a trarre vantaggio dall'integrazione del segnale visivo in aggiunta a quello acustico. Come possiamo vedere dalla Figura 4.5, tale ipotesi è confermata dall'andamento della Figura 4.5(b), confermando la nostra ipotesi cardine.

Nella Figura 4.5(b), sono state indebolite le connessioni *feedforward*, quindi possiamo notare che il segnale acustico è riuscito a predominare sul segnale visivo facendo spostare la percezione verso di esso. In questo modo la risposta riprodotta era sempre più fedele al fonema acustico, e ciò conferma la ridotta sensibilità all'effetto McGurk di un bambino autistico rispetto a un bambino normotipico, la cui risposta è rappresentata dal valore 1, corrispondente al caso in cui tutte le connessioni, sia cross-modali che *feedforward*, risultano integre.

In conclusione, una buona base di partenza per poter ottenere la risposta definitiva di un bambino autistico, consiste nel danneggiamento delle sinapsi *feedforward*.

4.3 Risposta del bambino autistico

Nel secondo capitolo, abbiamo illustrato i vari studi condotti per poter indagare le problematiche relative all'integrazione multisensoriale di un bambino con DSA. Abbiamo visto come, per un bambino autistico, la percezione e il riconoscimento dei fonemi acustici, quando questi vengono presentati da soli, sono abbastanza elevati, quasi al pari di un bambino normotipico. Quando invece, ad un bambino affetto da disturbi dello spettro autistico viene chiesto di riconoscere un fonema visivo, le differenze tra quest'ultimo e un bambino normotipico diventano rilevanti.

Per poter quindi ottenere dei risultati per confrontare la risposta di un bambino normotipico con quella di un bambino autistico, abbiamo cercato di identificare in cosa consistesse la differenza tra questi due soggetti, come descritto nel paragrafo precedente.

Ora, per poter concludere l'analisi affrontata in questo progetto di Tesi, possiamo riportare i risultati ottenuti a seguito delle simulazioni volte a riprodurre le caratteristiche salienti emerse sia dallo studio di Foxe [6] che dallo studio di Smith e Bennetto [8], fatte attraverso la nostra rete neurale.

Abbiamo deciso inoltre, a seguito delle informazioni ottenute durante l'analisi delle simulazioni, di lasciare inalterate le sinapsi *feedforward* che collegano l'area unisensoriale acustica a quella multisensoriale, e anche le sinapsi cross-modali che si stabiliscono tra le due aree unisensoriali, e di danneggiare invece solo le connessioni in avanti che collegano l'area unisensoriale visiva a quella multisensoriale, sulla base dell'ipotesi delle scarse capacità di comprendere i segnali visivi, che caratterizzano i soggetti affetti da disturbi dello spettro autistico.

Le simulazioni sono state condotte per 4 livelli di danneggiamento delle sinapsi in avanti, ovvero le sinapsi sono state indebolite del 10%, 20%,

30% e 40%.

Tali percentuali rispecchiano i vari gradi di *deficit* riscontrati in un bambino autistico.

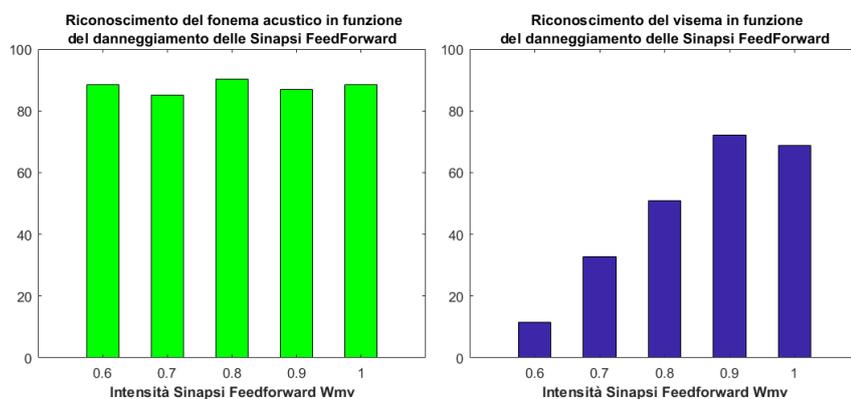
Sappiamo che l'autismo comprende una serie di *deficit* non identificabili in maniera univoca, quindi le percentuali di danneggiamento possono riflettere i vari livelli di autismo che caratterizzano un soggetto a cui è stata diagnostica il DSA.

Come per ognuna delle simulazioni discusse nei paragrafi precedenti, alla rete sono stati forniti tre diversi ingressi appartenenti alle tre condizioni solo acustica, solo visiva e audiovisiva.

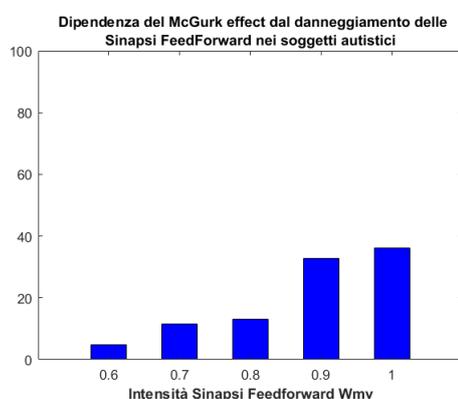
Per ognuno di questi ingressi sono state registrate le uscite, cercando di comprendere come per ogni percentuale di danneggiamento delle sinapsi, la rete modificasse le proprie risposte. Abbiamo quindi registrato la percentuale di riconoscimento del fonema acustico nella condizione solo acustica, la percentuale di riconoscimento del fonema visivo e, infine, la risposta in condizioni cross-modali, in funzione del segnale acustico, per poter valutare, di fatto, come cambiasse la sensibilità all'effetto McGurk a seconda della percentuale di danneggiamento.

Una volta ottenute, tutte queste informazioni sono state raccolte e riportate su una serie di istogrammi, dove l'asse delle ascisse riportava la percentuale di intensità delle sinapsi *feedforward* che collegano l'area unisensoriale visiva all'area multisensoriale W^{mv} , e sull'asse delle ordinate la percentuale di riconoscimenti.

Tali grafici sono riportati nella Figura 4.6, dove il valore 1 presente sull'asse delle ascisse (cioè sinapsi *feedforward* al valore normale) rappresenta il caso normotipico a cui vengono confrontate le risposte dei casi autistici. I grafici riportati nella Figura 4.6, mostrano i tre parametri che abbiamo osservato per ognuna delle simulazioni affrontate in questo progetto di Tesi. Queste simulazioni finali confermano le nostre ipotesi principali e sono in accordo con gli studi effettuati nel corso degli anni.



(a) Riconoscimento fonema acustico (b) Riconoscimento fonema visivo



(c) Sensibilità effetto McGurk

Figura 4.6: Mostrano le percentuali di riconoscimento a seguito del danneggiamento delle sinapsi *feedforward* in funzione del caso normotipico (Valore 1 in ascissa)

Se infatti osserviamo la Figura 4.6(a) possiamo vedere come la percentuale di riconoscimento del fonema acustico tende a restare pressoché simile in tutti i casi simulati. Di fatto, le percentuali vanno da 86.9% a 90% e le variazioni sono unicamente imputabili al rumore. Nel caso normotipico, utilizzato come caso campione, che ricordiamo è identificato dal valore 1 nell'ascissa, la percentuale di riconoscimento del segnale acustico è circa 88.5%. Un bambino autistico con età compresa fra i 9-11 anni ha delle prestazioni nel caso del riconoscimento del fonema acustico molto simili rispetto ad un bambino normotipico, in accordo con i nostri risultati.

Per quanto riguarda invece, il riconoscimento dei fonemi visivi o visemi, quando questi vengono presentati da soli, vediamo che la Figura 4.6(b) rispecchia proprio ciò che noi ci aspetteremmo. Infatti quando le sinapsi in avanti sono state danneggiate del 40%, una simulazione abbastanza grave che potrebbe riprodurre un bambino con gravi *deficit* di integrazione multisensoriale, la percentuale di riconoscimento del visema è dell'11.5%. Mano a mano che le sinapsi si intensificano, la percentuale di riconoscimento tende a crescere, fino ad arrivare ad una percentuale di circa 70% nel caso in cui le sinapsi sono state indebolite solo del 10%. Questo risultato potrebbe riflettere il caso di un soggetto con autismo ad alto funzionamento, le cui capacità integrative risultano di poco inferiori rispetto ad un soggetto normotipico. Il medesimo risultato è riscontrabile anche nel riconoscimento del fonema acustico.

Infine, se osserviamo la risposta multisensoriale dell'area STG/S riprodotta in funzione del segnale acustico, e poi riportata nella Figura 4.6(c) in funzione dell'intensità delle sinapsi *feedforward*, possiamo notare la stretta dipendenza tra la sensibilità all'effetto McGurk e il danneggiamento delle sinapsi.

Infatti i valori riportati nel nostro grafico evidenziano e confermano le ipotesi sostenute in questo progetto di Tesi. La sensibilità all'effetto McGurk è minore nei soggetti autistici e tende a diminuire sempre più a causa del danneggiamento delle connessioni che si stabiliscono tra l'area unisensoriale visiva e l'area multisensoriale.

Sappiamo infatti che, attraverso i vari studi condotti, il fenomeno di integrazione multisensoriale avviene anche ai livelli più bassi di elaborazione delle informazioni; tuttavia, a causa della ridotta risposta ai segnali visivi tipica dei deficit mostrati dai bambini autistici, non avverrà la fusione sensoriale che invece si può riscontrare in un bambino normotipico.

Se invece il danneggiamento di tali connessioni, è minore, la rete, e quindi un bambino autistico, tende a riconoscere più facilmente il fonema visivo. Ciò si traduce in un *input* visivo più forte verso la rete multisensoriale,

capace di influenzare il riconoscimento e quindi produrre la fusione sensoriale. A conferma, si riscontra una sensibilità più elevata all'illusione audiovisiva, pari a circa il 32% nel caso in cui le sinapsi sono danneggiate del 10%. Un valore che è molto vicino a quello del caso normotipico, il quale si aggira invece intorno al 36.1%.

Conclusioni

L'autismo o, per meglio dire, i disturbi dello spettro autistico (DSA), costituiscono una grave patologia che si manifesta entro i primi 30 mesi di vita dell'individuo. I bambini affetti da questo disturbo sono caratterizzati da un tardivo sviluppo delle capacità comunicative e sociali.

Vari studi di analisi dei processi sensoriali, hanno indagato le cause di questi *deficit* comunicativi che affliggono i soggetti autistici, giungendo alla conclusione che una concausa di questa patologia può essere rappresentata da un ritardo nello sviluppo dei meccanismi di elaborazione e integrazione multisensoriale.

Tradizionalmente, si pensava che le interazioni di tipo multisensoriali avvenissero solo nelle aree associative di più alto livello, mentre oggi, in base agli studi citati nel corso di questa Tesi, si è giunti alla conclusione che una prima forma di integrazione si verifica anche nei livelli corticali inferiori.

Per comprendere se i *deficit* dei soggetti autistici sono da attribuire a cause attenzionali o neurofisiologiche, sono stati effettuati alcuni studi in letteratura, indagando sulle cause delle differenze che si possono riscontrare tra i soggetti autistici e i soggetti normotipici, ovvero soggetti con capacità intellettive e neurologiche intatte. Quando i soggetti autistici eseguivano compiti semplici, quali l'ascolto di fonemi solo in condizioni acustiche, le loro performance erano pressoché simili a quelle dei soggetti normotipici.

Quando invece, il bambino autistico si trova di fronte ad un compito complesso, come ad esempio la comprensione di un discorso, egli non riesce a svolgere il compito in maniera adeguata rispetto al bambino normotipico. Riuscire a comprendere un discorso, soprattutto in condizioni rumorose, richiede di integrare diverse informazioni, quali ad esempio la voce e i movimenti labiali dell'interlocutore, ed è proprio questo il fenomeno dell'integrazione multisensoriale.

Ora, sulla base di questi risultati, in questo progetto di Tesi, attraverso l'utilizzo di una rete neurale, si è indagato sulle possibili cause delle differenze esistenti tra i soggetti normotipici e i soggetti autistici, valutando l'integrazione tra gli stimoli audiovisivi attraverso l'utilizzo dell'illusione audiovisiva, conosciuta con il nome di effetto McGurk.

In questo progetto si è analizzata non solo la risposta multisensoriale, ma anche la capacità di riconoscere i fonemi nelle condizioni unisensoriali solo acustiche e solo visive.

Per valutare le risposte multisensoriali, si è utilizzato l'effetto McGurk, poiché nei soggetti autistici si può riscontrare una minore sensibilità a tale effetto. In base a questo assunto, si sono comparati i risultati dei soggetti autistici con la loro controparte normotipica.

Inizialmente, le differenze tra i soggetti autistici e normotipici erano state attribuite ad un indebolimento delle sinapsi cross-modali, ovvero le connessioni che si stabiliscono tra le aree unisensoriali. In base ai risultati ottenuti, le ipotesi sono risultate errate, poiché le risposte dei soggetti autistici risultavano quasi simili a quelle dei soggetti normotipici.

Successivamente, l'attenzione si è spostata verso l'ipotesi di un indebolimento delle connessioni *feedforward*, ovvero le connessioni che si stabiliscono tra ogni regione unisensoriale e la regione multisensoriale. In questo progetto di Tesi, l'area multisensoriale è stata identificata con l'area corticale associativa STG/S, adibita alla comprensione del linguaggio e all'integrazione multisensoriale.

Si sono analizzati diversi livelli di indebolimento delle sinapsi *feedforward*, comparandoli con il caso normotipico. Si è evidenziato che un danneggiamento di tali connessioni di circa il 40%, rispetto al caso normotipico, produce una scarsissima sensibilità all'effetto McGurk. Mentre, con indebolimenti minori, la sensibilità all'illusione audiovisiva tende ad aumentare.

I risultati ottenuti si sono dimostrati incoraggianti e in linea con i risultati osservabili in letteratura.

Inoltre, si è giunti alla conclusione che le possibili cause dei *deficit* riscontrabili nei soggetti con DSA, sono appunto da attribuire a una minore intensità delle connessioni *feedforward* e non alle connessioni cross-modali, come inizialmente si era ipotizzato.

Di fatto, tali risultati hanno mostrato come il riconoscimento dei fonemi nei bambini autistici trae poco vantaggio dall'aggiunta di informazioni visive. Il *deficit* integrativo è assai debilitante, e influenza altamente le esperienze quotidiane del bambino. Per poter migliorare tali esperienze, sia i genitori che gli insegnanti, potrebbero utilizzare delle tecniche di apprendimento e delle pratiche di motivazione, aiutando il bambino nella comprensione del discorso.

Bibliografia

- [1] M. M. Murray, D. J. Lewkowicz, A. Amedi, and M. T. Wallace, “The merging of the sense 2.0: Development, plasticity & learning,” Manuscript Number: TICS-D-14-00306.
- [2] M. A. Meredith and B. E. Stein, “Interactions among converging sensory inputs in the superior colliculus,” *Science*, vol. 221, no. 4608, pp. 389–391, 1983.
- [3] M. A. Meredith and B. E. Stein, “Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration,” *Journal of neurophysiology*, vol. 56, no. 3, pp. 640–662, 1986.
- [4] R. A. Stevenson, D. Ghose, J. K. Fister, D. K. Sarko, N. A. Altieri, A. R. Nidiffer, L. R. Kurela, J. K. Siemann, T. W. James, and M. T. Wallace, “Identifying and quantifying multisensory integration: a tutorial review,” *Brain topography*, vol. 27, no. 6, pp. 707–730, 2014.
- [5] C. R. Fetsch, G. C. DeAngelis, and D. E. Angelaki, “Bridging the gap between theories of sensory cue integration and the physiology of multisensory neurons,” *Nature Reviews Neuroscience*, vol. 14, no. 6, pp. 429–442, 2013.
- [6] J. J. Foxe, S. Molholm, V. A. Del Bene, H.-P. Frey, N. N. Russo, D. Blanco, D. Saint-Amour, and L. A. Ross, “Severe multisensory speech integration deficits in high-functioning school-aged children

- with autism spectrum disorder (asd) and their resolution during early adolescence,” *Cerebral Cortex*, p. bht213, 2013.
- [7] N. Taylor, C. Isaac, and E. Milne, “A comparison of the development of audiovisual integration in children with autism spectrum disorders and typically developing children,” *Journal of autism and developmental disorders*, vol. 40, no. 11, pp. 1403–1411, 2010.
- [8] E. G. Smith and L. Bennetto, “Audiovisual speech integration and lipreading in autism,” *Journal of Child Psychology and Psychiatry*, vol. 48, no. 8, pp. 813–821, 2007.
- [9] C. Cuppini, M. Ursino, E. Magosso, L. A. Ross, J. J. Foxe, and S. Molholm, “A computational analysis of neural mechanisms responsible for the maturation of the multisensory speech integration in neurotypical children and those on the autism spectrum,” 2017.

Ringraziamenti

Mi avevano avvisato, scrivere i Ringraziamenti non sarà la cosa più semplice. Non so se confermare o smentire, ma di certo è una cosa che sento il bisogno di fare.

Il mio percorso di studi universitari non è iniziato nel migliore dei modi, per svariati motivi, ma oggi sono qui, a scrivere le ultime pagine della mia Tesi.

Quindi, mi sembra giusto iniziare così:

Le prime persone che voglio ringraziare sono i miei genitori, Mamma e Papà. Mi avete fornito tutti i mezzi per poter affrontare questo percorso, mi avete supportato e sopportato, cosa non facile. Non avete preteso nulla in cambio, se non la mia felicità.

Questo Traguardo è anche vostro.

Inoltre, voglio ringraziare il resto della mia famiglia. Mi avete fatto crescere, regalandomi splendidi ricordi e donandomi 4 piccoli, grandi pezzi di cuore.

Ora non è facile continuare, ma la persona che più merita una menzione in queste righe sei tu, Giuseppe. Mi hai accolto nel tuo mondo, mi hai guidato, mi hai supportato. Tu sei la mia ancora, la parte migliore di me. Ci sei sempre stato e hai fatto tanto per me, per noi. Ed è per tutto

questo e per molto altro ancora che ti dico: Grazie.

Volevo inoltre ringraziare il mio relatore, il professor Mauro Ursino. Mi ha fatto conoscere una parte della Biomedica a me sconosciuta. Con pazienza e gentilezza, mi ha insegnato tantissime cose e reso questo progetto di Tesi una bella esperienza.

Infine, per ultimi, ma non per importanza, voglio ringraziare voi, amiche e amici miei, vicini e lontani. Avete sopportato le mie sfuriate, i miei crolli, avete condiviso con me gioie, risate e tutto ciò che un'amicizia può portare nelle nostre vite. Vi prego, non cambiate mai, spero di avervi sempre al mio fianco.

Tutti voi avete reso questo percorso meno arduo, ed è per questo che vi dico un grande, sincero e commosso GRAZIE.

Vostra, Daniela