



SAPIENZA
UNIVERSITÀ DI ROMA

Facoltà di Filosofia, Lettere, Scienze Umanistiche e Studi Orientali

Dottorato di ricerca in
Linguistica Storica e Storia della Linguistica Italiana e Fonetica
CICLO XXIV

PARAMETRI NEUROLINGUISTICI PER L'*unspoken* SPEECH

Dottorando

Dott.ssa **PAMELA MATTANA**

Tutori

Prof. Ing. Paoloni Andrea

Prof. Di Giovine Paolo

Prof.ssa Giunchi Paola

Novembre 2008-2012

A mio figlio Tiziano...

“Chi tace spaventa.”

Alda Merini

INDICE

0. Introduzione

1. La teoria e le tecniche

1.1. Linguistica: modelli per la produzione e interpretazione del parlato

1.1.1. Le teorie passive

1.1.1.1 *Fuzzy-logical model of perception*

1.1.2. Le teorie attive

1.1.2.1. *La motor theory of speech perception*

1.1.2.2. *L'Analysis-by-Synthesis Theory*

1.1.3. Alcune teorie sull'esistenza delle invarianti fonetiche

1.1.3.1. Cole e Scott

1.1.3.2. Stevens e Blumstein

1.1.3.3. La teoria di Lindblom

1.1.4. Le teorie autonome

1.1.4.1. *Forester Model*

1.1.4.2. Il modello *the language processing module*

1.1.5. Le teorie di interazione

1.1.5.1. Il *logogen model*

1.1.5.2. Il *cohort model*

1.1.5.3. La teoria percettiva di Cole e Jakimik

1.1.5.4. *L'interactive Activation Model* e il TRACE

1.1.5.5. LAFS: *Lexical Access from Spectra*

1.2. Elementi di neurofisiologia

1.2.1. Morfologia generale di un neurone

1.2.1.1. Segnalazione neuronale: il flusso di informazioni attraverso il s.n.c.

1.2.1.2. La segnalazione neuronale attraverso le sinapsi

1.2.1.3. La trasmissione sinaptica

1.2.2. Il sistema nervoso centrale (s.n.c.)

1.2.2.. Area di Broca e Area di Wernicke

1.3. L'elettroencefalografia

1.3.1. Strumentazione e funzionamento

1.3.1.1. Gli elettrodi

1.3.1.2. Il posizionamento

1.3.2. I ritmi

1.3.3. Problematiche

1.4. Altre tecniche d'indagine

1.4.1. La TMS

1.4.1.1 Il funzionamento

1.4.1.2. Caratteristiche

1.4.1.3. Applicazioni in neurolinguistica

1.4.2. La fMRI :La risonanza magnetica funzionale

1.4.2.1. Il funzionamento

1.4.2.2. Caratteristiche

1.4.2.3. Applicazioni in neurolinguistica

1.4.3. La PET: Tomografia ad emissione di positroni

1.4.3.1 Il funzionamento

1.4.3.2. Neurolinguistica e PET

1.4.4. MEG

1.4.4.1. Neurolinguistica e Meg

1.4.4.2. I limiti

1.5. Gli ERP: potenziali evento correlati

1.5.1. Il problema inverso

1.5.2. Breve classificazione dei potenziali eventocorrelati o ERP

2. Le nuove frontiere della comunicazione

2.1. *Brain Computer Interfaces*

2.1.1. Tipologie di BCI

2.2. *Silent speech interface*

2.2.1. Quadro storico

2.2.2. Cattura del movimento articolatorio attraverso EMA

2.2.3. Caratterizzazione in tempo reale del tratto vocale utilizzando l'*imaging* a ultrasuoni (US) della lingua e delle labbra

2.2.4. Registrazione del segnale audio attraverso sensori di vibrazione ed elettromagnetici

2.2.4.1. Analisi di attività glottale con sensori di vibrazione ed elettromagnetici

2.2.4.2. Sensori di vibrazione: laringofono

2.2.4.3. Microfono fisiologico, pmic

2.2.4.4. Microfono da orecchio

2.2.4.5. Sensori elettromagnetici

2.2.4.6. GEMS

2.2.4.7. Elettromiografia di superficie (emg) e riconoscimento vocale

2.2.5. Interpretazione dei segnali elettroencefalografici (EEG)

2.2.5.1. Wester e Schultz (2006)

2.2.5.2. DaSalla et al (2009)

2.2.5.3. DIVA

2.2.5.4. I problemi comuni

3. La ricerca: tre protocolli sperimentali

3.1. Il primo protocollo

3.1.1. I soggetti

3.1.2. La raccolta del *corpus*

3.1.3. Gli strumenti: *Presentation*

3.1.4. Gli strumenti: *Brain Vision*

3.2. Il secondo protocollo

3.2.1. I soggetti

3.2.2. La raccolta del *corpus*

3.2.3. *Gli strumenti: Brain Vision*

3.2.4. *Gli strumenti: Eprime*

3.3. Il terzo protocollo

3.3.1. I soggetti

3.3.2. La raccolta del *corpus*

3.3.3. Gli strumenti: ATES

3.4. I materiali

4. Dal trattamento dei corpora ai dati

4.1. Gli strumenti: *EEG Lab*

4.2. *Preprocessing*: rimozione degli artefatti e filtraggi

4.2.1. Il problema degli artefatti

4.2.2. ICA, *independent Component Analysis*

4.2.3. *Epoching*

4.2.4. *Averaging*

4.2.5. *Common Spatial Patterns*

4.3. *Pattern recognition*: estrazione delle *features*

4.4. Le SVM

5. Dai dati ai Risultati

5.1. Introduzione

5.2. La prima fase: elaborazione nel dominio della frequenza (primo esperimento)

5.3. Ulteriore metodo di classificazione: *common spatial patterns* (primo esperimento)

5.4. Ulteriori elaborazioni nel dominio del tempo

5.5. I risultati

5.5.1. Il primo esperimento: sessione 1 e 2

5.5.1.1. Valori di riconoscimento della macchina con vettore a media di dati

5.5.1.2. Valori di riconoscimento della macchina con algoritmo a vettore di dati singoli

5.5.1.3. Riflessioni sui dati del primo esperimento

5.5.2. Il secondo esperimento: sessione 1 e 2

5.5.2.1. Riflessioni sui dati del secondo esperimento

5.5.3. Il terzo esperimento: sessione unica

5.5.4. Test risposte audio

5.6. Riflessioni conclusive sui risultati

5.7. Analisi neurofisiologiche dei segnali: mappe di attivazione e Average

6. Conclusioni

6.1. Introduzione

6.2. Primo esperimento

6.3. Secondo esperimento

6.4. Terzo esperimento

6.5. Contributi alla conoscenza del funzionamento del linguaggio articolato

6.6. Per il futuro

7. Bibliografia

8. Appendice e tavole

0. INTRODUZIONE

Il lavoro da noi intrapreso ci ha svelato un ambito nuovo della ricerca di tipo linguistico. Si presenta come una “ricerca di frontiera” sia per gli aspetti tecnici ma, in modo più evidente per gli aspetti teorici, poiché il carattere pionieristico di essa porta inevitabilmente a confrontarci con i risultati delle scienze cognitive sempre più pervasive nei diversi ambiti della ricerca sull'umanità, che hanno stravolto le ipotesi sulla comunicazione e il linguaggio. Abbiamo adottato la metodologia del Brain Computer Interface inserendoci nel filone di ricerca del Silent Speech, cioè della comunicazione tra parlanti senza produzione di segnale vocale udibile. Oggi numerosi istituti di ricerca si occupano di mettere a punto tecnologie per “catturare” il segnale connesso con il parlato, privilegiando due metodologie: la registrazione dei segnali elettromiografici (segnali elettrici provenienti dai muscoli connessi con l'apparato fonatorio) e attraverso il segnale neurofisiologico a livello della corteccia cerebrale. Queste tecniche per la comunicazione silenziosa troverebbero molteplici impieghi rispondendo ai diversi usi, nel campo medico e in quello della sicurezza: in ambienti rumorosi (veicoli spaziali, luoghi affollati, catene di montaggio), per la trasmissione di informazioni riservate (operazioni di polizia o azioni di guerra), per ridurre l'inquinamento acustico, per permettere a soggetti patologici di comunicare anche quando hanno perso l'uso dell'apparato vocale. Questo lavoro si propone di decodificare il segnale connesso con la voce prima che esso diventi segnale vocale. In sostanza la ricerca tenta la decodifica, tramite opportune registrazioni EEG, del segnale connesso al pensiero di articolazione di alcuni foni vocalici dell'italiano ([a],[i],[u]); ai soggetti è stato richiesto di pensarli senza poi produrli acusticamente. La metodologia scelta, di tipo non-invasivo, per la cattura del segnale cerebrale, da una parte svincola da alcune problematiche connesse alle pratiche invasive (introduzione di elettrodi all'interno della corteccia cerebrale) dall'altro, pone di fronte alla difficoltà dell'elicitazione dei dati con elettrodi sullo scalpo, non garantendo quella che genericamente potremmo definire la qualità dei segnali.

Sono stati adottati 3 differenti protocolli al fine di ottenere una visione completa delle possibilità di analisi e di elaborazione dei dati. In questo modo si è raccolta un'ampia quantità di dati anche rispetto alla letterature di riferimento. I primi due esperimenti sono stati condotti presso il Laboratorio del CRIL di Lecce. Il primo esperimento è stato eseguito utilizzando il protocollo suggerito dalla bibliografia di riferimento (Da Salla, 2009) sono state apportate alcune modifiche e adattamenti: infatti abbiamo utilizzato una cuffia con un numero maggiore di elettrodi (64 elettrodi) e un numero più elevato di soggetti, 9 volontari (6 uomini e 3 donne). Il secondo esperimento ha previsto la modifica del protocollo iniziale inserendo una fase che permettesse durante l'analisi di effettuare anche delle valutazioni sulla

percezione. Il terzo e ultimo esperimento è stato condotto presso l' Unità Ospedaliera di Neurologia di Belcolle (VT), in questo caso ha partecipato un solo soggetto e il protocollo è stato modificato per meglio adattarsi alla strumentazione. Nel procedere con l'analisi dei dati che via via venivano acquisiti ed elaborati, ci si è resi conto di come essi siano strettamente tipici e dipendenti dai singoli soggetti. Evidenziandosi una così una grande varietà si è passati da un sistema *speaker-independent* a uno *speaker-dependent*.

Il lavoro, traendo ispirazione dagli studi di Guenther (2010), ha previsto l'adozione del posizionamento degli elettrodi il più simile possibile pur rinunciando alla tecnica invasiva dell'ECOG (con elettrodi nella corteccia cerebrale). Altro riferimento è stato Da Salla (2009), abbiamo infatti fatto riferimento al suo protocollo sperimentale adottando in fase di analisi il metodo CSP (common spatial pattern), nello stesso tempo ci siamo discostati per la componente evidenziata la P300 nel caso di Da Salla nel nostro caso la P200 (si tratta di onde cerebrali con potenziale positivo che appaiono con una latenza di 300ms o 200ms rispetto allo stimolo). In questo lavoro sembra più rilevante la P200, quindi una componente che si presenta prima nell'asse temporale e, di conseguenza, i risultati ottenuti si discostano in maniera evidente da Da Salla. Non è possibile effettuare un confronto con i tempi ottenuti da Guenther: la tecnologia utilizzata al MIT ha una sensibilità e una risoluzione temporale molto superiore all'EEG. Questo lavoro cerca di rispondere, con l'impiego di tecniche nuove all'indagine sperimentale in campo linguistico, a come e dove avvenga la produzione degli stimoli motori, che comandano gli articolatori dei foni utilizzati nella comunicazione verbale e se essi siano da considerare, come vuole la teoria motoria di Lieberman (1985), delle **invarianti** linguistiche. I risultati ottenuti appaiono statisticamente convincenti, riteniamo quindi che sia questa la strada da percorrere. In futuro si potrà anche pensare di addestrare un sistema artificiale che traduca il segnale neurofisiologico in voce. La relazione si articola in 5 sezioni: capitolo 1 basi fisiologiche e teorie della produzione del parlato, capitolo 2 una panoramica sui principali studi e ricerche sul *silent speech* con le relative tecniche, capitolo 3 la descrizione dei protocolli sperimentali e i software utilizzati con elenco dei materiali nel CD, capitolo 4 in questa sezione sono raccolte tutte le procedure di elaborazione del segnale e gli algoritmi utilizzati per giungere ai risultati.

La sezione 5 racchiude i risultati dei test a cui sono stati sottoposti i segnali.

1. LA TEORIA E LE TECNICHE

1.1. Linguistica: modelli per la produzione e interpretazione del parlato

Nell'ambito delle teorie percettive del linguaggio possiamo distinguere principalmente due indirizzi: quello delle teorie attive e quello delle teorie passive. Le teorie attive sono così denominate perché prevedono la partecipazione attiva alla decodifica del messaggio. Le teorie passive ritengono che il soggetto che percepisce il messaggio non svolga nessuna funzione su esso bensì, si limiti a riceverlo e quindi la decodifica passi attraverso la sola elaborazione dei parametri fisici del segnale. Questo processo viene considerato esclusivamente sensorio.

1.1.1. Le teorie passive

La teoria di Gunnar Fant, di tipo acustico, è sicuramente tra quelle più note e che ha avuto maggior successo e applicazione soprattutto nell'ambito dello sviluppo di sintetizzatori vocali come il suo *Orator Verbis Electricis* (OVE, fig. 1).

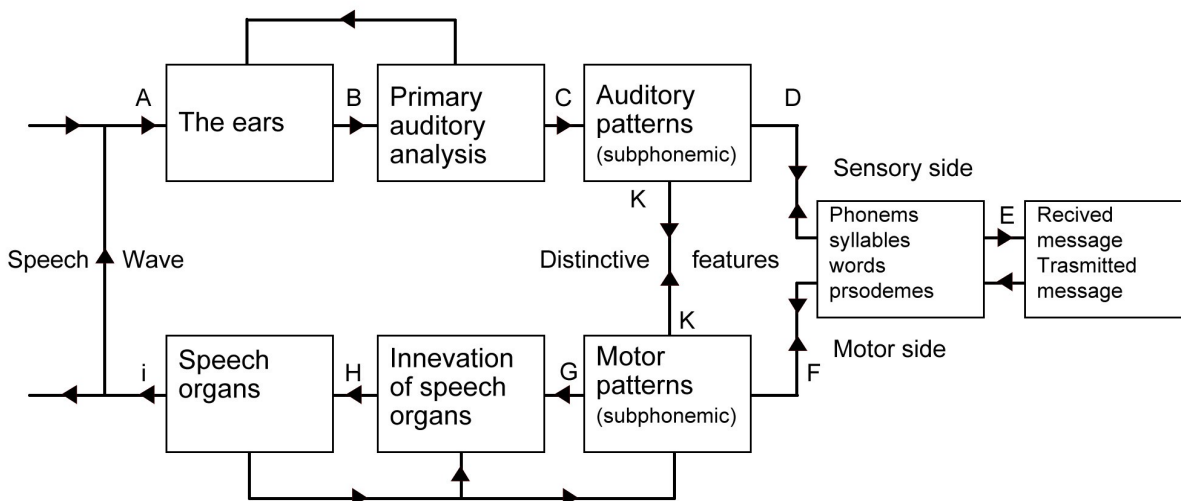


Figura 1: modello Gunnar Fant

Questa teoria si basa sul concetto di tratto distintivo; infatti l'ascoltatore attraverso le proprie strutture sensoriali periferiche e centrali sarebbe in grado di riconoscere all'interno del segnale acustico quelle caratteristiche fisiche proprie dei diversi tratti distintivi.

Queste caratteristiche fisiche sarebbero sufficienti a veicolare l'informazione contenuta nella porzione prodotta dal parlante nella sua interezza e complessità. Secondo questa teoria l'ascolto è essenzialmente un processo sensorio, vi è quindi un rapporto diretto e univoco tra le informazioni contenute nello stimolo acustico e la risposta neurale che esse attivano.

1.1.1.1. Fuzzy-logical model of perception

È un altro modello passivo elaborato da Massaro a partire dagli anni 70, esso è costituito da dei nodi centrali (fig. 2):

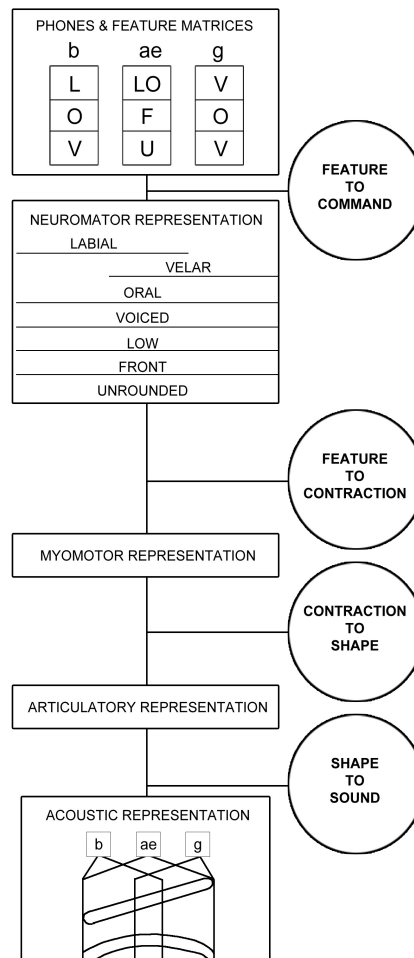


Figura 2: schema Fuzzy

- "each source of information is evaluated to give the continuous degree to which that source specifies various alternatives; the sources are integrated to provide an overall degree of support for each alternative;
- perceptual identification and interpretation follows the relative degree of support among the alternatives."
- (Campbell & Massaro, 1997; Massaro, 1987, chapter 1; Massaro & Friedman, 1990)
- ciascun nodo è indipendente da ciascun altro.

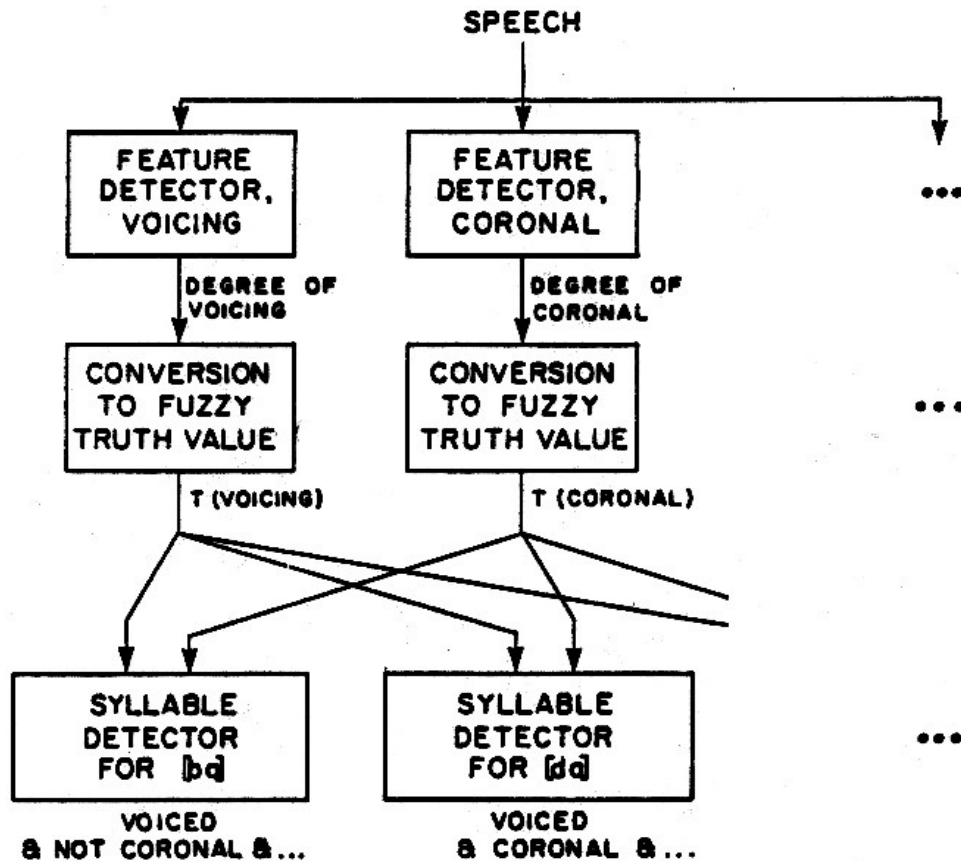


Figura 3: sistema automatico rivelazione dei tratti Massaro (1987)

Più esplicitamente, questo modello si basa sul riconoscimento di prototipi di andamenti, siano essi prosodici che semplici tratti distintivi dei singoli fonemi: essi vengono estratti dal *fuzzy*, cioè il *continuum*; in base al modello probabilistico ad ogni tratto viene assegnato un valore di esistenza in termini di probabilità, da 0 a 1, nel continuum dello stimolo. Successivamente, tramite i gradi con cui si presentano i tratti (fase del processamento dell'informazione) si giungerebbe all'associazione con i prototipi in memoria e quindi al riconoscimento. Cioè il grado di probabilità di un tratto azionerebbe il riconoscimento del corrispondente prototipo di cui l'ascoltatore ha la rappresentazione in memoria (fig.3).

Questo paradigma permette di determinare come una sorgente di informazione venga processata ed integrata con altre fonti d'informazione. Infatti questo paradigma è utilizzato per l'analisi delle modalità di comprensione dell'audio e del video e in particolar modo nella comunicazione bimodale. Di contro a queste due tesi si pone l'evidenza che la variazione è intrinseca nel segnale linguistico e la decodifica attraverso parametri così costituiti porterebbe alla comprensione di appena il 33% del messaggio con la lettura dei sonogrammi

(Klatt e Stevens, 1973) o al massimo del 56% (Shokey e Reddy, 1974) per la sola trascrizione fonetica di una lingua sconosciuta. In aggiunta se così fosse non sarebbe spiegabile la comprensione del messaggio e quindi la comunicazione in ambienti rumorosi o nel caos di segnale deteriorato. Quest'ultimi casi ci portano a riflettere sul ruolo fondamentale che svolge l'ascoltatore nel decodificare il messaggio mettendo in atto delle strategie che da una parte possono definirsi ontogenetiche e dall'altro filogenetiche.

1.1.2. Le teorie attive

Come già anticipato, queste teorie sostengono una partecipazione attiva dell'ascoltatore alla decodifica del messaggio; infatti egli non solo percepisce dal punto di vista fisico, ma elabora, estrae, opera comparazioni e ricostruisce in base alla sua conoscenza, nel caso il segnale sia giunto incompleto. L'ascoltatore agisce a più livelli: dalla "conoscenza" del gesto articolatorio che porta dalla percezione del segnale acustico, alla sua identificazione e categorizzazione come suono linguistico e, successivamente come dato fonema; alla "conoscenza" delle regole della lingua che gli permettono di ricostruire in *absentia* (segnale rumoroso, o semplice parlato ipoarticolato), il significato del messaggio.

Le teorie attive che hanno avuto maggior impiego e risonanza sono la *Motor Theory of Speech Perception* di Liberman (Liberman, 1967) e la *Analysis-by-Synthesis Theory* proposta da Stevens, Halle e House, entrambe partono dai medesimi presupposti:

- i suoni linguistici hanno una struttura acustica specifica e diversa da quella dei suoni non-linguistici (toni puri, suoni musicali, rumori) ciò è stato confermato da studi neurolinguistici;
- decodifica e di codifica sono proprietà ontogenetiche cioè specifiche dell'ascoltatore/parlante immerso nel proprio mondo;
- il fonema non ha un invariante perché è esso prodotto del contesto fonetico, del contesto prosodico, dello stile, nonché delle caratteristiche sociolinguistiche del soggetto che produce il continuum fonico.

1.1.2.1. La *motor theory of speech perception*

Presso gli Haskins Laboratories, Liberman e i suoi collaboratori, cercarono di spiegare il come e perché alla variabilità acustica corrisponda un'invariabilità percettiva. L'ipotesi generata fu che l'invarianza fosse da ricercare ad un livello alto della produzione, cioè dei comandi neuromotori e che di contro la variabilità del segnale insorgesse a livello muscolare,

dovuta a condizionamenti muscolo-articolatori quindi di tipo meccanico come, l'inerzia neuromuscolare. Questa teoria ipotizza che i soggetti comprendano il parlato tramite l'identificazione dei *gesti* motori del tratto vocale che contribuiscono alla produzione del parlato. Questa capacità sarebbe innata e specifica dell'uomo, per cui la capacità di articolare il parlato è strettamente legata alla capacità di identificare/comprendere lo stesso. Nella pratica per decodificare un segnale l'ascoltatore ripeterebbe internamente i movimenti che il parlante fa per produrre quel dato messaggio di natura verbale. L'ipotesi si basa su un principio di economia: poiché il parlante è anche ascoltatore, è economica l'esistenza di un solo processo che si attiva al livello profondo sia in produzione che in ricezione. Liberman sostiene che gli "*intended phonetic gestures*" cioè "*i movimenti articolatori intenzionali*" del parlante hanno una propria rappresentazione, sotto forma di comandi motori linguisticamente significativi di tipo invariante; dal punto di vista dell'ascoltatore gli stessi invarianti sono oggetto della percezione. Questa caratteristica è di natura innata nel parlante-ascoltatore. Le fasi della comprensione del segnale prevedevano che esso fosse sottoposto in primo luogo ad un'analisi delle sue componenti fisiche (frequenza, tempo, in

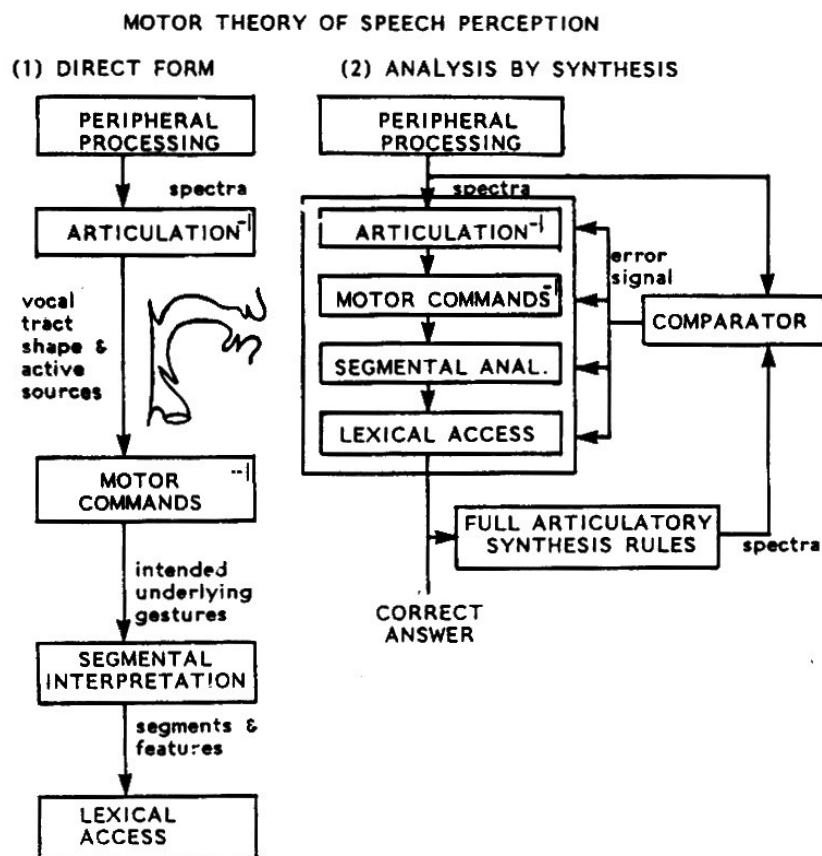


Figura 4: diagramma di funzionamento della Motor Theory

tensità) a questo modulo acustico viene associato uno di tipo neurale. Il modulo è quindi una struttura neurale che opera una relazione tra il segnale acustico verbale e le caratteristiche del gesto articolatorio. L'analisi uditiva del segnale viene effettuata al solo scopo di riconoscere la natura dei suoni, tra vocalici e non-vocalici e le varie tipologie di suoni naturali (fig.4). Ad oggi questa teoria è supportata dalla scoperta dei neuroni specchio, che lega la produzione con la percezione dei movimenti, compresi quelli del tratto vocale (Rizzolatti, 2009) La teoria che inizialmente era di tipo associazionista per cui il bambino imitava il parlato che aveva udito operando così un'associazione tra articolazione e ciò che il suo comportamento comportava, successivamente passa ad un approccio di tipo cognitivista con l'ipotesi dello *speech module*. Nel 1970 Studdert-Kennedy apportò una modifica alla teoria motoria. Ipotizzò quattro stadi di decodifica del segnale in entrata. L'analisi del segnale di input è condotta in modo che si proceda contemporaneamente seguendo le fasi dell'analisi l'una dopo l'altra ma tenendo sempre il controllo sulla coerenza con ciò che avviene in contemporanea.

- stadio uditivo: una serie di sistemi neurologici parzialmente indipendenti estraggono le proprietà fisico-acustiche relative alla struttura dello spettro, alla F0, all'intensità e alla durata del segnale di tipo linguistico. Quindi le trasformano in una mappa di eventi neurologici che varia nel tempo. Questi *feature-detectors* sono legati ai tratti linguisticamente rilevanti del segnale acustico.
- stadio fonetico: il segnale, ora non è più considerato meramente un fatto percettivo-uditivo, è identificato come generato dagli organi fono-articolatori.
- stadio dell'analisi fonologica: il segmento fonetico viene trasformato in un fonema del sistema linguistico proprio dell'ascoltatore, egli per questo compito utilizza le proprie regole fonologiche per determinare le caratteristiche fonologiche e fonotattiche della sequenza sonora percepita (ad esempio se si tratta della propria varietà di lingua o meno).
- stadio della comprensione: ad essa si giunge attraverso la decodifica del significato del messaggio linguistico per mezzo delle regole lessicali, sintattiche e semantiche.
- A fronte dei progressi in ambito neurolinguistico questa teoria da un lato è stata oggetto di rinnovato interesse, dall'altro ha visto la propria riformulazione che a oggi, proprio perché segue di pari passo le nuove scoperte, non ha ancora visto una sua veste definitiva.

Vedremo più avanti come Mac Neilage e Lindblom sostengano che la percezione non pre-

suppone la presenza di invarianti nel segnale acustico e propongono di ricercare l'invarianza a livello della meta finale del parlatore, ossia in termini di accesso lessicale o comprensione dell'ascoltatore.

1.1.2.2. *L'Analysis-by-Synthesis Theory*

La teoria *Analysis by Sythesis* nasce attorno al 1967 da Stevens, Halle e House, si può considerare come uno sviluppo più particolareggiato e formale, seppur con differenze, del modello di Lieberman. In accordo con il modello motorio troviamo l'idea che i pattern che sono generati dalle attività neuromotorie non solo vengono analizzati preliminarmente, ma vengono successivamente stoccati in uno spazio di memoria, per poi essere passati ad un'analisi di livello più alto: la decodifica vera e propria, che prevede il confronto tra gli esempi/modelli neurali in memoria e, quelli allocati nella memoria breve termine (fig.5). La differenza sostanziale tra i due approcci risiede nelle regole a cui deve rispondere il pattern per essere tale e per essere inserito in memoria, ma anche nelle regole cui deve sottostare il segnale da decodificare. Queste regole rispondono alla teoria generativa e il riconoscimento del pattern non si affida quindi alla memoria motoria- neurale, bensì alla soddisfazione delle regole. Infatti, il segnale entrante viene confrontato ciclicamente con tutti i pattern fino a trovare quello per regola simile. Il processo di produzione prende avvio da una serie di istruzioni inviate al meccanismo articolatorio per generare il suono, queste istruzioni corrispondono alla rappresentazione astratta dei segmenti fonetici e dei loro corrispondenti tratti; queste vengono definite matrici astratte, le regole provocano la trasformazione delle rappresentazioni astratte in istruzioni concrete, queste regole sono necessarie alla relazione tra il gesto articolatorio e le caratteristiche acustiche dei suoni che realizziamo. Una volta arrivati gli input agli organi articolatori e attivata la produzione, il segnale diventa un continuum non più segmentabile; ecco perché Stevens parla della necessità di ricercare l'invarianza nei cosiddetti Target, o bersagli cioè in quelle rappresentazioni astratte non ancora "corrotte" dall'inerzia degli organi articolatori motori e che sono inoltre influenzati da diversi fattori involontari.

Il processo percettivo, più schematicamente, procede così:

- l'analisi spettrale, o analisi uditiva periferica del segnale acustico: produce degli andamenti del segnale acustico che vengono allocati nella memoria a breve termine. A questo punto, il segnale in ingresso viene suddiviso in una serie di segmenti temporali, in base ai loro attributi acustici.

- analisi preliminare: estrazione dei tratti svincolati
- attivazione della comparazione dei tratti nella memoria breve termine con i pattern che rispondono alle regole generative.

Il legame tra produzione e percezione sta nelle rappresentazioni astratte comuni ai due processi; infatti ogni parlante ha nella propria memoria quelle rappresentazioni astratte e quella serie di regole che permettono le operazioni sui tratti, ma ha in se anche una descrizione di quei processi atti alla trasformazione delle istruzioni articolatorie in pattern uditivi.

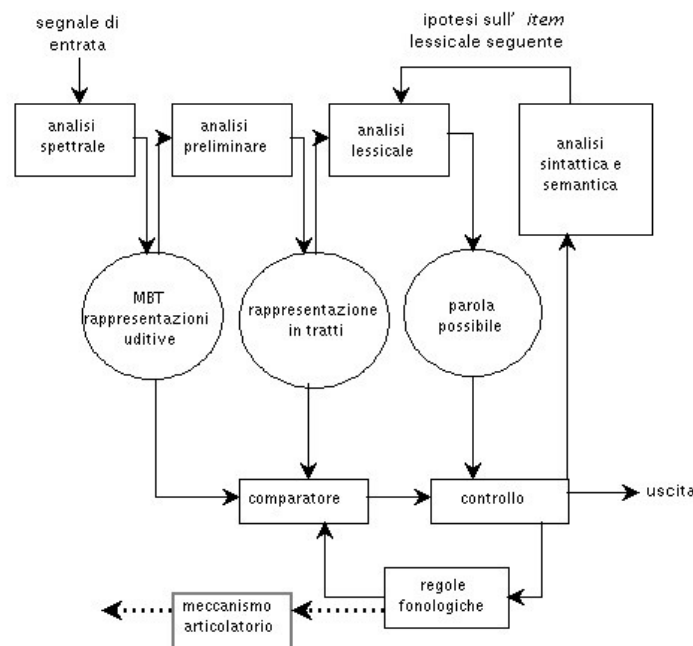


Figura 5: Modello dei processi percettivi proposto da Stevens e Halle (1967, 1972)

1.1.3. Alcune teorie sull'esistenza delle invarianti fonetiche

In questo paragrafo vogliamo ricordare brevemente l'esistenza di altre teorie che hanno come nucleo fondamentale il concetto di invarianti.

1.1.3.1. Cole e Scott

Nel 1974, in *"The phantom of the phoneme: invariant cues for stop consonants"* teorizzano, partendo da una serie di esperimenti percettivi, che la percezione implichi la simultanea identificazione di almeno tre tipi di parametri differenti: p. invarianti, p. condizionati dal contesto e p. contenuti nell'involuppo spettrale. Infatti attraverso questi esperimenti i due autori hanno riscontrato che alcuni foni sono caratterizzati da parametri invarianti. Ad esempio per le occlusive abbiamo la presenza del burst e del VOT, e di parametri contestuali; infatti

la distorsione delle formanti nell'inizio della vocale che segue l'occlusiva in questione, sarebbe una componente che permette, o meglio con i termini degli autori, agevola il riconoscimento anche in assenza della consonante stessa. Per le fricative la fase di rumore e di distribuzione spettrale delle componenti aperiodiche è da considerarsi un'invariante e solo sulla base di tali parametri che vengono percepiti, discriminati e soprattutto categorizzati questi suoni.

1.1.3.2. Stevens e Blumstein

Le loro attività hanno inizio intorno agli anni '80, condurranno una serie di esperimenti volti ad elaborare la loro teoria dell'invarianza acustica. Secondo quanto emerge dai loro lavori esisterebbe un rapporto di tipo invariante tra le proprietà acustiche del fono e il giudizio di tipo categorico- fonetico dell'ascoltatore. Ciò dimostrerebbe l'esistenza di *integrated acoustic cues*, esse corrisponderebbero in maniera diretta ai tratti fonetici distintivi delle lingue naturali. Nei loro studi una posizione di rilievo è rivestita dal burst delle consonanti occlusive, esaminando lo spettro nei vari contesti fonetici hanno concluso che esso costituisca l'indicatore principale per operare la discriminazione del luogo di articolazione delle occlusive. In particolare hanno così definito e categorizzato la forma dello spettro del *burst* nelle diverse tipologie di occlusive: nelle labiali con uno spettro di tipo *flat* o *diffuse-falling* (picchi di energia soprattutto nelle zone di bassa frequenza e man mano decrescenti), nelle dentali con uno spettro *diffuse-rising* (con energia crescente progressivamente verso le alte frequenze, nelle velari con uno spettro compatto con picchi di energia concentrati in una zona frequenziale ristretta e piuttosto centrale). L'ipotesi finale è che il segnale acustico sia ampiamente strutturato contenendo in se dei modelli acustici invarianti per ciascuno dei tratti fonetici: essi sono invarianti rispetto al parlatore, al fonetico e alla lingua.

La percezione usa queste invarianti costruendo un quadro fonetico per le lingue naturali e di conseguenza per comprendere ed elaborare i suoni linguistici all'interno del parlato continuo. Nonostante la similarità dei tratti proposti, si tratta di due approcci diversi perché Kewley-Port si basa su un modello psicofisico, mentre Blumstein adotta un approccio di linguistica formale. Considerare il *burst* e le transizioni formantiche come componenti di "*invarianti acustiche integrate*" risulta abbastanza interessante perché hanno proprietà acustiche simili e possono essere elaborati in modo analogo dal sistema uditivo, che è in grado di condurre solo analisi approssimate. Concordano con l'ipotesi sostenuta da Kewley-Port, anche i risultati di un recente esperimento sulla percezione delle occlusive singole e geminate della lingua italiana attraverso la presentazione di opportuni stimoli acustici,

che mostrano che le consonanti occlusive in contesto sillabico sono correttamente percepite dagli ascoltatori, che ne riconoscono il 95%, mentre gli stimoli acustici relativi al solo *burst* non sono sufficienti per una corretta percezione ed identificazione delle consonanti che, infatti, sono riconosciute solo nel 25% dei casi. Purtroppo questi approcci sono limitati nel fatto che non offrono un quadro completo delle caratteristiche acustiche invariante per tutte le categorie fonetiche.

1.1.3.3. La teoria di Lindblom

Secondo questa teoria la percezione è vista come frutto di un processo di elaborazione di due tipi di *input*: *signal dependent* cioè quella informazione interna al segnale e quindi di natura strettamente fonetica e *signal independent* cioè quella informazione esterna "independente" dal segnale e mediata dal contesto linguistico ed extra-linguistico. Secondo questo modello anch'esso di natura attiva, la percezione non ha bisogno di categorie rigide come potevano essere gli invarianti di tipo articolatorio ed acustico dei prototipi di Massaro; bensì si richiede che i gesti siano differenziati in maniera sufficiente da poter dare all'ascoltatore quelle informazioni acustiche che si sommano alla sua conoscenza del contesto da intendersi anche in senso grammaticale, in questo modo il soggetto che ascolta sarà in grado di ricostruire il segnale e cosa più importante, anche se disturbato o incompleto. Va da sé che questa teoria trova argomenti fondanti nel concetto di ridondanza ed economia della lingua.

1.1.4. Le teorie autonome

Le teorie autonome considerano l'elaborazione linguistica come un processo completamente seriale o modulare, in cui non avviene interazione tra i vari livelli di analisi e in cui le diverse fonti di conoscenza non possono intervenire nel processo di riconoscimento prima che questo sia completo.

1.1.4.1. Forester model

Tra queste la più rappresentativa è quella di Forester. Egli ritiene, infatti, che il processo percettivo sia costituito da tre moduli linguistici separati:

- modulo lessicale: le informazioni dal sistema percettivo periferico sono analizzate. Il sistema cerca di collocare un'entrata lessicale in tre file di accesso diversi: un file ortografico (per gli ingressi visivi) un file fonetico (per gli ingressi fonetici) e un file

sintattico-semantic (per entrambi i tipi di ingressi);

- modulo sintattico: la parola riconosciuta e le informazioni avviano la ricerca nel lessico principale finalizzata all'accesso lessicale;
- semantico;
- modulo generale per la decisione finale.

Questi moduli o processori in termini volutamente informatici, sono configurati in maniera seriale e indipendente, ogni modulo può ricevere ed elaborare solo l'output elaborato del modulo precedente. Le informazioni viaggiano in maniera univoca e lineare dal modulo lessicale a quelli successivi, nessuna nuova informazione può immettersi nel processo avviato o può risalire a ritroso i livelli. Si presuppone quindi l'indipendenza dell'intero sistema linguistico dal sistema cognitivo generale.

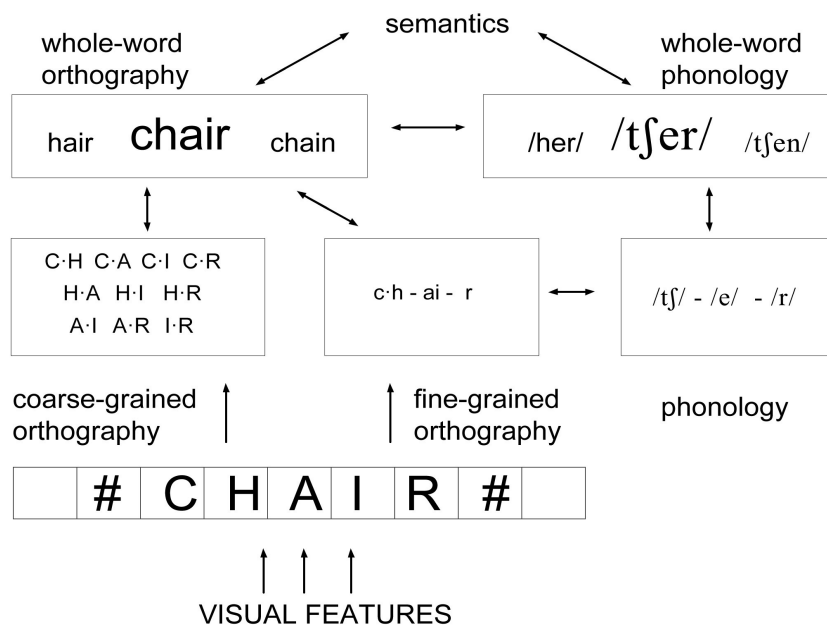


Figura 6: Diagramma funzionamento del Forester Model

1.1.4.2. Il modello *the language processing module*

Questo modello proposto da Fodor, si discosta dal precedente in due punti fondamentali: l'esistenza di un solo processore autonomo e la libera circolazione delle informazioni tra i vari sottocomponenti del modulo. Anche in questo caso non è permesso l'apporto di informazioni esterne al modulo, quindi le informazioni date dal contesto in cui si esprime e vive

il soggetto.

1.1.5. Le teorie di interazione

Queste teorie partono tutte da un'idea comune d'interazione dei vari livelli di analisi e d'integrazione delle molteplici *urquellen* per il riconoscimento della parola, discostandosi l'una dalle altre per la quantità di informazione necessaria in ciascuna delle fasi della produzione e del riconoscimento del parlato. In breve i candidati lessicali, generati dalle matrici, sono attivati in misura direttamente proporzionale alla loro similitudine con il candidato in ingresso.

1.1.5.1. Il *logogen model*

E' un modello ideato da Morton nel 1969. Esso utilizza delle unità chiamate logogheni per spiegare la produzione e la comprensione del parlato da parte dell'essere umano. I logogheni sono numerosi ciascuno specializzato nel riconoscimento una specifica parola; bisogna però chiarire che essi non sono parole o chiavi o ancora radici ma racchiudono delle proprietà specifiche e necessarie all'identificazione di un infatti essi posseggono una forma, un suono e un significato. Il logogheno si attiva quando sono riconosciute tutte le sue caratteristiche a questo punto sale di livello e il sistema è pronto per l'output. In alcuni casi si generano più opzioni ma non è chiarito il sistema di selezione, ne' sono ben chiariti gli aspetti di tipo psicolinguistico (fig.7).

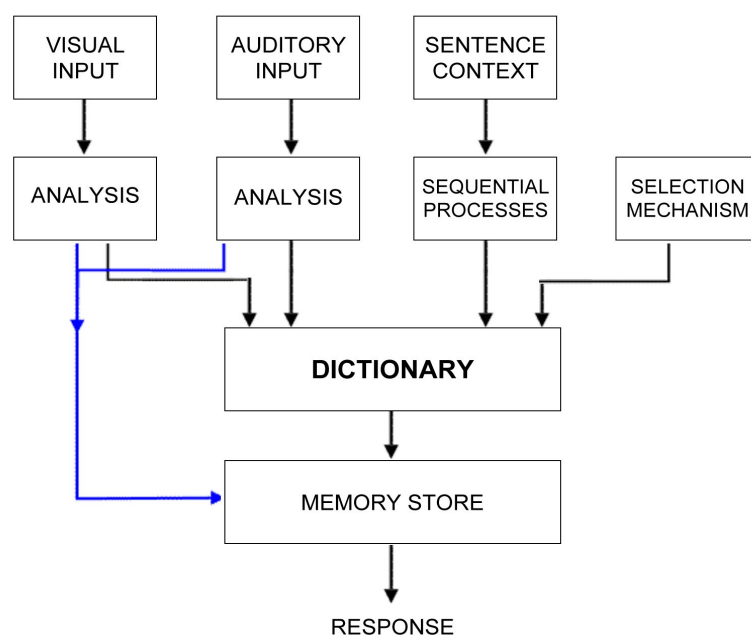


Figura 7: diagramma funzionamento del logogen model

1.1.5.2. Il *cohort model*

Si tratta di una teoria introdotta da William Marslen-Wilson nel 1983 e adottata sia in psicolinguistica che in neurolinguistica per spiegare il funzionamento del riconoscimento lessicale. Questa teoria può essere applicata a tipi di input sia di natura visuale che di tipo auditiva quindi sia scritto che parlato. Si parte dal presupposto che il fonema iniziale della parola stimoli i neuroni che intraprendono la formulazione di candidati in competizione, questi gruppi vengono definiti coorti. Man mano che si procede nella pronuncia il sistema eliminerà quei candidati che si discostano dalla forma presentata fino al riconoscimento finale. Gli studiosi portano a sostegno di questa teoria l'esperimento condotto su dei soggetti cui veniva richiesto di produrre la parola riprodotta dalla macchina, il più velocemente possibile. Risultava che i soggetti ripetessero la parola ancor prima che la macchina avesse finito la riproduzione. I passaggi effettuati per giungere al riconoscimento sono detti "*activation and selection*" o "*recognition and competition*." Questi processi continuano fino al momento in cui non si raggiunge il *recognition point*, a questo punto dopo che tutti i competitori sono stati rigettati rimane solo una parola.

1.1.5.3. La teoria percettiva di Cole e Jakimik

Dalla proposta originale delle coorti si passa ad un modello che rende conto anche del ruolo giocato dal contesto nella generazione ed eliminazione dei competitori. Infatti la conoscenza della grammatica della lingua, nella sua accezione più ampia (dalle regole fonologiche a quelle sintattiche) e la conoscenza del contesto in cui avviene la produzione del segnale aiutano l'ascoltatore a decodificare e in alcuni casi a disambiguare il messaggio. Il segnale viene elaborato seguendo l'ordine lineare della produzione, parola per parola, in un sistema a catena che procede effettuando la stima probabilistica dell'opportunità della comparsa dell'elemento successivo. Concorrono all'individuazione il suono e il significato della parola precedente e il rapporto sintagmatico tra essa e quella da decodificare. La decodifica parte dal piano fonetico fino a quello semantico.

1.1.5.4. L' *Interactive Activation Model* e il TRACE

È anch'esso di tipo connessionista, prodotto da McClelland e Rumelhart tra il 1981-1982. Si basa sul concetto di nodi, essi possono avere tra di loro rapporti di natura eccitatoria o inibitoria. Come per i due modelli presentati qui sopra, la fase di decodifica prende avvio dalla prima lettera della parola che suggerisce tutte quelle parole che iniziano per quella lettera e procede lungo la parola. Un'elaborazione successiva è il TRACE proposto da James

Mc Clelland e Jeffrey Elman nel 1986 è un modello di tipo neurale sviluppato per poter essere utilizzato in associazione con un computer. Si producono così delle simulazioni atte a dare un'immagine di come avvengano i processi di produzione e comprensione del parlato all'interno del cervello umano. Questo modello cerca inoltre di rispondere molte complesse problematiche comprese quelle poste dalla varianza:

- il concetto di continuum fonico ;
- lo scambio di fonemi;
- la coarticolazione;
- la variabilità della lingua in tutte le sue dimensioni (linguistiche e paralinguistiche)

Il riconoscimento della parola segue i seguenti step: nel momento in cui la parola è percepita si attivano nella memoria una serie di parole che hanno la stessa iniziale, così a seguire per le altre lettere che compongono la parola.

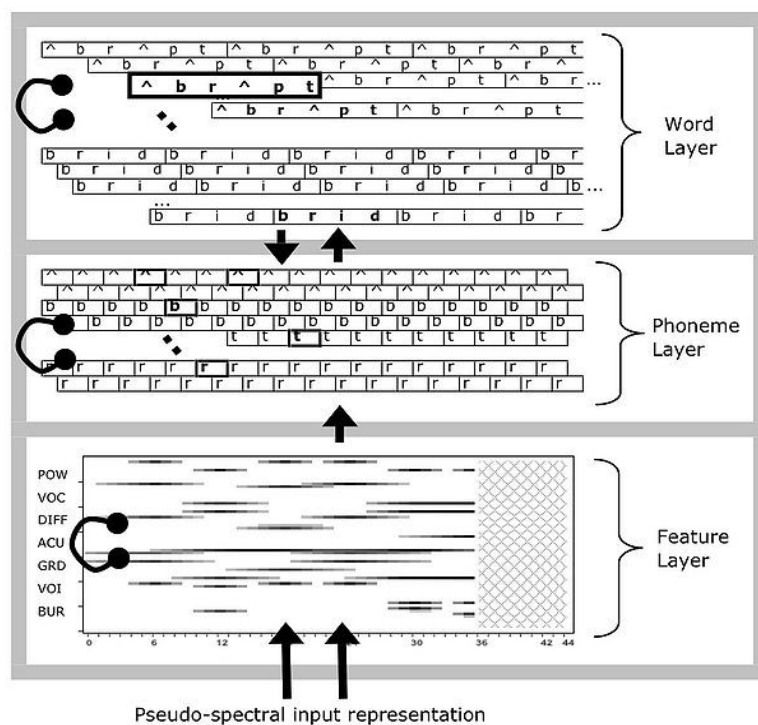


Figura 8: diagramma schematico dell'architettura TRACE

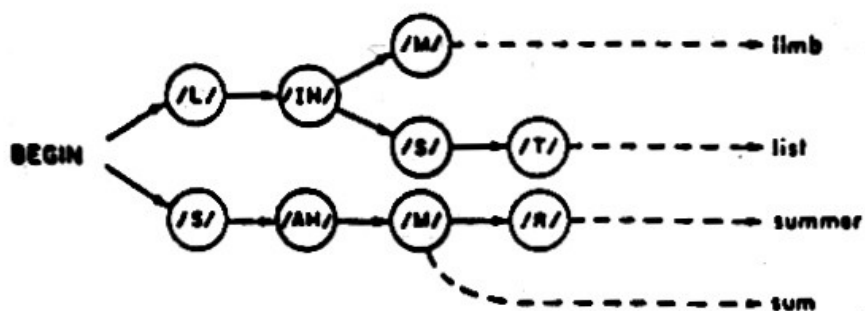
I nodi del TRACE hanno un livello di riposo e una soglia di attivazione. Se l'ingresso lessicale ha un grado di concordanza sufficiente con l'unità che il nodo rappresenta, il livello di attivazione del nodo raggiunge la soglia, se non c'è concordanza, il nodo resta a riposo.

Tra i nodi esiste un'interconnessione e il salire a soglia di uno può implicare gli altri nodi connessi. I rapporti di connessione possono essere di due tipi: eccitatorie o inibitorie e di conseguenza possono esistere tre tipi di rapporti di connessione diversi a seconda che i nodi siano o meno sullo stesso livello. I nodi di livello diverso hanno connessioni progressive eccitatorie bidirezionali: un nodo fonemico può eccitare nodi di parole o sillabe; connessioni laterali inibitorie bidirezionali: l'attivazione di dato fonema inibisce quella di un altro fonema, lo stesso vale per le parole. Infine la connessione di *top-down feedback* dalle parole al fonema. La novità di questa teoria sta nella Traccia, che queste analisi e di conseguenza informazioni lasciano nella memoria a breve termine e che sono a disposizione per le operazioni successive e che quindi possono essere richiamate nel caso si debba disambiguare, o riprendere un input lasciato in sospeso perché poco chiaro. struttura molto più complessa rispetto alle altre unità del sistema, i nodi delle parole non solo riflettono l'attivazione della parola considerata come unità, ma anche l'attivazione di ognuno dei fonemi costituenti e di ognuno dei tratti costituenti i fonemi.

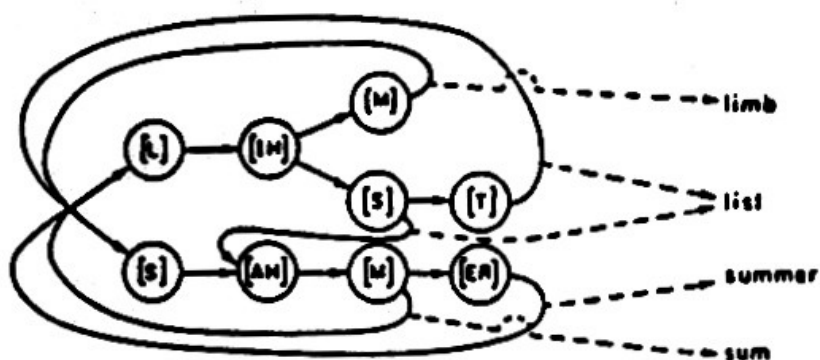
1.1.5.5. LAFS: *Lexical Access from Spectra*

Il modello di Klatt non prevede l'interazione tra i vari livelli di elaborazione, presume un accesso diretto degli input lessicali nella loro forma fisica spettrale. Klatt sostiene che nella memoria a lungo termine del parlante risieda un *tesaurus* di tutte le sequenze spettrali nella forma di difoni (es.CV: /ta/,/da/...) fonotatticamente possibili. Ogni sequenza difonica è associata a un prototipo spettrale, una sorta di stampato. Klatt risponde al problema della variabilità che sorge di fronte a delle rappresentazioni così rigide inserendo nel tesaurus le variazioni contestuali non le istruzioni per esse. Ciò crea una folta generazione soprattutto in contesti di variazione. Il processo di riconoscimento prevede il confronto, da parte del soggetto che percepisce, tra le rappresentazioni spettrali della parola in entrata con quelle prototipiche in memoria; quindi si passa dalla similarità tra la forma dello spettro della parola e la forma dei difoni. Pertanto il riconoscimento si basa solo sull'analisi fonetico-acustica. E' evidente che questo modello non tenga conto di tutte le variabili extra linguistiche, ma ciò è in parte spiegabile con l'uso di tipo ingegneristico che si è fatto; infatti da questo modello deriva l'applicazione della sintesi vocali per difoni (Fig.9).

STEP 1: LEXICAL TREE (PHONEMIC)



STEP 2: LEXICAL NETWORK (PHONETIC)



STEP 3: LEXICAL ACCESS FROM SPECTRA (SPECTRAL TEMPLATES)

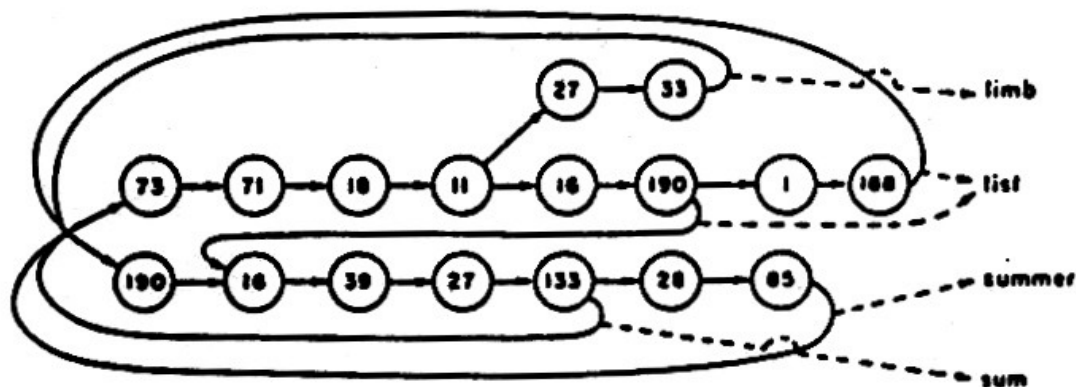


Figura 9 : diagramma a tre livelli del LAFS

1.2. Elementi di neurofisiologia

In questa sezione ci si limiterà fornire, a chi legge questa tesi, senza avere interessi specifici per la neurolinguistica e la neurologia, gli elementi di base per muoversi agevolmente

tra queste pagine. Innanzi tutto dobbiamo dire che la percezione, il pensiero e il movimento sono possibili grazie alla comunicazione tra le cellule neuronali. Nel nostro sistema nervoso ci sono principalmente di due tipi di cellule: neuroni e cellule gliali. Le prime sono cellule funzionali le seconde strutturali. I neuroni come abbiamo appena accennato sono cellule dotate di proprietà tali da renderle fondamentali e specifiche per il compito della segnalazione. I neuroni si distinguono per forma, per funzione, per localizzazione e per le interconnessioni all'interno del sistema nervoso.

1.2.1. Morfologia generale di un neurone

- a) Corpo cellulare: contiene il nucleo e la maggior parte del citoplasma e degli organuli
- b) Dendriti: processi corti e molto ramificati specializzati nella ricezione e nell'invio di messaggi al corpo cellulare
- c) Assone: processo lungo che conduce gli impulsi nervosi lungo il neurone (può anche ramificarsi in diramazioni chiamate assoni collaterali)
- d) Rami terminali: terminazioni dell'assone che a loro volta si ramificano in terminazioni sinaptiche, che secernono i neurotrasmettitori
- e) Nodi di Ranvier: interruzioni tra le cellule di Schwann nelle quali l'assone non è mielinizzato

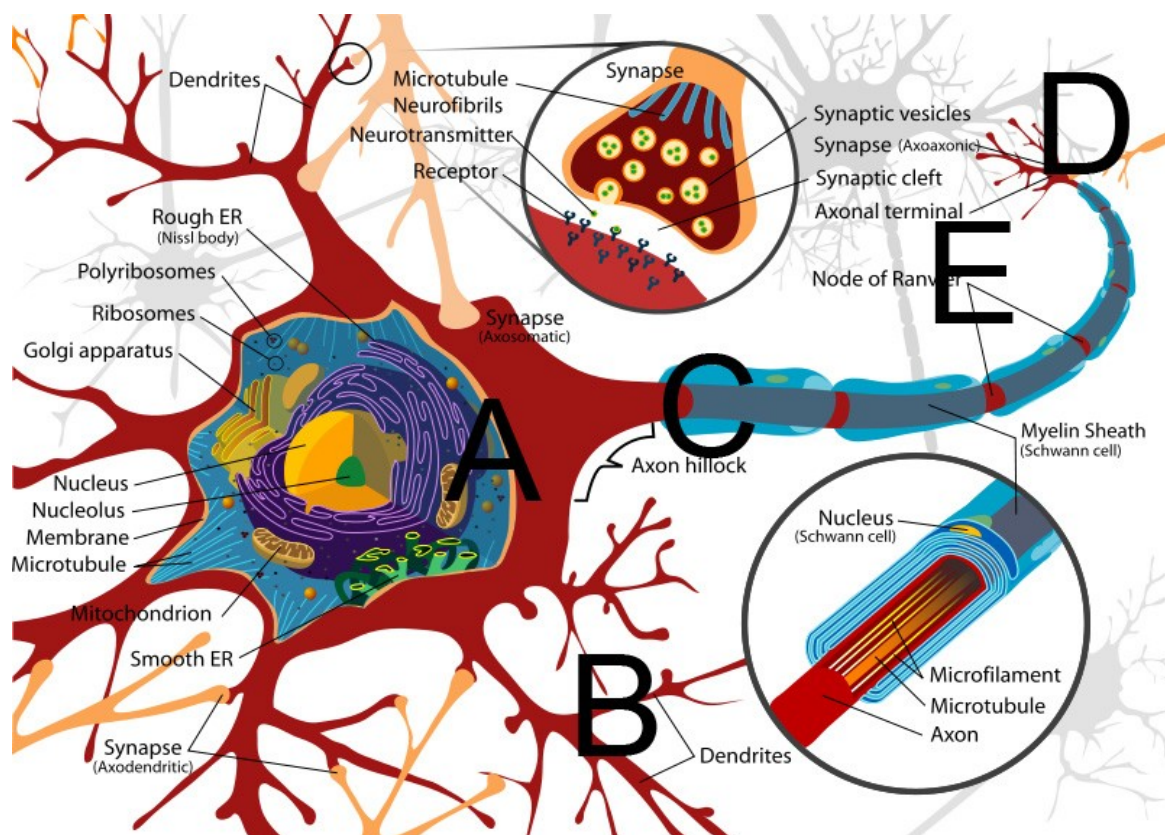


Figura 10: schema del neurone e suo funzionamento

Un neurone è costituito da un corpo detto appunto soma o pironoforo contenente l'apparato metabolico per la produzione ad esempio delle proteine che alimentano la cellula. Il soma a sua volta contiene racchiusi nella membrana plasmatica, sospesi nel citoplasma, quegli organuli intracellulari quali il nucleo, i ribosomi, i mitocondri, l'apparato del Golgi propri della maggioranza delle cellule.

Dal corpo si dipartono due tipi di prolungamenti contenenti citoplasma, i dendriti e l'assone. I primi hanno una forma ramificata più o meno complessa pronta a ricevere gli impulsi provenienti da altri neuroni a livello delle sinapsi. Per questo motivo i dendriti sono detti post-sinaptici. L'assone è invece detto pre-sinaptico poiché si trova all'inizio del flusso d'informazione. Assoni e terminazioni assoniche formano le sinapsi. Queste terminazioni sono dotate di caratteristiche peculiari e specializzate che rendono possibile la comunicazione tra un neurone e l'altro tramite il rilascio dei neurotrasmettitori.

1.2.1.1. Segnalazione neuronale: il flusso di informazioni attraverso il s.n.c.

Ogni neurone comunica con le altre cellule nervose a livello delle sinapsi, che sono giunzioni specializzate tra il bottone di un neurone (cellula presinaptica) e il dendrite di un altro (la cellula post sinaptica (fig. 11)).

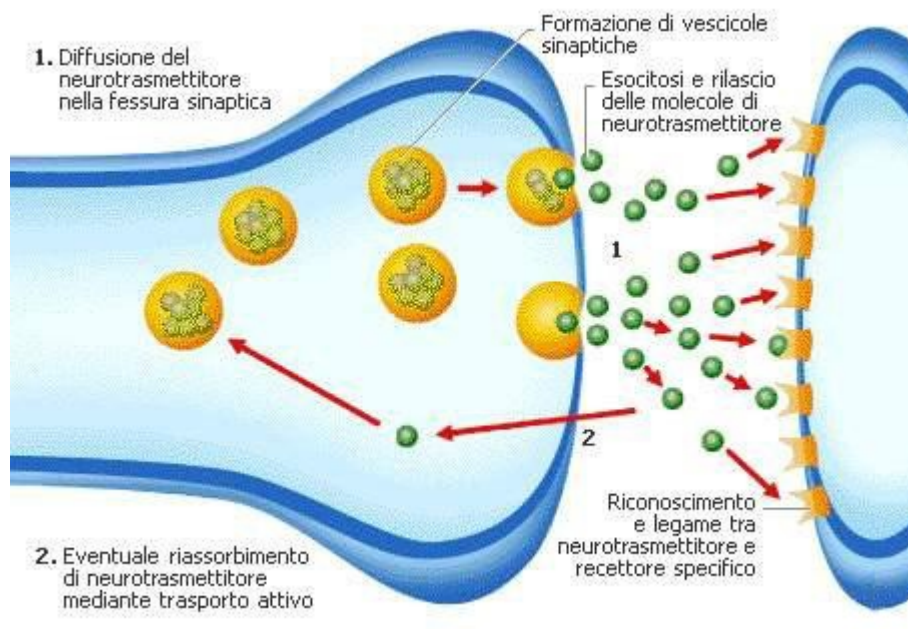


Figura 11: conduzione dell'impulso nervoso

Centinaia di migliaia di assoni tenuti assieme da tessuto connettivo costituiscono un nervo. Il processo di segnalazione è composto da più fasi, vediamole qui di seguito.

- La ricezione: in questa fase i neuroni ricevono il segnale di uno stimolo che può avvenire sotto forma chimica (neurotrasmettitori oppure una sostanza chimica che stimola l'olfatto). Oppure il segnale può essere mediato in una forma fisica per cui avremo il coinvolgimento degli organi di senso, ma anche dei segnali fisici di tipo elettrico recepiti a livello delle sinapsi elettriche.
- La trasmissione: invio di un messaggio attraverso un neurone afferente (o sensoriale), il messaggio viaggia tra 2 neuroni o tra un neurone ed una cellula muscolare/ghiandola, il messaggio è inviato al sistema nervoso centrale (SNC).
- Integrazione: smistamento e interpretazione del messaggio e determinazione della risposta. Avviene nel SNC ad opera degli interneuroni (o associativi), che ricevono il messaggio dai neuroni afferenti.
- Risposta: gli interneuroni trasmettono il messaggio integrato ai neuroni efferenti (o motoneuroni) → Trasmissione. I neuroni efferenti lo trasmettono a loro volta ad un effettore (muscolo, ghiandola, ecc), che mettono in atto la risposta.

I neuroni come abbiamo accennato sono immersi in una soluzione ad elevata concentrazione di ioni sodio (Na^+) e una concentrazione relativamente bassa di ioni di potassio (K^+). Il fluido all'interno del neurone ha invece concentrazioni invertite e in prossimità della membrana cellulare esterna abbiamo una prevalenza di cariche positive mentre all'interno una prevalenza di quelle negative. Si tratta di un'energia potenziale che una volta "liberata" genera un impulso nervoso (Fig. 11). Il potenziale di membrana di un neurone a riposo, ossia non eccitato è detto potenziale a riposo consiste nella differenza di carica elettrica tra i due lati della membrana plasmatica di un neurone ed è uguale a -70 mV .

Il potenziale della cellula a -70 mV , è determinato dalla differenza di concentrazione di ioni (soprattutto Na^+ e K^+) tra l'interno della cellula e il fluido al suo esterno, dalla permeabilità selettiva della membrana a questi ioni (è più permeabile ai K^+ che ai Na^+). A regolare questo equilibrio nel transito degli ioni, si pongono tre canali ionici: i canali ionici passivi, i canali ionici voltaggio-dipendenti, i canali ionici chemio-dipendenti.

I gradienti che determinano il potenziale di riposo sono mantenuti costanti dalle pompe sodio-potassio: ogni 3 Na^+ pompate fuori dalla cellula, 2 K^+ sono pompate dentro.

Vediamo ora più nello specifico come avviene la trasmissione attraverso neurone innanzi tutto l'impulso nervoso è prodotto dal potenziale d'azione a sua volta generato dall'inversione della polarità della carica elettrica all'interno della cellula. I processi alla base di questa inversione di polarità si definiscono: polarizzazione e depolarizzazione. Le fasi di trasmissione sono Stato di riposo- stimolo che causa la depolarizzazione - potenziale d'azione ripolarizzazione e ritorno allo stato di riposo.

Uno stimolo elettrico, chimico o meccanico può alterare il potenziale di riposo. Cioè aprire quei canali posti sulla membrana e far entrare il sodio rendendo il potenziale di membrana (-70 mV) meno negativo, il potenziale elettrico decresce e si dice che la membrana è depolarizzata e subisce un'azione eccitatoria.

Con la depolarizzazione si genera un potenziale d'azione (impulso nervoso). Se uno stimolo rende il potenziale di membrana più negativo, si dice che la membrana è iperpolarizzata e subisce un'azione inibitoria.

L'apertura del primo canale ionico per il sodio voltaggio-dipendente causa l'apertura di quelli adiacenti innescando una reazione a catena: il potenziale di azione si autopropaga come un'onda di depolarizzazione che percorre tutto l'assone. Si tratta di una risposta del tipo tutto o nulla: non esistono variazioni nella forza di un singolo impulso. L'intensità della sensazione provocata è data dal numero di neuroni eccitati e dalla loro frequenza di scarica. La conduzione dell'impulso nervoso si distingue nelle seguenti fasi:

Riposo: ai valori di potenziale di riposo, i canali ionici per il sodio voltaggio-dipendenti sono chiusi e gli ioni Na^+ non possono entrare nella membrana. Depolarizzazione: se uno stimolo depolarizza la membrana al di sopra un determinato valore (valore di soglia = -55 mV ca) i canali ionici per il sodio voltaggio-dipendenti si aprono e il voltaggio impenna fino a 35 mV o anche più (Pike). L'entrata di Na^+ causa un'improvvisa inversione della polarità: la membrana acquisisce carica + all'interno e carica - all'esterno Durante la depolarizzazione l'assone si trova in uno stato di refrattarietà assoluta: l'assone non può infatti trasmettere nessun altro potenziale d'azione Nel potenziale d'azione si presenta la fase di ripolarizzazione: man mano che il potenziale d'azione passa, dietro di esso si aprono i canali ionici per potassio voltaggio-dipendenti. L'uscita di ioni K^+ causa una ripolarizzazione che ripristina la polarità originale. Durante la ripolarizzazione la membrana passa ad uno stato di refrattarietà relativa: l'assone può trasmettere altri potenziali d'azione, ma il valore di soglia è più alto, i canali ionici per il potassio voltaggio-dipendenti rimangono aperti finché non si torna allo stato di riposo.

1.2.1.2. La segnalazione neuronale attraverso le sinapsi

Le sinapsi fungono da giunzione tra due neuroni o tra un neurone ed un effettore.

Il neurone che al livello di una sinapsi ha le proprie terminazioni è detto neurone presinaptico. Il neurone che al livello di una sinapsi ha i propri dendriti è detto neurone postsinaptico.

Le sinapsi possono essere elettriche o chimiche (la maggior parte sono chimiche).

Le cellule pre e postsinaptiche sono separate da uno spazio sinaptico di circa 20 nm che non riesce ad essere superato dal potenziale d'azione.

Il segnale nervoso viene quindi trasportato al neurone postsinaptico attraverso i neurotrasmettitori.

I neurotrasmettitori sono dei messaggeri chimici che portano il segnale nervoso al di là dello spazio sinaptico. Ne esistono oltre 60, eccone alcuni:

- Acetilcolina: effetto eccitatorio sui muscoli scheletrici, inibitorio sul muscolo cardiaco;
- Noradrenalina: coinvolta nella fase REM del sonno, influenza l'umore;
- Dopamina: importante per le funzioni motorie, influenza l'umore;
- Serotonina: eccitatoria per le vie muscolari, inibitoria per le sensazioni, regola l'ingestione del cibo, influenza l'umore;
- Glutammato: funzioni della memoria e dell'apprendimento, eccita le vie di segnalazione del cervello;
- Aspartato: funzioni della memoria e dell'apprendimento, eccita le vie di segnalazione del SNC;
- Glicina: inibisce le vie di segnalazione del SNC.

1.2.1.3. La trasmissione sinaptica

I neurotrasmettitori si trovano nei terminali sinaptici all'interno di piccole vescicole dette vescicole sinaptiche. La trasmissione sinaptica avviene attraverso il rilascio del neurotrasmettitore: quando arriva un potenziale d'azione si attivano dei canali ionici - calcio sensibili al voltaggio, tali canali permettono l'entrata di ioni Ca^{2+} nel terminale sinaptico.

Gli ioni di calcio aiutano le vescicole contenenti il neurotrasmettitore a fondersi con la membrana presinaptica e a rilasciare il neurotrasmettitore nello spazio sinaptico.

1.2.2. Il sistema nervoso centrale (snc)

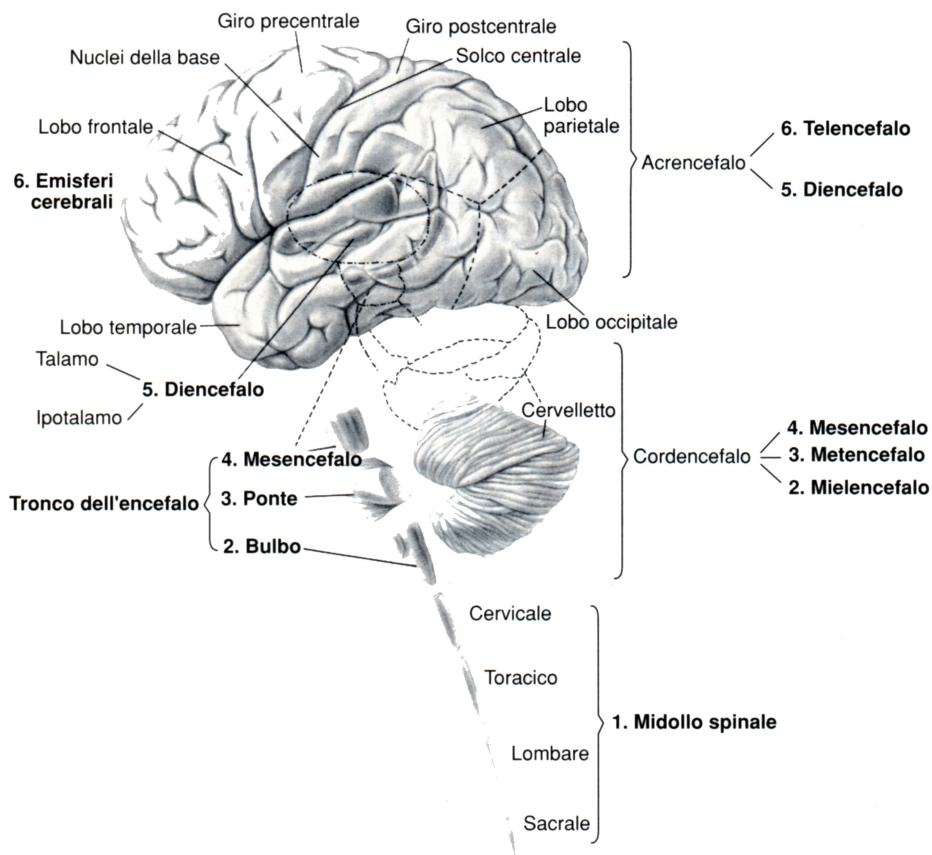


Figura 12: il sistema nervoso centrale

- Le meningi: consistono in tre strati di membrane protettive costituite da tessuto connettivo che ricoprono e proteggono l'encefalo e il midollo spinale
- Dura madre: strato più esterno e resistente
- Aracnoide: strato intermedio
- Fluido cerebrospinale (CSF): contenuto nello spazio subaracnoideo, protegge l'encefalo ed il midollo spinale dai traumi
- Pia madre: ultimo strato, sottile e vascolarizzato adeso al tessuto cerebrale ed al midollo
- Il midollo spinale si estende dalla base dell'encefalo fino alla seconda vertebra lombare. Il suo compito è quello di trasmettere informazioni da e per l'encefalo la sua struttura è la seguente:
- Canale centrale: piccolo canale circondato da un'area composta di materia grigia a forma di H

- **Materia grigia:** materia composta da grandi ammassi di corpi cellulari, dendriti, assoni non mielinizzati e cellule gliali
- **Materia bianca:** situata all'esterno della materia grigia, è composta da assoni mielinizzati raccolti in fasci detti tratti o vie.

L'encefalo è composto da:

- **Tronco encefalico:** su di esso poggia il cervello. È connesso al midollo spinale e al cervelletto.
- **Talamo:** situato sulla sommità del tronco encefalico, svolge la funzione di centro di smistamento delle informazioni tra midollo spinale e cervello. I neuroni del talamo selezionano e interpretano tutte le informazioni afferenti, eccetto quelle olfattive. Tra le sue funzioni principali c'è la regolazione dello stato di coscienza
- **Ipotalamo:** è situato subito sotto il talamo, connesso con l'ipofisi, la regola il centro di controllo della temperatura, dell'appetito, del metabolismo dei grassi e di alcune emozioni.
- **cervelletto:** è il centro per la coordinazione ed il perfezionamento dei movimenti muscolari.

Ma vediamo ora più nello specifico la **corteccia cerebrale**: si tratta dello strato superficiale del cervello costituita dalla materia grigia, una massa non regolare con anse e circonvoluzioni che permettono così di aumentarne la superficie. Su di essa abbiamo i solchi, delle pieghe superficiali nella materia e delle fessure che sono più profonde. Al di sotto della materia grigia vi è la materia bianca. Essa si prolunga nel midollo spinale. Tra le sue funzioni vi è quella di connettere i due emisferi tramite un fascio detto corpo calloso. Infine abbiamo i gangli basali costituiti dalla materia grigia che ricevono segnali dalla substantia nigra. Ai fini della nomenclatura generale possiamo far riferimento a tre assi: orizzontale, sagittale e frontale. In base alla direzione e al piano essi si distinguono ulteriormente così da avere per il piano orizzontale le porzioni dorsali e ventrali; il piano coronale diviso in rostrale e caudale; infine il piano sagittale che divide l'encefalo nelle due metà, sinistra e destra. Da questa divisione deriva anche la nomenclatura utilizzata per il montaggio della cuffia EEG. Gli emisferi si differenziano dal punto di vista strutturale e come vedremo alcune di queste differenze sono la spia di differenze anche sul piano funzionale.

Si ricorda innanzi tutto che l'emisfero destro controlla la parte sinistra del corpo, mentre il

sinistro controlla la parte destra del corpo. Possiamo individuare i lobi presenti per ogni emisfero, distinguibili in **frontale**; in corrispondenza della fronte; **temporale**; subito sopra le orecchie, zona dell'udito; **parietale**; parte posteriore alta; **occipitale**; parte posteriore bassa, zona della vista. Quindi le **aree**: le sensoriali ricevono segnali dagli organi di senso, le motorie controllano i movimenti volontari, le associative sono addette alla connessione tra aree motorie e sensoriali. Riguardo l'asimmetria strutturale-funzionale, citata sopra, dobbiamo far presente che per . l'area del linguaggio (che vedremo nello specifico più avanti) nell'emisfero sinistro la scissura di Silvio è molto più lunga e il suo dispiegamento porta ad un' estensione maggiore alla zona specializzata nel linguaggio.

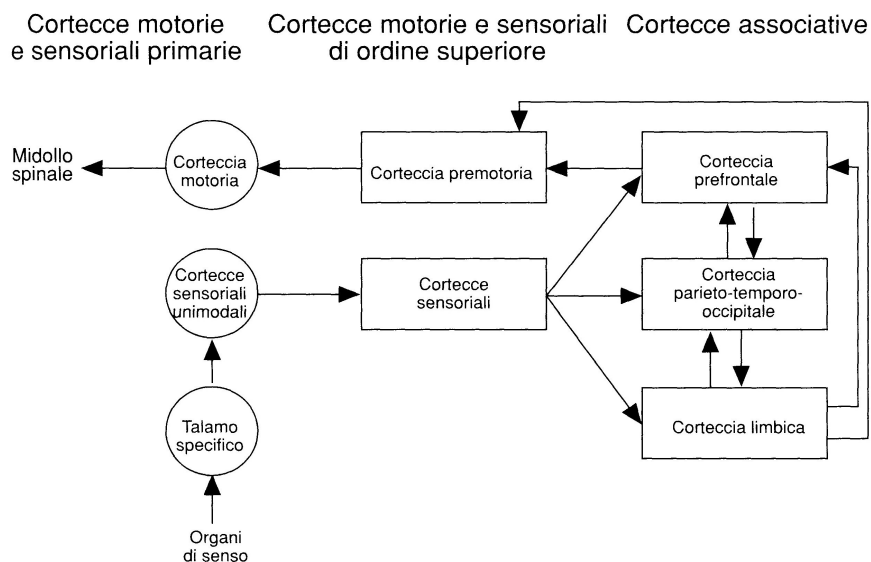
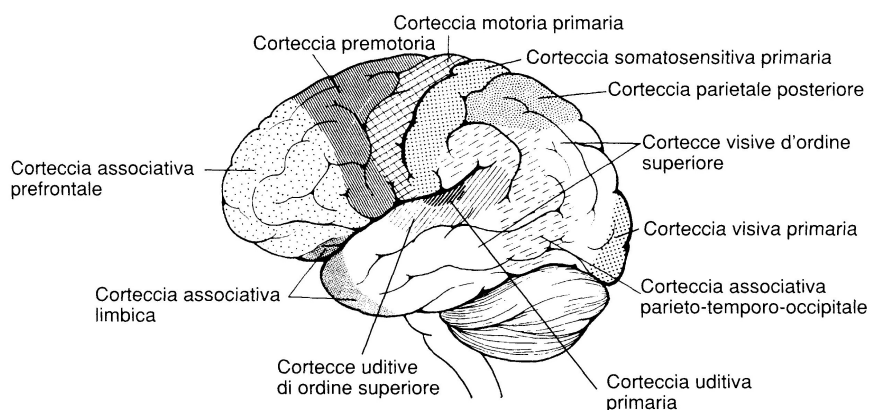


Figura 13: divisione delle aree che compongono le cortecce motorie

1.2.2.1. Area di Broca e Area di Wernicke

Queste due aree sono state collegate, dai due studiosi dai quali prendono il nome, P.P. Broca e C. Wernicke, alla produzione ed elaborazione del linguaggio.

Nel caso di Broca egli studiò l'afasia occorsa in due pazienti con lesioni nella regione del giro frontale inferiore sinistro (in Brodmann aree 44 e 45). Per lungo tempo si è ritenuto che lesioni in questa area compromettessero solo la produzione, ma in recenti studi che adottano l'fMRI si è scoperto che i pazienti erano incapaci di comprendere il significato di frasi utilizzando le informazioni sintattiche e di disambiguare (fig. 14).

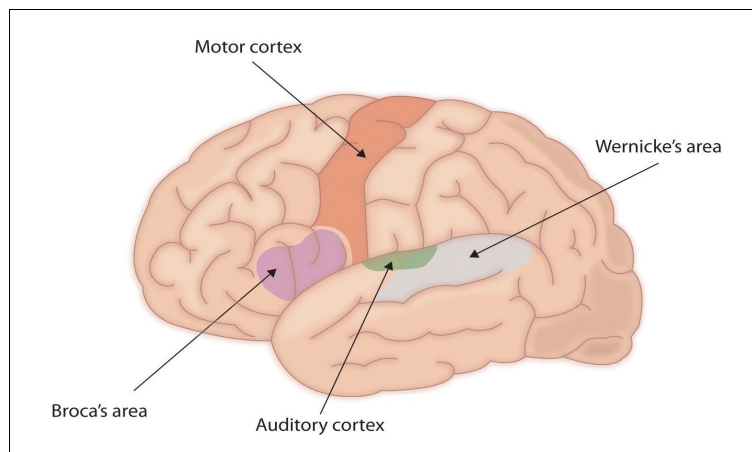


Figura 14: aree coinvolte nella produzione e interpretazione del linguaggio

1.3. L'elettroencefalografia

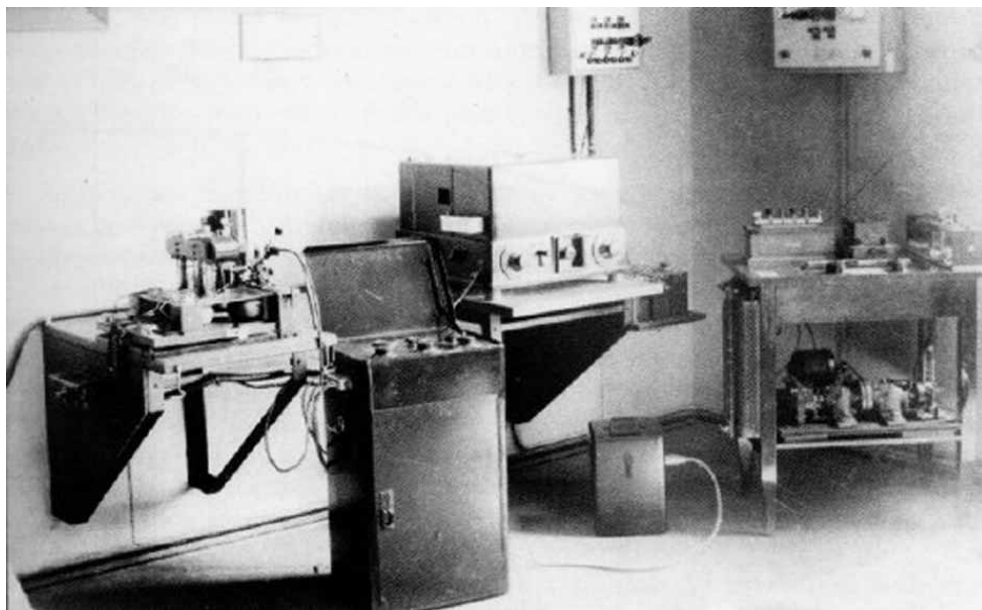


Figura 15: primo elettroencefalografo

Questa tecnica deve le sue origini agli studi di Hans Berger, e fu in seguito perfezionata da Herbert Jasper intorno agli anni '20.

Durante i suoi esperimenti Berger scoprì come tra due aghi infissi nello scalpo vi fosse una differenza di potenziale elettrico direttamente correlata all'attività cerebrale. Ad oggi è una tecnica non invasiva che permette di misurare l'attività elettrica superficiale, di correnti intracellulari, rilevabile dalla superficie dello scalpo, permettendo di monitorare l'attività elettrica del cervello nel suo divenire temporale e spaziale, così da poter studiare e "segmentare" ad esempio le fasi di elaborazione di un vocabolo letto.

Ma cosa misura l'EEG? Misura il flusso di corrente extracellulare che viene generato dalla somma delle attività di un elevato numero di neuroni. L'elettrodo registra in superficie l'attività dei neuroni corticali piramidali disposti in corrispondenza dell'area corticale sottostante l'elettrodo. Ciò che viene rilevato è la differenza di potenziale elettrico tra un elettrodo attivo, posto al di sopra della sede dove si svolge l'attività neurale, e un elettrodo non-attivo di riferimento, collocato ad una certa distanza dal primo. Le onde EEG sono caratterizzate da una frequenza, espressa in cicli per secondo (Hz) e da una ampiezza, espressa in microvolt (μV). La registrazione dei segnali viene fissata tramite rappresentazione grafica (detta elettroencefalogramma) su supporto cartaceo o un file in formato digitale. In particolare nel nostro caso abbiamo ad esempio tre file: uno con estensione "*.vhdr" per i settaggi e degli elettrodi in registrazione, uno con estensione "*.vmrk" per i *trigger* relativi agli stimoli e infine uno con estensione "*.eeg" che contiene il segnale corrispondente al tracciato EEG. Un normale tracciato EEG si compone di una somma di ritmi a frequenze diverse compresi nella banda 0-70Hz con un'ampiezza massima di $30\mu\text{V}$. Questa tecnica ha come vantaggio una risoluzione temporale elevata, siamo nell'ordine dei microsecondi; mentre la risoluzione spaziale è piuttosto bassa, circa 10-20mm, ciò è causato soprattutto dalle distorsioni che subiscono i segnali elettrici nel giungere allo scalpo dalle zone di origine sottostanti. Per correggere queste distorsioni si utilizza l'ICA, un processo basato su di un algoritmo che cerca di pervenire ad una stima del segnale di un elettrodo scomponendolo in componenti che influiscono su di esso, riuscendo così ad identificare meglio la sorgente. Vedremo meglio di cosa si tratta nella sezione dedicata al trattamento del segnale.

1.3.1. Strumentazione e funzionamento

Per la registrazione di un tracciato con elettroencefalografo avremo degli elettrodi posizionati sulla testa del soggetto collegati ad un amplificatore, questo amplificatore è collegato ad un convertitore analogico/digitale (A/D) per poi inviarlo al computer. Una volta sul

compute la registrazione verrà registrata e visualizzata in tempo reale. Potrà essere archiviata per essere analizzata e visionata successivamente oppure essere stampata come nel caso delle refertazioni.

1.3.1.1. Gli elettrodi

Per la registrazione vengono utilizzati dei piccoli elettrodi di materiale conduttore (placche di argento rivestite elettroliticamente da cloruro d'argento), posizionati sul cuoio capelluto, prima si sarà proceduto con la rimozione delle cellule epiteliali morte e lo sgrassaggio della cute al fine di ridurre l'impedenza; per aumentare la conducibilità viene aggiunta una pasta di composizione salina di tipo abrasivo, ma ciò dipende essenzialmente dal tipo di elettrodi che vengono utilizzati e dal tipo stato in cui si trova il soggetto sottoposto all'analisi. Proprio per questo gli elettrodi possono presentarsi sotto diversa forma. Quelli a ponte sono utilizzati in diagnostica e il posizionamento è effettuato a mano mentre il fissaggio-mantenimento dagli elettrodi è ottenuto mediante delle guide di silicone.

Nel caso degli elettrodi a coppetta li vediamo montati liberi da cuffia e solitamente attraverso pasta adesiva, questa tecnica viene impiegata in pazienti in età infantile e neonatale o nel caso di monitoraggi continui. (fig. 16, 17, 18). Gli elettrodi si presentano spesso con cavi individuali, nelle cuffie di nuova generazione, invece, possiamo trovare un fascio di fibre che esce direttamente dalla cuffia che termina con una presa.(Fig. 19).



Figura 16: elettrodo ad ago



Figura 16: elettrodo a coppetta



Figura 18: elettrodo a ponte



Figura 19: cuffia con elettrodi premontati

1.3.1.2. Il posizionamento

Il nome e il posizionamento degli elettrodi è stabilito dallo Standard Internazionale **10-20**: questa denominazione deriva dal valore percentuale delle distanze che separano gli elettrodi; infatti la distanza che separa gli elettrodi l'uno dall'altro può essere del 10% o del 20% rispetto al 100% cioè la distanza tra la prominente alla base dell'osso occipitale e l'attaccatura superiore del naso (da 30 a 36 cm) (Fig.20).

La cuffia da noi utilizzata per il 1° e 2° procollo ha 64 elettrodi posizionati, 2 di massa (posizionati sui lobi dell'orecchio) i restanti sono posizionati lungo cinque linee: 1) **P1**: longitudinale esterna e 2) **P2**: long. interna di destra, 3) centrale, 4) **P1**: longitudinale esterna e 5) **P2**: long. interna di sinistra. La linea trasversale, che va da un orecchio all'altro (T4-C4-Cz-C3-T3) è detta montaggio **P3**. Ogni sigla è correlata alla posizione che ha l'elettrodo sullo scalpo: sono formate da una/due lettere, Fp (frontopolare), F (frontale), C (centrale), P (parietale), T (temporale), O (occipitale), e sono seguite da un numero (o una z) che identifica l'emisfero (numeri dispari: sinistra; numeri pari: destra; z: linea mediana).

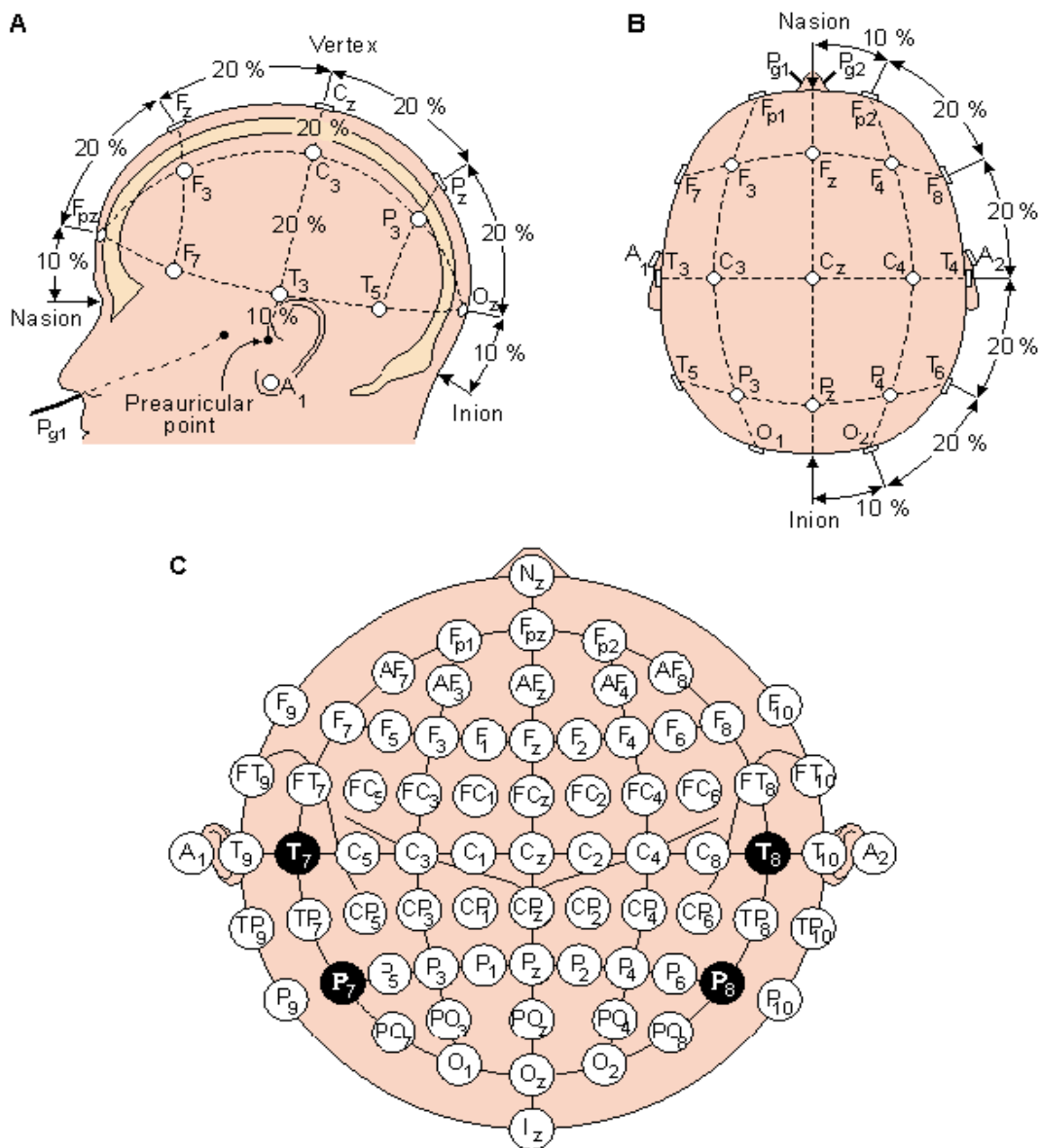


Figura 20: standard di montaggio elettrodi 10-20

La registrazione e di conseguenza la rappresentazione del segnale varia a seconda del tipo di settaggio effettuato (*montage*), cioè scegliendo quale differenza di potenziale vada registrata. Per i nostri esperimenti è stato scelto un montage monopolare per cui ciascun canale rappresenta la differenza di potenziale tra un elettrodo sullo scalpo e uno posto in una zona inerte (i lobi dell'orecchio) denominato *referenza*.

1.3.2. I ritmi

Il risultato della registrazione sarà un elettroencefalogramma che rappresenta le onde cerebrali.

- **Onda Alfa** ("ritmo di Berger"), si distingue in alfa *lento* (8-9 Hz, alfa *intermedio* (9-11.5 Hz) ed alfa *rapido* (11.5-13 Hz), ed ha un'ampiezza media di 40-50 μV che viene registrato ad occhi chiusi in un soggetto sveglio e negli istanti precedenti l'addormentamento. Una delle caratteristiche delle onde alfa è la loro configurazione regolare e sincronizzata. All'apertura degli occhi, l'attività alfa scompare ed è sostituita da un'attività di basso voltaggio e più rapida, di tipo beta, questo fenomeno è detto desincronizzazione.
- **Onda Beta o ritmo beta** è distinto in beta *lento* (13.5-18 Hz) e beta *rapido* (18.5-30 Hz) possono raggiungere i 35 Hz in caso di forte attività. Le onde beta sono dominanti in un soggetto ad occhi aperti.
- **Onde Delta o ritmo delta:** sono caratterizzate da una frequenza che va da 1 a 4 Hertz. Sono presenti nei bambini e in adulti con forte stress o affetti da disturbi metabolici.
- **Onde Theta o ritmo theta:** vanno dai 5 agli 8 Hertz, caratterizzano gli stadi 1 e 2 del sonno.

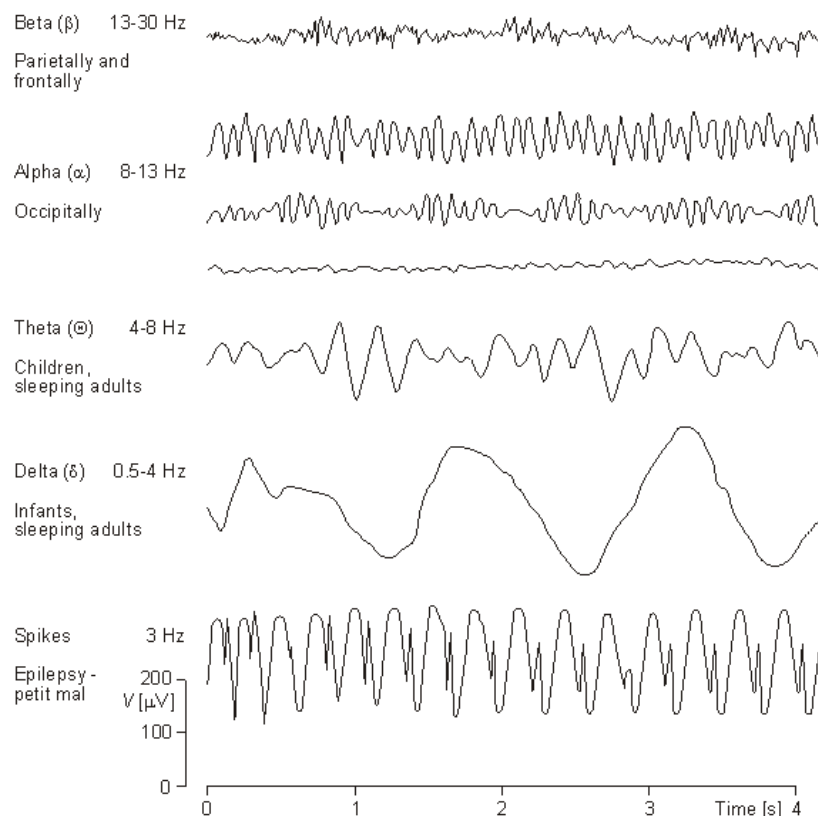


Figura 21: ritmi EEG

Vi sono inoltre dei segnali che appaiono come distorsioni del ritmo: gli **Erp**, sono potenziali elettrici che vengono registrati mediante EEG, all'apparire di uno stimolo sia esso visivo che uditivo. La comparsa sul tracciato è in sincrono con la presentazione dello stimolo, al fine della elicitazione sono formulati differenti paradigmi di somministrazione. Più in generale rispondono tutti alla logica per cui si misura la differenza di potenziale elettrico tra il treno di stimoli categorialmente uguali e lo stimolo nuovo.

1.3.3. Problematiche

Gli artefatti biologici sono dei segnali che vengono registrati dall'EEG, ma non sono prodotti dal cervello e inquinano il tracciato.

- movimento degli occhi e saccadi
- EKG (cardiac) il battito cardiaco
- EMG (*muscle activation*) movimenti della lingua

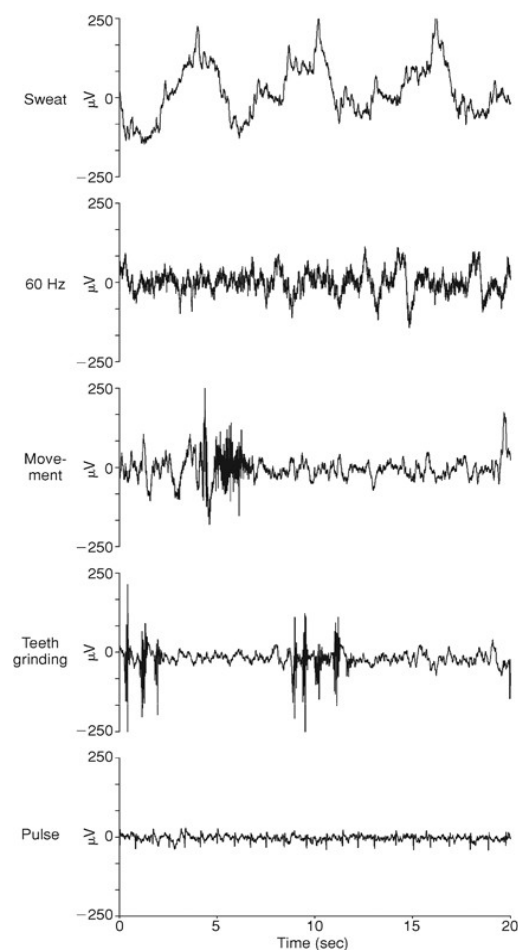


Figura 22: esempi andamenti di artefatti

Oltre agli artefatti prodotti dall'interno, altri provengono dall'esterno, gli artefatti ambientali. I movimenti del soggetto, la sudorazione del cuoio capelluto durante la registrazione, ma anche le interferenze di campi magnetici nella stanza di registrazione.

1.4. Altre tecniche di indagine

Offriamo qui di seguito una panoramica degli strumenti che a oggi vengono adottati per le ricerche in campo neurolinguistico, riportando anche a margine gli studi di maggior interesse. Ci soffermeremo in particolare sull'elettromiografia in quanto tecnica concorrente nella corsa alla decodifica del parlato non-parlato.

1.4.1. La TMS

Stimolazione magnetica transcranica (TMS) è una pratica non invasiva di stimolazione dei neuroni: utilizza campi magnetici rapidamente varianti per attivare la corteccia cerebrale con una risoluzione temporale inferiore ad un secondo, ed una spaziale di 1 centimetro, ha capacità d'indurre un'area di depolarizzazione della profondità di "solamente" 1.5 - 2 cm sotto la superficie dello scalpo. L'impulso porta alla generazione, temporanea e reversibile, di una lesione (**lesione virtuale**). La TMS stimola le aree motorie, le risposte elettromiografiche indotte possono essere registrate a livello epidurale o meno invasivamente a livello del muscolo target (attraverso *elettromiografo*) e costituiscono un potenziale motorio evocato (MEP). In questo modo possono essere esaminate in dettaglio le proprietà di propagazione delle fibre cortico-spinali che controllano il muscolo target, e mapparne la rappresentazione corticale rendendo la TMS una tecnica efficace anche per lo studio della plasticità cerebrale (p.es. nel recupero *post-stroke*).

1.4.1.1 Il funzionamento

L'apparecchio è costituito da un generatore di scariche elettriche e da una bobina (*coil*) che viene posta a contatto con lo scalpo del soggetto stimolato. Il principio alla base del funzionamento di questo apparecchio è l'Induzione elettromagnetica. Al passaggio di corrente nella bobina si genera un campo elettromagnetico perpendicolare al flusso stesso. Lo scalpo, anch'esso all'interno del medesimo campo magnetico, fungerà da conduttore e al suo interno verrà indotta corrente: essa produrrà dei potenziali di azione in neuroni eccitabili della corteccia. Gli effetti non dipendono dal campo magnetico ma dal campo elettrico indotto. Il campo elettrico generato dipende da diversi fattori, primo fra tutti i tipi di bobine. La forma delle bobine varia a seconda dello studio o indagine/terapia che si vuole ef-

fettuare. Le bobine si differenziano per: il materiale di cui sono fatte, per la configurazione geometrica delle spire e infine per le caratteristiche biofisiche che ogni specifica bobina produce con il suo campo magnetico. Ciascuna differente bobina produce un diverso tipo di campo magnetico:

- Bobina circolare: fu la prima ad essere introdotta,
- Bobina a"8" (o *butterfly coil*): il più impiegato oggi, genera un campo elettrico più focalizzato;
- Bobina a doppio cono: ricalca la curva della testa ed è utilizzata per la stimolazione profonda,
- Deep TMS (o H-coil): utilizzata in ambito clinico per il trattamento di pazienti affetti da depressione cronica.
- E' possibile controllare gli impulsi attraverso il generatore. I metodi di stimolazione sono nella sostanza tre:
 - singolo impulso: singole scariche ad intervalli >3sec;
 - forma d'onda: monofasica o bifasica;
 - treno di impulsi o rTMS: l'impulso consiste in un n° di scariche in un tempo dato sotto 1 Hz (slow TMS), sopra 1 Hz (rapid-rate TMS).
- Gli effetti della TMS si possono dividere in due tipologie strettamente dipendenti dal tipo di stimolazione effettuata:
 - Impulsi singoli o a coppia. Se usati sulla corteccia motoria primaria producono dei potenziali motori evocati (MEP) ,come ad esempio la contrazione dei muscoli della mano, che possono essere registrati attraverso EMG. Se usati sulla corteccia occipitale provocano flash di luce che solo il soggetto sottoposto può vedere. Stimolando altre aree della corteccia i soggetti non sono consci di quello che gli sta accadendo e dell'alterazione dei propri comportamenti (es. incapacità di produzione di un eloquio fluente o con significato). Gli effetti non perdurano al termine della stimolazione.
- rTMS: produce effetti che vanno oltre il periodo di stimolazione. L'impiego è quindi di tipo terapeutico, ad esempio nei pazienti affetti da depressione cronica in quanto questa stimolazione aumenta la produzione di endorfina e serotonina.



Figura 23: 8-Coil



Figura 24: applicazione della TMS

1.4.1.2. Caratteristiche

Produce una sensazione di formicolio alla scarica dell'impulso e i rischi sono molto bassi eccetto che per i pazienti con epilessia con impianti. Non sono conosciuti, ad oggi, effetti negativi permanenti. Tutti gli esperimenti debbono essere condotti all'interno di una struttura medica e sotto il controllo di personale medico.

1.4.1.3. Applicazioni in neurolinguistica

La TMS è forse tra le tecniche più usate per l'indagine linguistica più nello specifico fonetica, sicuramente è quella che ha dato apporto più significativo a sostegno della teoria motoria. Sono di questi ultimi anni le risposte che provengono da questa metodologia adottata con costanza nelle ricerche dell'equipe del Prof. Fadiga. Vogliamo ricordare come in più occasioni è stato dimostrato come il sistema motorio giochi un ruolo fondamentale nella decodifica del parlato sia in ambiente rumoroso che non. Di notevole importanza sono i dati acquisiti sul ruolo dell'area motoria per la discriminazione del luogo e modo di articolazione come per la modulazione del *pitch* nelle vocali.

1.4.2. La fMRI: la risonanza magnetica funzionale

Questa tecnica visualizza la risposta emodinamica cioè i cambiamenti nel contenuto di ossigeno del parenchima e dei capillari, correlata all'attività dei neuroni nel cervello.

Le cellule nervose attive, consumano ossigeno trasportato dal sangue che irrorata tramite una rete di capillari tutto il cervello. Ad una maggiore attività cerebrale corrisponde una maggiore richiesta di ossigeno e di conseguenza un maggior afflusso di sangue. L'afflusso avviene con un ritardo, da 1 a 5 secondi circa, dal momento della richiesta. L'emoglobina quando ossigenata ha una proprietà detta diamagnetica ossia: esposta ad un campo magnetico prende la polarità opposta al campo. E' pertanto possibile visualizzare la parte del cervello coinvolta in una particolare attività come può essere il movimento di un arto.



Figura 25: apparecchiatura per fMRI

1.4.2.1. Il funzionamento

Lo scanner ha la forma di grande cilindro, inseriti al suo interno ci sono degli anelli metallici di materiale magnetico estremamente conduttivo: al passaggio dell'energia generano un forte campo magnetico. Questo campo fa sì che le particelle di ossigeno si polarizzino e soprattutto una direzione uniforme. A questo punto le spire che avvolgono il cilindro emettono un segnale radio che eccita i protoni, facendoli risuonare alla stessa frequenza. I protoni così stimolati emettono un segnale che sarà tanto più forte tanto più saranno vicini al sangue ossigenato. Aumentando la potenza del campo magnetico aumenta la localizzazione e la visibilità. Il tempo di acquisizione si aggira intorno al secondo per *frame* e la risoluzione spaziale è di circa 4 -1 mm.

1.4.2.2. Caratteristiche

Innanzitutto è una tecnica non invasiva, ha un'ottima risoluzione spaziale a fronte di una modesta risoluzione temporale che non dipende strettamente dallo scanner ma dal tipo di processo analizzato; infatti se l'EEG misura l'attività elettrica, in sé molto rapida, l'fMRI misura la risposta emodinamica per natura più lenta.

Nel tempo sono state avanzate delle critiche a questa tecnica:

- si tratta di una misura indiretta dell'attività neurale e la risposta potrebbe essere dettata da fenomeni fisiologici non neurali.
- Aree differenti del cervello potrebbero avere diverse risposte emodinamiche, che il modello lineare non descriverebbe opportunamente.

è spesso usata per visualizzare l'attivazione localizzata in regioni specifiche, senza rappresentare adeguatamente la natura distribuita del funzionamento delle reti neurali.

I soggetti possono essere sottoposti a vari test come visione di video, ascolto di suoni, percezione di odori, esecuzione di operazioni cognitive come memorizzare o immaginare qualcosa, premere alcuni pulsanti o altri tipi di compiti. A causa del forte campo magnetico (anche ad apparecchio spento) i soggetti non devono avere oggetti ferromagnetici (orologi, occhiali, pacemaker, viti ossee etc.) che renderebbe pericoloso anche solo l'avvicinamento allo scanner.

1.4.2.3. Applicazioni in neurolinguistica

Le ricerche che utilizzano questa tecnica, in Italia, si focalizzano su la neurolinguistica del bilinguismo e la rappresentazione delle diverse lingue nel cervello di soggetti bilingui o poliglotti. La ricerca già effettuata conferma l'importanza per tutte le lingue in soggetti poliglotti delle zone del cervello relative alla linguaggio, situate nell'emisfero di sinistra. Tuttavia, con l'uso di paradigmi linguistici sperimentali specifici, è stato possibile scoprire le differenze neurali legate alle variabili chiave, quali l'età di acquisizione, il livello di competenza ed il grado di esposizione a lingue differenti. Inoltre si sono dimostrate le attivazioni in strutture del cervello situate fuori delle zone classiche di linguaggio. Gli studi esteri si sono invece focalizzati sulla L1 e in particolare sulla localizzazione e individuazione delle aree deputate a determinati compiti linguistici.

Dallo studio Wildgruberé (1999) sembrerebbe che l'area 22/42 dell'emisfero destro sia deputata al compito di percezione della prosodia *affettiva*; la stessa area sembrerebbe coinvolta nella produzione della stessa tipologia di prosodia. Purtroppo con questa tecnica si riesce ad evidenziare solo il luogo di attivazione ma non l'intero network, quindi le varie connessioni e momenti di svolgimento di un compito.

1.4.3. La PET: Tomografia ad emissione di positroni

E' una tecnica che produce immagini tridimensionali o mappe dei processi funzionali all'interno del del nostro corpo. Questa tecnica come l'fMRI utilizza l'emodinamica per visualizzare e mappare, nel nostro caso nel cervello, le aree che maggiormente si attivano durante l'esecuzione di un compito specifico.

1.4.3.1. Il funzionamento

Al soggetto sottoposto ad esame viene somministrato per via endovenosa, e in rari casi per inalazione, un radio-isotopo con emivita breve, i radionuclidi sono incorporati in composti metabolizzati dall'organismo umano, di preferenza si sceglie il glucosio. Il soggetto viene posizionato sotto lo scanner dopo che il glucosio e di conseguenza il radioisotopo siano entrati nei tessuti da analizzare. L'isotopo di emivita vita media decade, emettendo un positrone; questo percorrerà pochi millimetri prima di collidere con un elettrone. Dalla collisione si spigioneranno una coppia di fotoni gamma che prenderanno due direzioni. Questi fotoni sono rilevati da un contatore: lo scintillatore. Le scintille così rilevate verranno utilizzate da un calcolatore che ricostruirà l'immagine delle aree studiate(fig. 23). Poche strutture sono dotate di questa tecnologia di indagine; infatti gli isotopi a causa della loro

emivita breve devono essere prodotti da un ciclotrone (apparecchio di per se molto costoso che serve per accelerare gli ioni) presente nella struttura dove verrà effettuato l'esame. Ad oggi le aziende sanitarie cercano di ovviare a questa spesa della produzione "casalinga", piuttosto rilevante dal punto di vista economico, con l'acquisto di preparati confezionati da aziende del settore nucleare. E' quindi una tecnica molto costosa e piuttosto invasiva.



Figura 26: macchina PET

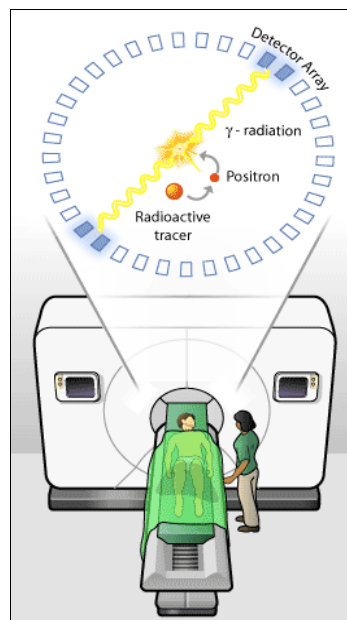


Figura 27: schema di funzionamento PET

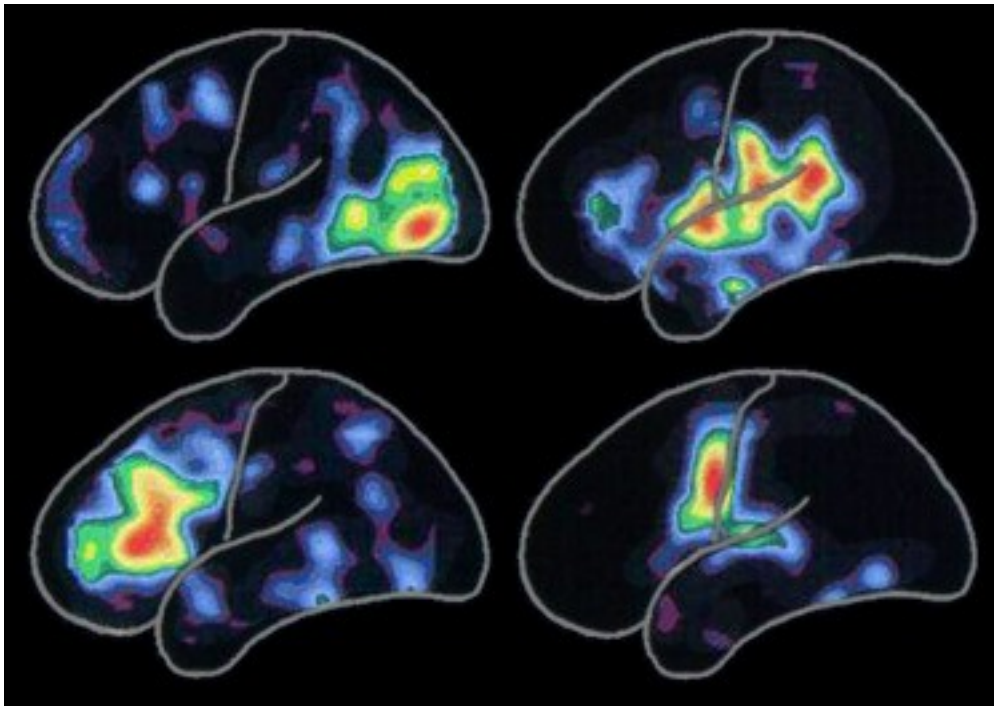


Figura28: esempio di scansione PET

1.4.3.2. Neurolinguistica e PET

Un esempio di applicazione è lavoro di A. Moro, (Moro, 2006) che attraverso una serie di esperimenti condotti su soggetti sottoposti a stimoli di frasi grammaticali e agrammaticali, e le cui reazioni erano registrate mediante PET, sembrerebbe dimostrare l'esistenza di una programmazione biologica che guida la struttura delle nostre lingue. Infine, che le grammatiche "sintetiche", costruite *ad hoc* per violare tutte le regole delle lingue conosciute, violerebbero dei principi biologici nella mente umana e sarebbero estranee al nostro programma genetico. Recentemente, questa tecnica è stata utilizzata per indagare i processi neurali quando si è sottoposti ad un parlato mascherato (da rumore, da altro parlato e da un rumore di parlato con spettro rotato). Le immagini dimostrano che il parlato mascherato da altro parlato comporta un'attivazione bilaterale del giro temporale superiore rispetto alla *baseline* del parlato mascherato da rumore, mentre per il rumore con spettro rotato abbiamo la sola attivazione del giro temporale superiore destro rispetto sempre alla *baseline*. Per meglio chiarire il concetto di *baseline* ricordiamo che gli esperimenti che utilizzano la PET prevedono dei protocolli complessi e che i risultati si basano differenza di immagini, tra una condizione di riferimento e una di attivazione; per cui bisogna prevedere uno stimolo base (*baseline*) poco complesso e uno più complesso.

1.4.4. MEG

La magnetoencefalografia è una tecnica di *imaging* funzionale utilizzata nella neurologia, che si basa sulla misurazione dei campi magnetici prodotti dall'attività elettromagnetica dell'encefalo. Ha il vantaggio rispetto all'EEG di captare le variazioni di campi magnetici evocati dai potenziali elettrici presenti negli spazi extracellulari dei neuroni piramidali. Infatti al contrario dei campi elettrici i campi magnetici sono scarsamente soggetti a distorsione concorrendo ad una migliore risoluzione spaziale dello strumento. La sua applicazione sperimentale si ha grazie al calcolo dei campi magnetici collegati ad eventi, una sorta di potenziali evocati magnetici, detti ERF (event-related field) e sono codificati anch'essi in base al tempo che intercorre tra lo stimolo e la comparsa del potenziale.



Figura 29: MEG

1.4.4.1. Neurolinguistica e Meg

Gli studi di tipo neurolinguistico finora condotti hanno come argomento principale la dislessia: la MEG infatti permette di visualizzare con una ottima risoluzione temporale e una buona risoluzione spaziale, tutti quei processi legati alle diverse fasi della produzione e della comprensione del linguaggio sia parlato che scritto. Ad esempio lo studio condotto sulla cronologia delle attivazioni cerebrali durante la denominazione delle figure, è stata condotta tramite tecnica magnetoencefalografica. I risultati sono così riassunti:

Fase elaborativa	Time (msec)
preparazione concettuale	175
accesso al lemma	75
accesso alle informazioni fonologiche	80
sillabificazione	125
codifica fonetica	145
TOT	600

Tabella 1: tempi di esecuzione dei vari compiti

1.4.4.2. Limiti

Ad oggi è poco impiegata e necessita di una stanza ad isolamento magnetico per effettuare i test, ed è inoltre soggetta ad artefatti di natura biologica. La tabella qui di seguito fornisce un quadro d'insieme delle caratteristiche delle apparecchiature.

	Risoluzione temporale	Risoluzione spaziale	Tipologia	Studi	Limiti
TMS	N sec dipendente dal tipo di stimolazione	1cm in superficie e fino a 2.5 in profondità	Elettromagnetica non è una tecnica di <i>neuroimaging</i>	Lesioni virtuali, reazioni delle strutture corticali	non fornisce immagini, non influisce sulla sottocorticale, richiede personale medico
EEG	1msec	10-20mm	<i>neuroimaging</i> elettromagnetica	Fenomeni evento correlati	la scarsa risoluzione spaziale
fMRI	0.1- 5sec	0.5- 2mm	<i>neuroimaging</i> emodinamica	Topografici- morfologici-funzionali	scarsa risoluzione temporale, non evidenzia le reti neurali
PET	30sec	<5mm	<i>neuroimaging</i> emodinamica	Topografici- morfologici-funzionali	paradigmi complessi, tempi lunghi di elicitazione, elevato costo della strumentazione, altamente invasiva
MEG	1msec	5-10mm	<i>neuroimaging</i> elettromagnetica	Fenomeni evento correlati	Gli artefatti biologici

Tabella 2: quadro sinottico delle tecniche e delle loro caratteristiche

1.5. ERP: potenziali evento correlati

Come abbiamo mostrato precedentemente l'uso dell'eeg, permette di rilevare l'attività elettrica spontanea e di registrare la risposta ad un determinato stimolo sensoriale.

Il potenziale di risposta è una oscillazione con caratteristiche proprie specifiche in **ampiezza**, **latenza** e **polarità** che ne permettono l'individuazione. Il potenziale si distingue in due categorie: nel caso di risposte fisiologiche evocate da un evento esterno per esempio uno stimolo sensoriale, ripetuto un numero indefinito di volte, si parlerà di **potenziali evocati stimolo correlati** anche detti **potenziali esogeni**. Nel nostro caso in cui il potenziale nervoso è registrato durante compiti cognitivi ed attentivi con un determinato significato, si parlerà di **potenziali evento correlati (ERPs)** detti anche **potenziali endogeni**.

I potenziali hanno caratteristiche d'onda specifiche, ma si presentano ridotti, che tuttavia sono sommerse nel tracciato eeg in cui il rapporto SNR (segnale rumore) è a svantaggio del primo. Quindi si procede con l'*averaging* cioè con la somma delle ripetizioni dell'evento che poi viene mediata. Nella rappresentazione per convenzione le deflessioni (o picchi)

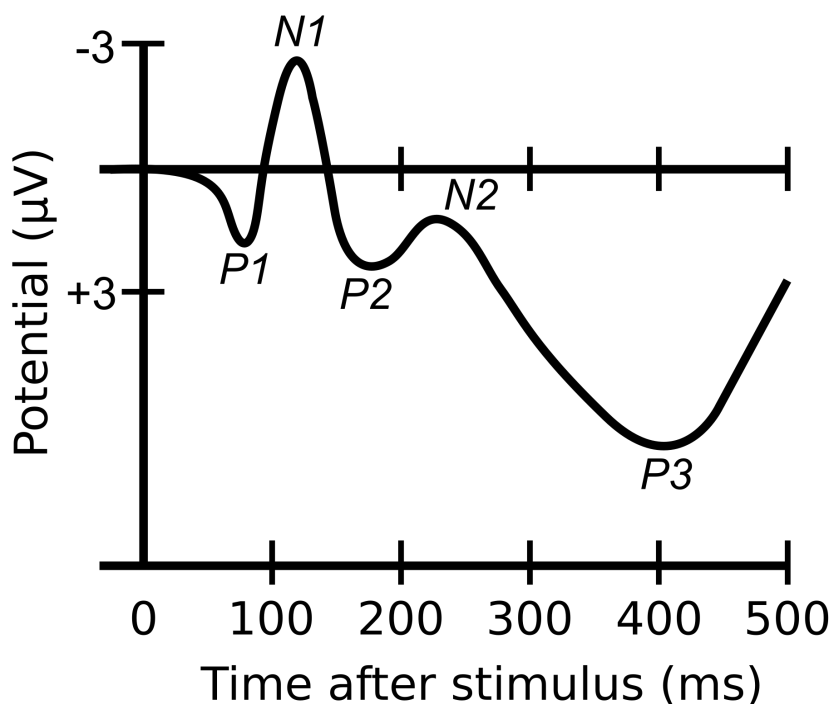


Figura 30: potenziali evento correlati (ERP)

positive volgono verso il basso quelle negative verso l'alto. Su una risposta si misurano i valori di ampiezza e latenza dei picchi. Essi vengono classificati e nominati in base alla loro polarità positiva o negativa (P o N), all'ampiezza del picco dell'onda, alla posizione sullo scalpo e alla latenza cioè il periodo che intercorre tra l'inizio dello stimolo e la com-

parsa della modificazione dell'EEG.

Quindi un potenziale positivo che compare dopo 100msec dallo stimolo sarà denominato P100. Vediamo nello specifico come vengono elicitati gli ERP, quali sono le caratteristiche generali. I potenziali evocati stimolo correlati sono delle risposte elettriche a stimoli sensoriali, registrabili con elettrodi sulla superficie del derma, spesso sullo scalpo.

Sono adottati soprattutto ambito clinico per la diagnosi di lesioni nello specifico sistema sensoriale esaminato.

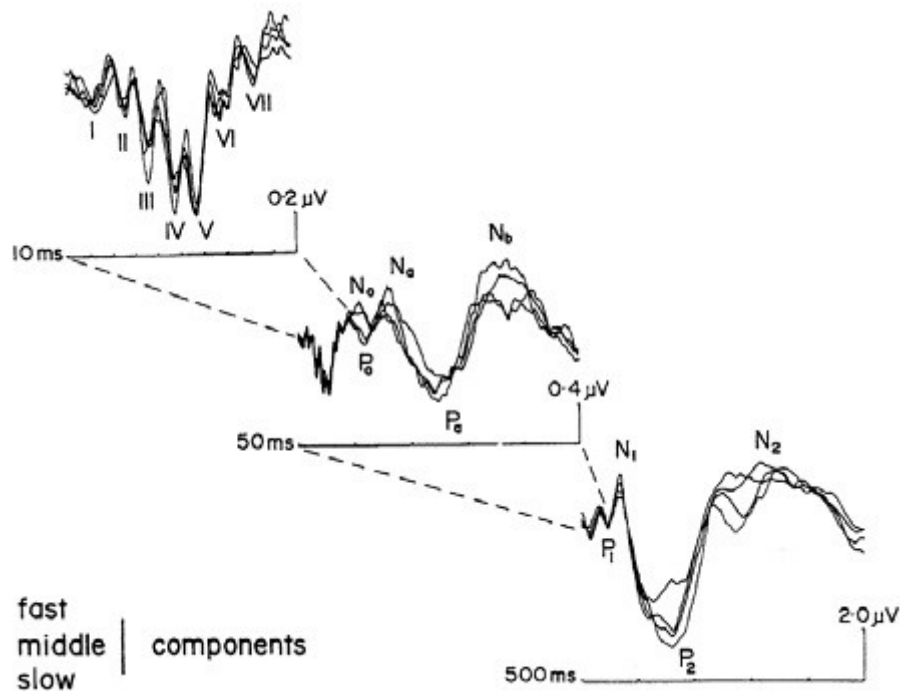


Figura 31: componenti veloci, medie e lente da Näätänen R & Winkler I (1999) *Psychophysiol*, 125, 826-859 Pauline e Hallowell Davis 1935-1936

I potenziali evocati sensoriali detti PES si originano in seguito alla stimolazione di un organo sensoriale quale l'occhio, l'orecchio la cute con uno stimolo visivo, acustico ed elettrico, rispettivamente. Grazie al potenziale evocato può essere studiato il percorso dello stimolo dalla periferia al sistema nervoso centrale. Parametro fondamentale nello studio dei potenziali è la latenza infatti è possibile scegliendo un'opportuna finestra di analisi (tPS) focalizzare lo stadio di analisi del segnale. Ad esempio per lo stesso stimolo avremo 3 componenti: la prima, detta precoce, con latenza ≤ 20 msec è indice di risposta delle vie nervose e del tronco encefalico. La seconda, detta intermedia, con latenza 20-200 msec essa è indice di risposta delle aree corticali sensoriali. La terza, con latenza lenta fino oltre i 300 msec, è indice di risposta delle vie nervose. Da ciò risulta evidente come la scelta di un

protocollo piuttosto che un altro e ancor prima la sua progettazione debba tener conto ciò che esattamente vuole evidenziare. Componenti del protocollo e quindi variabili sensibili al tipo di componente da elicitare sono: 1) il tipo di stimolo, 2) la posizione degli elettrodi, 3) il numero di ripetizioni, 4) la cadenza di ripetizione, 5) l'amplificazione, 6) la banda amplificata.

1.5.1. Il problema inverso

Per ciascun potenziale evocato è possibile individuare l'area superficiale di provenienza (topografia) della risposta dipendente dall'attivazione dell'area dedicata allo svolgimento di un compito sia esso di trasmissione che di elaborazione. La localizzazione da luogo a molteplici problematiche infatti per l'identificazione a livelli profondi del tessuto neuronale. Si tratta del **problema inverso**. Per chiarire il concetto possiamo partire dal più noto dei problemi inversi, cioè quello contenuto nel VII libro della Repubblica di Platone, si ricordi lo sforzo dei prigionieri della caverna nel tentativo di voler ricostruire il mondo reale, cioè che è fuori dalla caverna, sulla base delle informazioni limitate di cui dispongono, mere ombre proiettate sul fondo della caverna. Ma ciò porta solo a ricostruzioni distorte.

Il problema inverso si oppone per sua natura ad un tipo di problema diretto e sarà rispetto a quest'ultimo molto più complesso perché in generale non produce mai un'unica soluzione: significa che mancano informazioni sul modello, tali da permettere di scegliere quella giusta tra le soluzioni possibili. Ulteriori informazioni possono essere aggiunte ad un modello attraverso l'ICA.

1.5.2. Breve classificazione dei potenziali eventocorrelati o ERP

Qui di seguito riportiamo una breve classificazione degli ERP, in queste righe è impossibile essere esaustivi proprio per la natura dell'oggetto di trattazione data la mole di lavori sull'argomento e che spesso risultano essere tra loro discordanti. Abbiamo quindi tentato di offrire una trattazione agevole che raccogliesse interpretazioni univoche. Le prime componenti che sono anche le prime a comparire sulla temporale dell'eeg sembrano rispecchiare caratteristiche fisiche degli stimoli e la natura sensoriale primaria della percezione di questi. Ci soffermeremo ponendo particolare attenzione alla P300 perché ad oggi è la componente più utilizzata per il BCI.

- **P1:** si tratta di una componente positiva che ha una latenza rispetto all'apparire dello stimolo che può oscillare tra i 50ms per gli stimoli uditivi e i 100ms per gli stimoli

visivi. E' interpretata come indicatore neurofisiologico dell'attenzione. Dal punto di vista della distribuzione sullo scalpo, la P100 uditiva ha una maggiore ampiezza sulla linea centrale dello scalpo per intenderci la linea che congiunge un orecchio all'altro. La P100 visiva è individuabile nell'area occipitale e l'ampiezza di questa componente è determinata dall'attenzione.

- **N100:** tipicamente questa componente con polarità negativa appare 100ms dopo l'apparire dello stimolo visivo è possibile rilevare la sua ampiezza massima nelle aree frontocentrali e nel caso di stimoli uditivi aumenta proporzionalmente al grado di attenzione. Per gli stimoli visivi abbiamo una localizzazione con maggiore ampiezza nell'area occipitale.
- **P200:** ha una scarsa variabilità inter-individuale e per questo è tra le più studiate. Nei protocolli uditivi è spesso preceduta dalla N1, La sua ampiezza è maggiore nell'area frontale-centrale ed è legata alla complessità dello stimolo. Nel caso di paradigmi visivi abbiamo un'attivazione maggiore nella zona occipitale. La latenza varia tra circa 150 e 275 ms dopo l'inizio di un qualche stimolo esterno. Funzionalmente, sembrerebbe rappresentare un aspetto di ordine superiore di elaborazione percettiva, dipendente dall'attenzione.
- **N2:** questa componente ha una elevata variazione inter-individuale, sembra essere coinvolta nella discriminazione dello stimolo. Risulta avere un'ampiezza e una latenza proporzionali agli intervalli di inter-stimolo. Per i paradigmi uditivi è rilevabile T3, T4; invece per i paradigmi visivi è rilevabile intorno e sopra Oz. Sempre nell'ambito visivo è di notevole interesse il picco intorno compreso tra 256 189ms detto N170 che sembra essere caratteristico nel riconoscimento di facce famigliari.
- **P300:** si tratta di un potenziale positivo (maggiore di 10 μ V) endogeno, il picco del potenziale compare dopo circa 300ms dallo stimolo. Compare esclusivamente in coincidenza del riconoscimento dello stimolo prefissato (stimolo-*target*) ed ha distribuzione caratteristica sulle zone centro-parieto-occipitali. Questo particolare fenomeno elettrofisiologico si colloca al termine di un'elaborazione cognitiva che prevede stadi complessi, esso è definito come fenomeno attentivo. La latenza della P300 esprime il tempo impiegato dal soggetto per completare il compito di riconoscimento dello stimolo-*target*. L'ampiezza risulta essere inversamente proporzionale alla probabilità di comparsa del target e dalla quantità di informazione contenuta, cioè dalla complessità di esso ma anche in base alla distanza dal *target*. Facciamo notare che la P300 è più sensibile di altre componenti al livello di attenzione del soggetto.

to ciò influisce sulla concentrazione e sui meccanismi di elaborazione dell'informazione quindi sulla prestazione. Si è visto in sede sperimentale che la gratificazione del soggetto (dal pagamento dell'attività volontaria, all'offerta di una colazione ma anche il semplice "ben fatto") per lo svolgimento del compito richiesto migliora la prestazione. Il paradigma di elicitazione più utilizzato è quello oddball: composto da un treno di stimoli standard (non target) all'interno del quale vengono inseriti degli stimoli rari (o target) la comparsa dello stimolo target varia in maniera quantitativa su base percentuale e con casualità randomica.

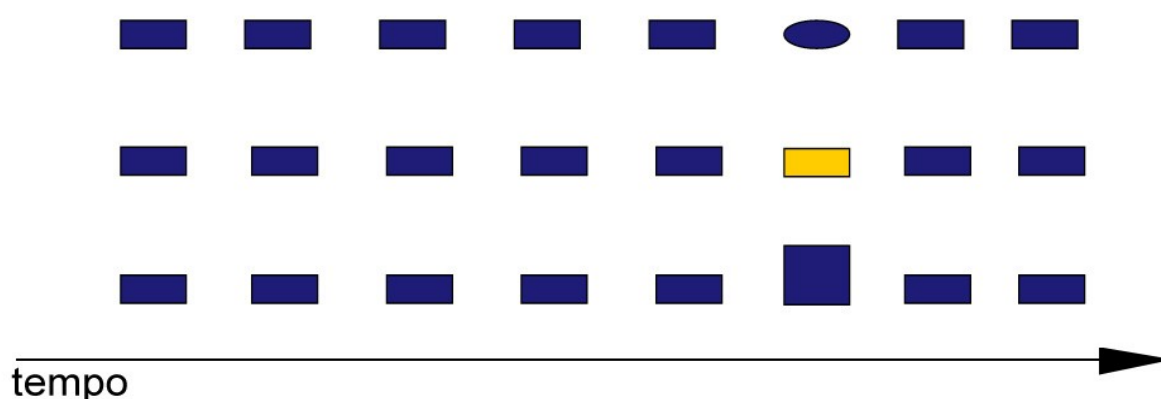


Figura 32: esempi di paradigma oddball

La task del *single stimulus* prevede un target raro (T) in assenza di altri stimoli. La task *oddball* invece prevede due differenti stimoli che vengono presentati con una sequenza casuale, dei quali uno ha una frequenza minore dell'altro (target = T, standard = S). Al soggetto viene chiesto di rispondere mentalmente o fisicamente agli stimoli target e di non rispondere a qualsiasi altro stimolo. In ogni caso, il target genera un ampio potenziale positivo che cresce in ampiezza muovendosi dagli elettrodi frontali verso i parietali ed ha una latenza di circa 300ms per gli stimoli uditivi e 400 ms per quelli visivi in soggetti giovani. Più nel complesso possiamo ipotizzare che per la P300 la memoria svolga un ruolo di base nel processo di comparazione: attivandosi nel momento in cui ci vengono sottoposti stimoli dall'esterno e concorrendo alla discriminazione degli stessi. Abbiamo la comparsa della componente endogena P300 quando avviene il riconoscimento di uno stimolo presentato diverso da quello precedente. Se lo stimolo non si differenzia dal precedente abbiamo la generazione di un'onda N100 seguita da una P200. A questo punto il processo sembrerebbe procedere prima attraverso l'attenzione poi coinvolgendo la working memo-

ry: la prima distingue lo stimolo raro dal treno di stimoli standard e la seconda supporta questo processo immagazzinando le caratteristiche degli stimoli rari e permettendo quindi un confronto tra target e non target. Ad oggi tuttavia i meccanismi di generazione della P300 non sono del tutto chiari. Per i contributi che gli studi che utilizzano queste componenti, hanno saputo offrire alla linguistica rimandiamo al capitolo sullo stato dell'arte.

- **N400:** essa si presenta dopo i 400ms dallo stimolo e sembra essere associata a compiti di comprensione sia di tipo uditivi che visivi. E' stata utilizzata soprattutto per indagini sulla semantica e sintassi di frase.
- **P600:** questa componente è associata a due funzioni principali, il primo legato ai processi di memoria il secondo al linguaggio in particolare alla prosodia. Appare tra i 400 ei 600ms è una componente ancora poco studiata.
- **La mismatch negativity MMN:** descritta per la prima volta nel 1978 da Naatanen (Näätänen & Al., 1978). , essa appare come una deflessione con latenza tipica compresa tra 150 e 250 ms. Appare maggiore in ampiezza nell'area fronto-centrale. Per attivare la MMN il paradigma utilizzata è di tipo oddball: un treno di stimoli standard e uno stimolo deviante, la risposta al deviante sarà la nostra componente mmn. E' stata molto utilizzata per stimoli uditivi ma poco per quelli visivi. La mismatch negativity avviene anche se il soggetto non presta consciamente attenzione allo stimolo. Per la MMN uditiva possiamo avere risposte a variazioni di tono, intensità o durata. L'onda della MMN visiva può verificarsi in risposta a vari aspetti devianti come il colore, le dimensioni, oppure la durata. Questa componente esogena viene interpretata come una " memoria di traccia", cioè la stimolazione ripetuta e regolare formerebbe un modello neurale nella memoria a breve termine e la violazione di esso porterebbe alla generazione della MMN. Al momento la generazione del modello mnemonico non è dimostrata (Näätänen & Winkler, 1999; Näätänen, Paavilainen, Rinne, & Alho 2007).

2. LE NUOVE FRONTIERE DELLA COMUNICAZIONE

2.1. Brain-Computer Interface

Con sistema BCI intendiamo una strumentazione in grado di permettere una connessione diretta fra cervello di un soggetto ed computer. Ciò permette di dotare il soggetto comunicante di una nuova modalità di comunicazione con un nuovo canale. Questa tecnologia ha risvolti importanti sul piano della qualità della vita per tutti quei soggetti con patologie neuromuscolari gravi che sono impediti nell'uso delle normali vie neurosensorie-comunicative: infatti con un sistema BCI è possibile trasformare “i pensieri in potenza” in “pensiери in atto”. Qui di seguito riportiamo la figura di un sistema BCI generico con le sue applicazioni più note.

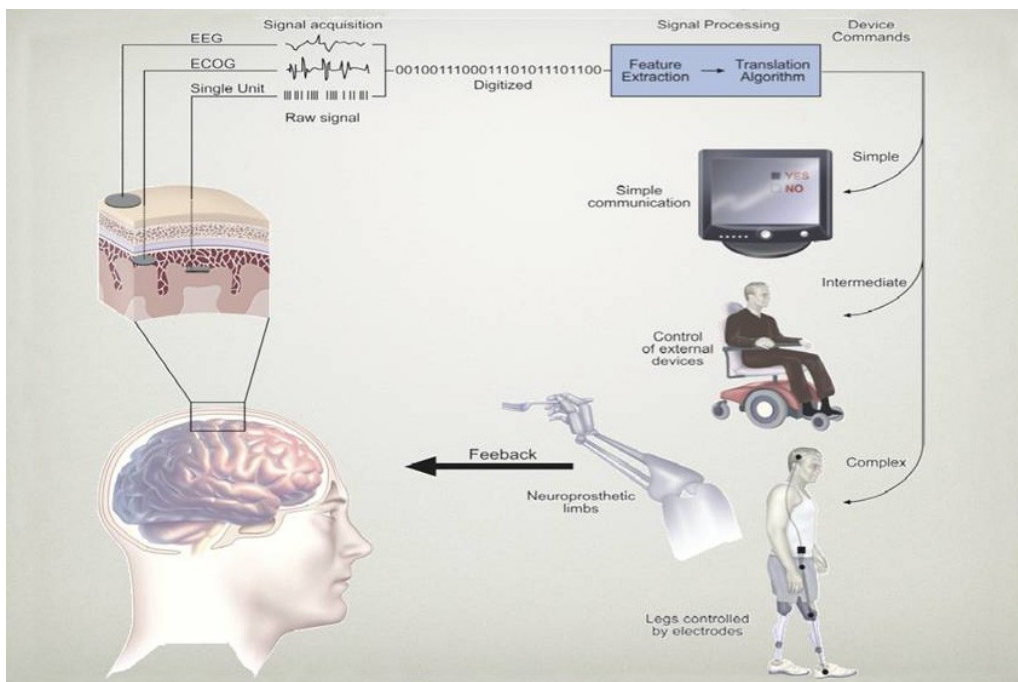


Figura 33: esempio sistema BCI; da sx abbiamo la rilevazione del segnale cerebrale, di seguito la digitalizzazione e la processazione del segnale che quindi giunge alle periferiche di comunicazione/azione

Vediamo come funziona un sistema BCI. Il soggetto esprime con il pensiero la volontà di compiere una determinata azione, ciò porta all'eccitazione dei neuroni corrispondenti all'area motoria specifica. Il segnale è captato dall'EEG che trasforma il segnale analogico in segnale digitale, il quale viene poi elaborato dallo stesso sistema BCI (blocco Signal Processing in figura 28), fornendo un una risposta in uscita. Avremo così un feedback per il soggetto: in tutti i sistemi BCI questa fase è sempre presente, proprio come accade nei sistemi naturali; per alcuni è l'illuminarsi di una parola sul monitor, per altri è il movimento di

un arto robotico. Il feedback fornito al soggetto, ha un ruolo fondamentale sia nel sistema BCI sia nel caso di comunicazione con normali canali neuromuscolari; infatti questo processo permette il controllo la verifica e quindi l'eventuale calibrazione dell'input.

Il sistema BCI dipende da un lato dal cervello del soggetto, che produce i segnali poi misurati dal primo segmento del sistema; dall'altro il secondo segmento che trasforma questi segnali in comandi specifici. Il sistema BCI dipende da un lato dal cervello del soggetto, che produce i segnali poi misurati dal primo segmento del sistema; dall'altro il secondo segmento che tramuta questi segnali comandi specifici. Per il primo segmento, è fondamentale la modalità di indurre la generazione dei segnali di comando da parte del cervello, cioè i task. Per il primo segmento, è fondamentale la modalità di indurre la generazione dei segnali di comando da parte del cervello. Esistono due tipologie principali di paradigmi per l'elicitazione delle componenti. Ad esempio è possibile indicare uno stimolo come target e altri come non-target e utilizzando un paradigma di somministrazione oddball: il soggetto dovrà focalizzare la sua attenzione solo sul target. E' possibile utilizzare anche la *motor-imagery*. Si tratta di un protocollo per cui il soggetto pensa di voler eseguire un determinato movimento, come nel nostro caso produrre una vocale, senza che questo gesto sia generato fisicamente come output. Infatti è dimostrato che la rappresentazione mentale di un movimento coinvolge le stesse aree cerebrali di quando il movimento viene realmente effettuato. Il *feedback*, può essere di livello *simple*, per cui il soggetto tramite il sistema BCI si limiterà a muovere un cursore su uno schermo; potremo avere un sistema di livello *intermediate* in cui l'output diviene comando per gli spostamenti della carrozzina in un ambiente reale; il livello *complex* è l'ultimo stadio che permette al soggetto di pilotare esso stesso le sue protesi come fossero i suoi arti naturali. Il sistema BCI più comune, è quello di livello *simple* per la comunicazione anche se ad oggi sono in molti a tentare di utilizzare il livello successivo e vedremo di seguito con quali risultati. Queste tecnologie sono dedicate in maniera molto specifica, a quei soggetti che affetti da patologie muscolari degenerative come la SLA (sclerosi laterale amiotrofica) ma anche quelli colpiti da ictus cerebrali o da lesioni spinali. Si tratta di soggetti che possono avere deficit parziali ma si può arrivare fino ai *locked-in*. Negli ultimi 10 anni in Italia come nel resto del mondo è cresciuta la sensibilità e l'attenzione verso la SLA, una patologia di tipo progressivo che porta il soggetto ad una completa paralisi motoria bloccando sia il movimento oculare sia quello respiratorio autonomo. Il soggetto mantiene però intatte tutte le facoltà emotive e cognitive quindi è come imprigionato in un guscio

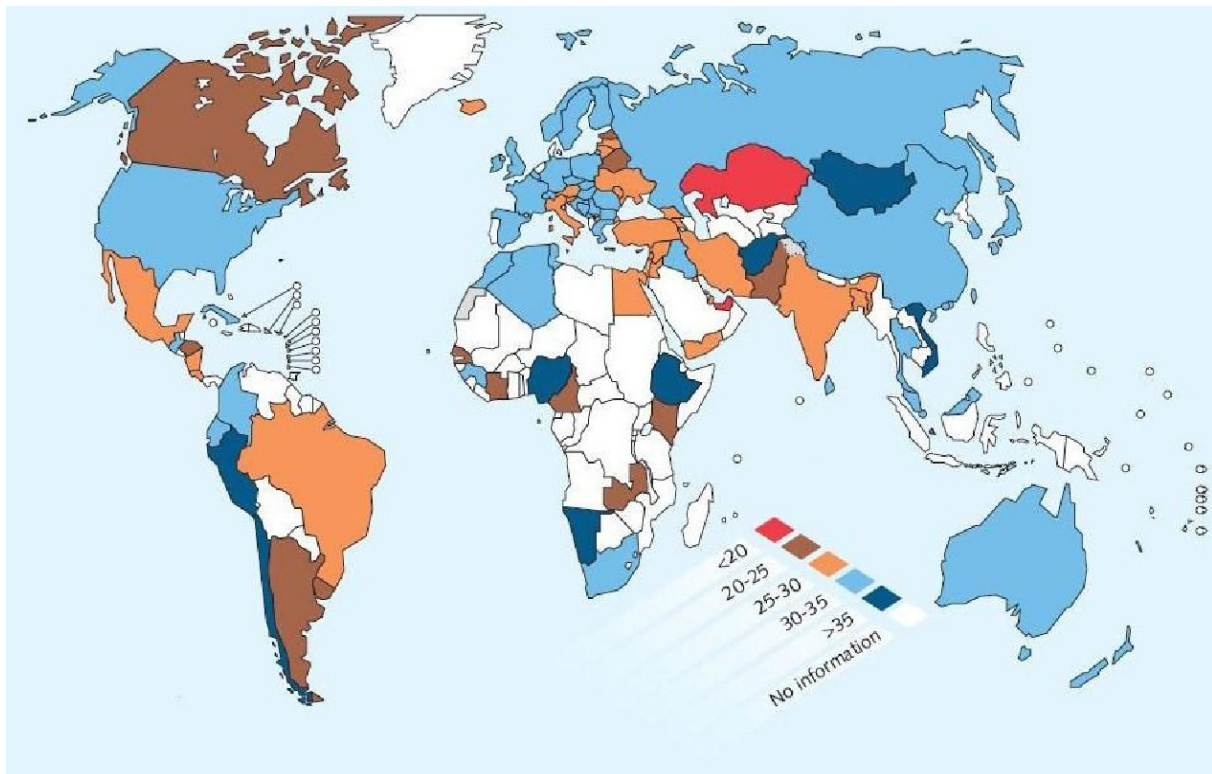


Figura 36: Atlante ALS nel mondo [<http://www.atlasofms.org/>].

La World Health Organization ha redatto nel 2008 un atlante raffigurante la diffusione della SLA [<http://www.atlasofms.org/>]. Qui si rende conto in particolare del fattore età, i sintomi si manifestano in età giovanile tra i 25 e i 35 anni, questo dato accompagnato da quello geografico ci mostra come i giovani, parliamo della generazione produttiva del nostro paese, siano i soggetti a rischio. Da qui nasce un'urgenza che non è solo quella di prevenire e curare in futuro, ma quella più immediata di rendere migliore la vita di questi cittadini.

2.1.1. Tipologie di BCI

Un sistema BCI è attivato tramite un segnale generato dall'attività cerebrale del soggetto (input). I sistemi si distinguono a partire dal luogo di elicitazione e dalla tipologia di segnale estratto [N. Birbaumer, 2005.], vediamo qui di seguito i segnali di input più diffusi.

- Potenziali sinaptici ed extra-cellulari: sono segnali con un'elevata risoluzione sia temporale che spaziale. Sono riferiti ai singoli potenziali d'azione neuronali (spikes, o LFPs), riflettono l'attività combinata fra i neuroni adiacenti e le relative sinapsi (JR. Wolpaw, 2006., N. Birbaumer, 2005.).
- SCP (*Slow Cortical Potential*, fig. 37): si tratta appunto di potenziali lenti (1, JR.

Wolpaw, 2002., E. Niedermeyer, 2004.), queste variazioni di potenziale sono registrate a livello della corteccia cerebrale. Presentano una frequenza compresa tra 1 e 2 Hz. Le variazioni negative del potenziale corticale sono associate alle attività motorie e alle funzioni che coinvolgono l'attività corticale; le variazioni positive sono associate alla diminuzione di tale attività. Si presentano all'interno della finestra temporale compresa tra 0.5 e 10 secondi.

- Potenziale ERP P300: di questa tipologia abbiamo trattato in larga misura nel paragrafo dedicato ai potenziali.
- I ritmi senso-motori mu e beta (fig. 38): sono associati alle aree corticali e strettamente connessi alle aree motorie e premotorie, ciò li rende particolarmente adatti non solo alla comunicazione ma in particolar modo per il comando delle protesi. Questi ritmi possono presentarsi in un *range* compreso fra gli 8 Hz e i 32 Hz; quelli utilizzati per il BCI sono il ritmo mu (8-12 Hz) e il ritmo beta (18-26 Hz). Questi ritmi sono soggetti ad un fenomeno detto desincronizzazione (ERD) in prossimità del movimento o della preparazione ad esso; infatti avremo una diminuzione della positività dei due ritmi. Nella fase successiva di rilassamento si registrerà un aumento dell'ampiezza dei suddetti ritmi, cioè la risincronizzazione (ERS).

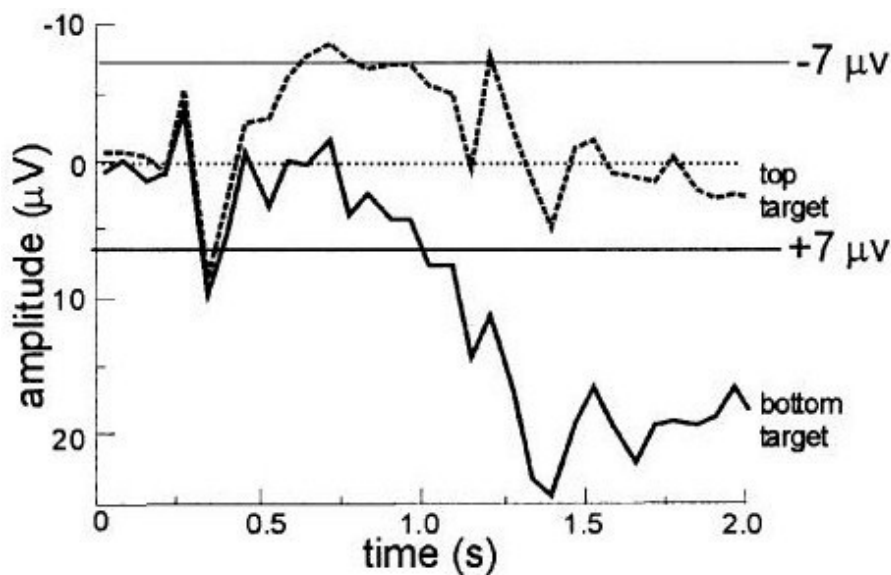


Figura 37: Slow Cortical Potentials, esempio di addestramento; il segnale tratteggiato è la risposta all'evento target. La linea continua è il segnale generato in corrispondenza di un diverso target posto in basso.

In base al luogo di estrazione del segnale possiamo avere BCI invasiva e BCI non invasiva. Per BCI invasiva intendiamo un segnale prelevato da elettrodi impiantati direttamente nel cervello.

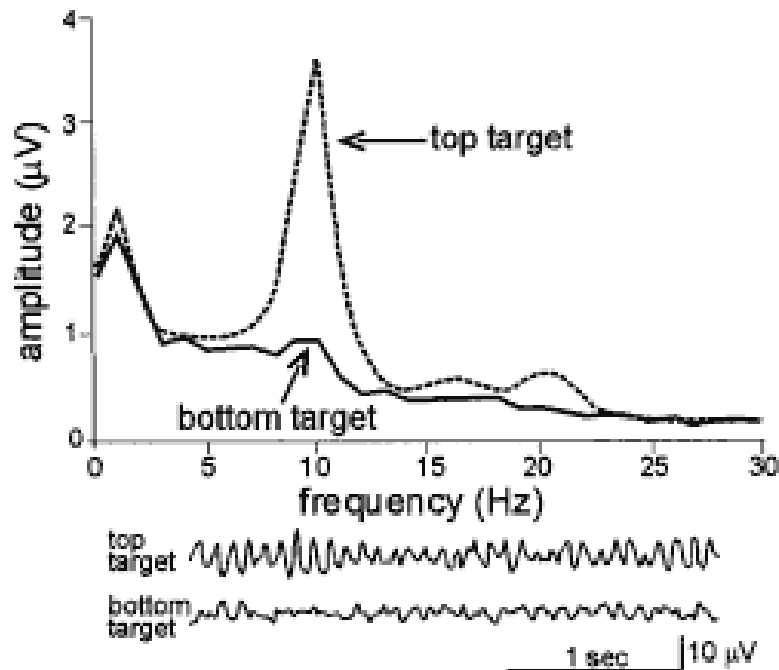


Figura 38: ritmo mu; l'immagine indica la diversa ampiezza del ritmo mu in funzione dello stimolo presentatosi al soggetto.

L'elettrocorticogramma (ECoG) è una tecnica sperimentata da due neurochirurghi canadesi Wilder Peneld e Herbert Jasper all'inizio degli anni '50 per degli studi areali sull'epilessia. Il segnale rilevato dall'ECoG è generato dai potenziali derivano dai neuroni corticali piramidali potenziali che si propagano perpendicolarmente alla corteccia cerebrale, dove sono posizionati gli elettrodi subdurali (sotto la dura madre). L'impianto degli elettrodi prevede un'operazione chirurgica di craniotomia.

E' una tecnica, quindi, notevolmente invasiva e costosa ciò comporta una scarsa possibilità nella riproduzione di esperimenti. Quindi è impensabile utilizzare il segnale elettrocorticografico per un dispositivo di tipo commerciale e con uso domestico su vasta scala. E' quindi evidente come il segnale elettroencefalografico diventi l'alternativa più utilizzata proprio per la sua non invasività e versatilità nelle applicazioni. Infatti l'EEG ricade nelle tecniche BCI non invasive. Il segnale viene prelevato da elettrodi posti sullo scalpo.

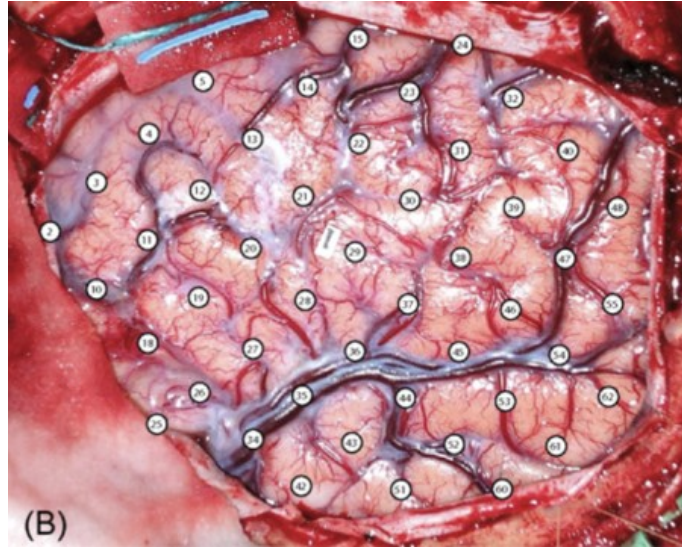


Figura 39: esempio di acquisizione del segnale ECoG

In base al tipo di segnali è possibile distinguere in BCI indipendente e BCI dipendente. BCI indipendente: quando il cervello genera autonomamente il segnale che media il messaggio da interpretare (P300). Per il BCI dipendente necessita di una generazione di segnali determinata dalla presentazione di stimoli esterni, ad esempio gli SSVEP, cioè i potenziali visivi (Steady State Visual Evoked Potential). Si tratta di potenziali che riflettono il processo di elaborazione dell'input visivo al livello cerebrale (fig.40).

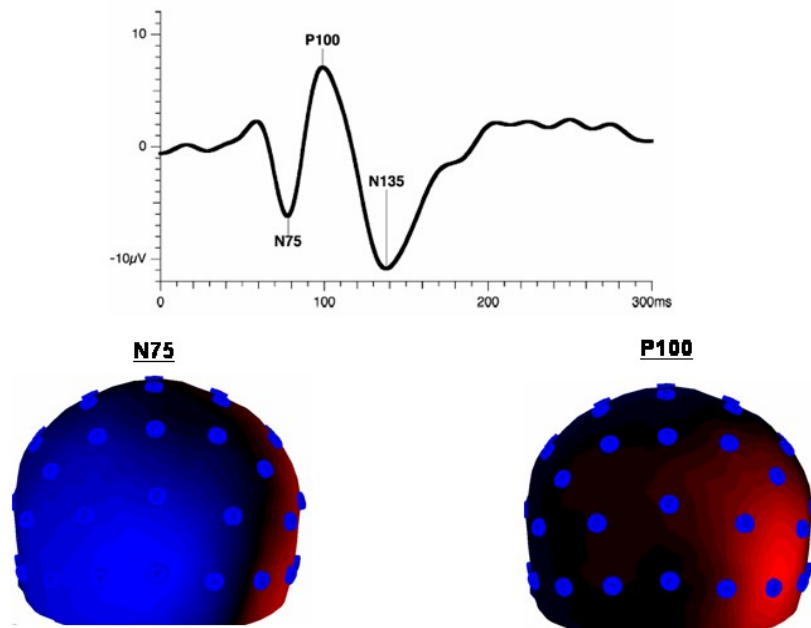


Figura 40: Esempio di un Visual Evoked Potentials

Il funzionamento di un sistema BCI basato su EEG è rappresentabile attraverso uno schema di blocchi. Per la trattazione più completa di tali blocchi rimandiamo al capitolo specifico sul trattamento del segnale (cfr. capitolo 4).

Il segnale generato dall'utente è captato dagli elettrodi quindi viene amplificato e convertito in digitale. I sistemi per l'acquisizione dati portatili grazie alle ridotte dimensioni e bassi consumi energetici sono la tecnologia utilizzata per il BCI domestico. I dati in formato digitale vengono elaborati e memorizzati se necessario sul computer. La gestione delle operazioni dall'esecuzione del task al *feedback* avviene attraverso l'implementazione di pacchetti anche *open source*. Da citare uno fra tutti in modo particolare per la sua capacità di gestione delle varie fasi e quindi anche il più usato anche nell'ambito della ricerca: BCI2000. Si passa cioè dall'acquisizione al pre-processing quindi all'estrazione delle features in fine alla classificazione traduzione in comando.

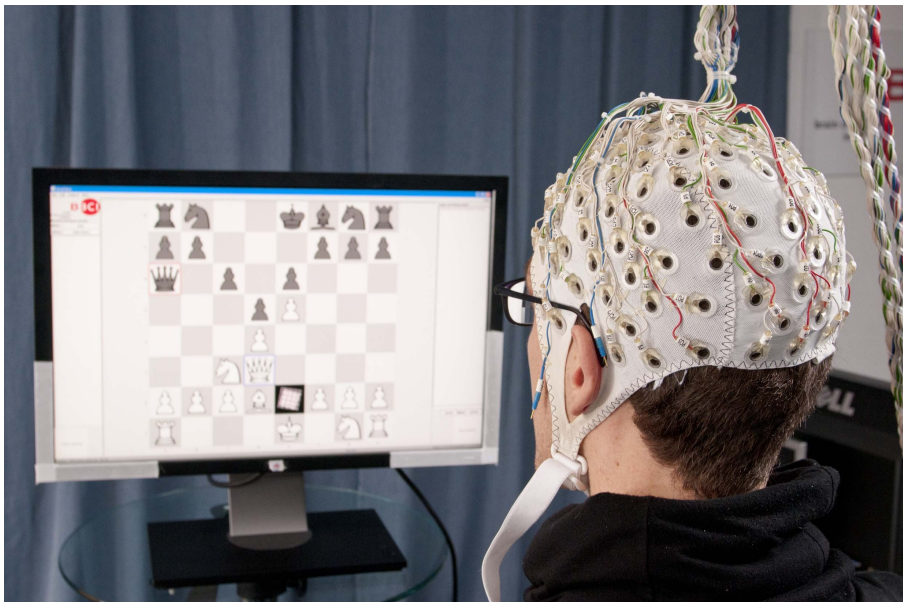


Figura 41: esempio di interazione in modalità gioco con sistema BCI

Ma come possiamo sapere quanto è efficiente la nostra macchina, quali sono gli indici di validazione del sistema BCI? I parametri sono (F. Cincotti et Al., 2006):

- Velocità di trasferimento dell'informazione: un dispositivo BCI dovrebbe essere in grado di trasportare informazioni ad alta velocità, ed avere un valore elevato di accuratezza.
- Facilità di utilizzo: un sistema che ha la possibilità di essere configurabile in modo molto ampio è molto utile in fase di ricerca ma, il dispositivo per l'utente finale do-

vrebbe rendere possibili solo un ridotto numero di operazioni. Ciò significa che tutta la macchina dai sensori, preferibilmente incorporati in un supporto leggero, al resto dell'apparecchiatura permetta un utilizzo semplificato che non porti disagio al soggetto e a chi lo accudisce.

- **Robustezza:** la possibilità che possa essere installato senza confliggere con i vari modelli di hardware.
- **Operabilità On-Demand:** prevede la possibilità che l'utente possa spegnere da solo la macchina.
- **Sicurezza:** si tratta di un punto fondamentale poiché i destinatari di queste tecnologie sono *in primis* soggetti con ridotta abilità motoria, la sicurezza dei sistemi risulta essere ancora un argomento fondamentale.

2.2. Silent speech interface

La possibilità di elaborazione del parlato in assenza di un segnale acustico intelligibile ha dato luogo all'idea di interfaccia per una comunicazione "silente", da utilizzare sia come ausilio per la comunicazione di soggetti con handicap, sia come parte di un sistema operativo per la comunicazione in contesti con forte disturbo del segnale.

Un'interfaccia *Silent Speech* (SSI) è un sistema che consente la comunicazione vocale quando il segnale acustico non è disponibile. Dei sensori acquisiscono dati dal processo di fonazione attraverso gli articolatori, i percorsi neurali; uno SSI produce una rappresentazione digitale del parlato che può venire sintetizzato direttamente, interpretato come insieme di dati, o indirizzato ad un network. I sistemi di comunicazione SSI sono ancora in fase sperimentale, ma sembrano evidenti le potenziali applicazioni.

I possibili utenti di questa tecnologia vanno dai malati che hanno difficoltà di fonazione e o per i quali la fonazione è preclusa dalla paralisi degli organi fonatori, a quei soggetti che per motivi di privacy, o che operano in settori particolari (dall'*intelligence* alle missioni nello spazio) hanno necessità di comunicare senza che il contenuto della conversazione venga ascoltato o che l'ambiente circostante possa creare interferenze.

Infatti queste tecnologie, essendo basate su segnali vocali non-acusticamente acquisiti, rendono i sistemi insensibili al rumore ambientale.

Ad oggi sono stati descritti in letteratura sette diversi tipi di tecnologia SSI basati su sistemi sperimentali:

- Cattura del movimento con sensori in punti fissi sugli articolatori tramite Articologra-

- fo elettromagnetico (EMA) (Fagan et al, 2008.);
- Caratterizzazione in tempo reale del tratto vocale utilizzando gli ultrasuoni (US) e l'imaging ottico della lingua e delle labbra (Denby e Stone, 2004;. Denby et al, 2006; Hueber et al, 2007a, b, c, 2008a, b, 2010);
 - Trasformazione digitale di segnali provenienti da un non-Audible Murmur (NAM) (Nakajima et al, 2003a, b, 2006;. Nakajima, 2005; Heracleous et al, 2007;. Hirahara et al, 2010;. Tran et al, 2008a, b, 2010);
 - Analisi dell' attività della glottide (Titze et al, 2000;. Ng et al, 2000;. Tardelli, 2003; Preuss et al, 2006;. Quatieri et al, 2006), o vibrazione (Bos e Tack, 2005;. Patil e Hansen, 2010);
 - Elettromiografia di superficie (sEMG) dei muscoli articolatori o della laringe (Jorgensen et al, 2003;. Maier-Hein et al, 2005;. Jou et al, 2006;. Hasegawa-Johnson, 2008;Jorgensen e Dusan,Schultz, 2010);
 - Interpretazione dei segnali provenienti da elettroencefalografo (EEG) (Porbadnigk et al, 2009.);
 - Interpretazione dei segnali provenienti da impianti nella *speech motor cortex* (Brumberg et al., 2010).

2.2.1. Quadro storico

La prima forma seppur embrionale di SSI nasce dall'idea di poter comprendere cosa si sta dicendo solo guardando le labbra (Petajan, 1984) abbiamo infatti intorno a questi anni un fiorire di corsi che insegnano questa tecnica ma anche di apparecchi che promettono di farlo in maniera automatica (Nakamura, 1988). Per comprendere la diffusione di questa idea si può fare riferimento ai molti film che citano questa tecnica si va da "2001: A Space Odyssey" (S. Kubrick, 1968) a "Los abrazos rotos"(P. Almodóvar, 2009) Il primo sistema di SSI, anche se con prestazioni molto limitate, è nato in Giappone. Nel 1985, attraverso i segnali elettromiografici provenienti da 3 sensori montati sul viso del parlante, è stato possibile riconoscere 5 vocali giapponesi con un'accuratezza del 71%. Le vocali sono state riprodotte *real-time* (Sugie e Tsunoda, 1985). A meno di 10 anni di distanza è stato sviluppato un sistema basato su immagini, con estrazione delle caratteristiche dei movimenti di labbra e di lingua estratti dal video, il risultato è stato del 91% di riconoscimento (Hasegawa e Ohtani, 1992). In Giappone nel 2002, un comunicato stampa NTT DoCoMo ha annunciato un prototipo di cellulare silenzioso grazie all'EMG e alla cattura ottica del movimento delle labbra (Fitzpatrick, 2002).

"La spinta allo sviluppo di un tale telefono cellulare", ha detto la società, "è stato liberare i luoghi pubblici di rumore", aggiungendo che "la tecnologia dovrebbe inoltre aiutare le persone che hanno definitivamente perso la loro voce." (Nakajima et al, 2003a;. Denby e Stone, 2004).

Da qui in poi cominciano ad essere pubblicati studi che hanno come oggetto i veri e propri SSI. I progressi tecnologici legati alla tecnologia vocale e alla bioingegneria hanno contribuito fortemente affinché si procedesse a passi più spediti nello sviluppo di questa nuova possibilità di comunicazione. Vi è quindi oggi una ricchezza di studi che prendono in esame le risorse tecnologiche disponibili per l'applicazione in ambito di studi per lo sviluppo del *Silent speech*. Si va dagli ultrasuoni (Stone et al, 1983;. Stone e Shawker, 1986; Stone e Davis, 1995; Scobbie, 2003; Stone, 2005; Davidson, 2005; Epstein, 2005), all' X-ray cine-radiografica (Arnal et al, 2000; Munhall et al, 1995), dall'fMRI (Gracco et al, 2005; NessAiver et al, 2006), agli EMA (Perkell et al ., 1992; Hoole e Nguyen, 1999), all'EMG (Tatham, 1971; Sugie e Tsunoda, 1985) all'EPG (Gibbon, 2005).

Ad oggi la frontiera è quella del SSI integrato in un sistema BCI che utilizza la tecnologia EEG (Wolpaw et al, 2002;. IEEE 2008;. Sajda et al, 2008).

Qui di seguito riportiamo alcuni studi di interfacce per il parlato silenzioso. Partendo da quelle con approccio fisiologico, le quali si servono delle tecniche che misurano i parametri fisici della produzione del parlato, arrivando fino a quelle che prevedono l'utilizzo di segnali di comando ottenuti direttamente dal cervello.

2.2.2. Cattura del movimento articolatorio attraverso EMA

Questa tecnica prevede di monitorare la produzione del parlato silenzioso attraverso elettrodi posizionati su gli articolatori e all'interno della cavità orale.

Il risultato saranno una serie di coordinate cartesiane poste in uno spazio a 2 o 3 dimensioni (Carstens, 2008; Scho NLE et al, 1987;. Hummel et al, 2006) ciò però riproduce quanto già sappiamo sulla posizione e forma dei tratti dei foni. Nel 2008 (Fagan et al., 2008) è stato studiato un sistema che utilizza magneti permanenti fissati ad una serie di punti distinti dell'apparato vocale, i sensori posti sul soggetto sono in rapporto con altri sensori magnetici posizionati intorno alla testa dell'utilizzatore. In questo caso magneti sono stati incollati alla lingua dell'utente, alle labbra e ai denti, oltre a 6 sensori.

Lo scopo dello studio era stabilire se fosse possibile estrarre informazioni sufficienti da una serie di sensori di questo tipo tali da permettere il riconoscimento di un corpus vocale

ridotto. L' algoritmo adottato, è basato su un adattamento del Dynamic Time Warping (DTW), che implementa la programmazione dinamica (DP) (Holmes e Holmes, 2001; Furui, 2001). Il protocollo prevedeva che il soggetto ripettesse 9 parole e 13 fonemi (per il codice di trascrizione dei fonemi si è utilizzato l' ARPAbet dello [Speech Understanding Project](#) ((1971–1976) (Levinson, 2005)) per i dati di addestramento della macchina. I risultati ottenuti in laboratorio con questo piccolo set di vocaboli ha portato ad un riconoscimento superiore al 90%, il che fa sperare che possa essere utilizzato con un vocabolario più ampio anche in situazioni reali.

2.2.3. Caratterizzazione in tempo reale del tratto vocale utilizzando l'*imaging* a ultrasuoni (US) della lingua e delle labbra

Imaging ad ultrasuoni è un altro modo per ottenere informazioni dirette sulla configurazione del tratto vocale: si tratta di una procedura non invasiva e clinicamente sicura che rende possibile la visione in tempo reale dei movimenti di uno degli organi dei articolatori fondamentali per la produzione del linguaggio, la lingua.

Il manopolo (sonda che emette gli ultrasuoni) è posizionato sotto il mento esso può fornire una vista parziale della superficie della lingua mentre una video camera posta frontale al soggetto inquadra le labbra. I dati di questo progetto SSI (Ouisper, 2006), sono serviti a comandare un sintetizzatore che doveva estrarre le coordinate accoppiando segnale video e segnale degli ultrasuoni. Il corpus di frasi era piuttosto ampio (720 frasi) ma per stessa ammissione dei ricercatori “la capacità di estrarre i fonemi in modo discreto da dati fisiologici è ancora da realizzare” ma “ si vedono le possibili applicazioni che questa tecnologia può offrire a soggetti che hanno subito laringectomie”. Un sintetizzatore vocale da parlato silenzioso che cercasse di tradurre in segnale acustico le sole informazioni visive, è stato tentato più volte; ad esempio attraverso l'uso del *mapping* ad ultrasuoni (Denby e Stone, 2004) o con l'utilizzo dello spettro di frequenza (Denby et al., 2006). L'estrazione e parametrizzazione del profilo lingua è stata sostituita da una tecnica di codifica globale: EigenTongues (Hueber et al., 2007a). Questa tecnica usa le immagini ad ultrasuoni della lingua, una volta registrate vengono estratte e proiettate all'interno di uno spazio rappresentato dalle configurazioni possibili all'interno del tratto orale.

Questa tecnica prevede la sintesi mediante LPC e il range di errore nella predizione della descrizione acustica va dall'11% al 16%. Per rendere migliore il riconoscimento e più naturale la sintesi nella forma di difoni associati alle relative configurazioni visive (Hueber et al., 2007b,c, 2008a, 2010). Si è così ottenuto un ampio *thesaurus* di audio-video.

Questo sistema è in grado di prevedere correttamente circa il 60% delle sequenze con solo input video. Al momento della sintesi viene aggiunto un pattern prosodico che è stato ritenuto accettabile 90% dei casi. Il miglioramento di questa tecnica è strettamente legata al miglioramento dei sistemi di cattura dell'immagine. Tanto è più alto il numero di fotogrammi al secondo tanto è più sarà efficiente il sistema di riconoscimento e quindi più accurata la sintesi (Hueber et al., 2008b). Ciò è previsto affinché la macchina possa essere utilizzata con un vocabolario molto ampio, ma una restrizione a 250 vocaboli la rende già operativa nel contesto reale.

2.2.4. Registrazione del segnale audio attraverso sensori di vibrazione ed elettromagnetici.



Figura 42: registrazione trasformazione di non-audible murmur (nam)
www.youtube.com/watch?v=I5dVILReCHQ

Per *Non Audible Murmur* (NAM) intendiamo il suono generato a bassa ampiezza dal flusso d'aria attraverso laringe e la sua risonanza nel tratto vocale (Nakajima et al, 2003 a,b; Otani et al, 2008). Il NAM è suono irradiato dalla bocca che può a malapena essere percepito dagli ascoltatori nelle vicinanze, ma è un segnale facilmente individuabile usando un microfono ad alta sensibilità attaccato alla pelle sopra il tessuto molle al di sotto della mandibola lungo il collo verso i mastoidi. Il microfono NAM è progettato per la rivelazione selettiva delle vibrazioni dei tessuti durante il parlato ed è insensibile al rumore ambientale. Si presenta come un dispositivo di input utile per le telecomunicazioni (telefoni cellulari) ma

anche per i disabili. La soluzione commerciale è composta da un microfono miniaturizzata incapsulato all'interno di un involucro di silicone che serve ad attenuare anche il rumore da fregamento sulla pelle. Per quanto concerne le sue prestazioni presenta un picco di 500-800 Hz, e una larghezza di banda di circa 3 kHz., presenta una lieve distorsione spettrale e una maggiore risposta ai suoni consonantici. Ad oggi il suo utilizzo è diffuso soprattutto nell'ambito della comunicazione tra utenti non-patologici, per la telefonia è stato dotato di un auricolare per la ricezione.

2.2.4.1. Analisi di attività glottale con sensori di vibrazione ed elettromagnetici

Nei primi anni del 2000, gli Stati Uniti Dipartimento della Difesa DARPA ha avviato l'Advanced Program Encoding (ASE), che prevedeva finanziamenti per sviluppare sensori non-acustici per la codifica del parlato in condizioni acustiche difficili, come gli interni dei veicoli da combattimento e aerei, operazioni militari ecc. Si sono valutati sensori specifici utilizzati in combinazione con un normale microfono a distanza ravvicinata. Per questi studi è stato creato un apposito data-base il "*DARPA Pilot Speech Encoding avanzata Corpus Speech*" (Tardelli, 2003).

Programmi con obiettivi simili sono stati intrapresi in Canada, dal dipartimento di Difesa Ricerca e Sviluppo (Bos e Tack, 2005), e in Europa, con il progetto europeo SAFIR (Speech Interface Research Automatic Friendly, IST-2002-507427) (Dekens et al., 2008).

La forma d'onda necessaria può essere ottenuta o tramite rivelatori sensibili alle vibrazioni trasmesse attraverso un tessuto o dall'interazione di movimento glottale con un campo elettromagnetico (Titze et al, 2000; Dupont e Ris, 2004; Quatieri et al, 2006).

2.2.4.2. Sensori di vibrazione: laringofono

I laringofoni hanno la forma di un archetto ai cui lati sono appunti i due microfoni da posizionare sul collo, uno su ogni lato del pomo d'Adamo sono a contatto diretto con la pelle. Sono progettati per avere una bassa risposta direzionale verso l'esterno, tale da rimanere insensibile al rumore di fondo. I Bone sono, come dice la parola stessa, (osso) progettati per raccogliere il segnale vocale che si propaga attraverso le ossa del cranio, come i precedenti sono anche resistenti alla contaminazione da rumore ambientale. Talvolta si presentano incorporati nei caschi hanno sempre applicazioni militari (Bos e Tack, 2005).

2.2.4.3. Microfono fisiologico, pmic

E' una forma più sofisticata del laringofono. Si indossa come un colletto, è costituito da un

senore piezoelettrico immerso in un gel all'interno di una vescica, è progettato per avere una forte aderenza che permette un guadagno del segnale rispetto all'aria, e quindi elimina il rumore di fondo. (Bos e Tack, 2005; Quatieri et al, 2006;. Preusset al, 2006;. Patil e Hansen, 2010)

2.2.4.4. Microfono da orecchio

E' inserito nel condotto uditivo (Bos e Tack, 2005;. Dekens et. al, 2008) e proprio grazie alla sua posizione garantisce una cattura del segnale molto buona.

2.2.4.5. Sensori elettromagnetici

L'elettroglottografo (egg) (Rothenberg, 1992; Titze et al, 2000; Quatieri et al, 2006), è uno strumento di ricerca standard progettato per rilevare i cambiamenti di impedenza elettrica attraverso la gola durante il discorso espresso. È costituita da 2 elettrodi placcati oro tenuti in posizione su entrambi i lati della laringe per mezzo di un collare, con un potenziale assegnato. Quando le corde vocali sono chiuse, l'impedenza elettrica diminuisce, mentre quando sono aperte, aumenta. La vibrazione glottale in questo modo induce un segnale di circa 1 V RMS su un supporto di 2-3 MHz, che è abbastanza facilmente rilevabile. Un inconveniente della tecnica è la sua sensibilità al posizionamento degli elettrodi.

2.2.4.6. GEMS

Si basa su una tecnologia radar in miniatura (<1 mW) relativamente recente (Burnett et al, 1997;. Titze et al, 2000) con un'alta risoluzione per brevi impulsi elettromagnetici nella banda dei 2,4 Ghz ISM. L'antenna GEMS può essere attaccata alla gola o in altre posizioni. Rileva in modo accurato la frequenza della glottide dal movimento della stessa. Il corretto posizionamento è un fattore determinante per la qualità del segnale. **TERC** (Brown et al, 2004, 2005;. TERC, 2009) misura le variazioni della capacità elettrica intrinseca della glottide. Il dispositivo sfrutta la variazione di permittività (permeabilità) della sezione trasversale del collo. Il dispositivo è costituito da un collare di tessuto su cui sono posizionati elettrodi di rame sintonizzati a una particolare frequenza di diverse decine di megahertz.

2.2.4.7. Elettromiografia di superficie (emg) e riconoscimento vocale

L'elettromiografia di superficie (emg) è una tecnica che prevede la registrazione dell'attività elettrica muscolare catturata da elettrodi di superficie non-impiantati. I dati rilevati sono co-

stituiti dalla differenza di potenziale tra regioni attive e regioni non-attive.

E' quindi possibile adottare questa tecnologia per la rilevazione dell'attività muscolare durante la produzione del parlato e quindi giungere alla sintesi, inoltre si tratta di una tecnica che richiederebbe il minimo sforzo da parte del parlante con un ottimo rapporto segnale/rumore. Le prime rilevazioni risalgono agli anni '80 (Sugie eTsunoda,1985;Morse e O'Brien, 1986,Morse et al., 1989, 1991). Sebbene i primi risultati siano stati promettenti, con un tasso di precisione del 70% su un vocabolario di dieci parole, la performance è crollata al 35% con l'uso vocabolari di 17 parole. In seguito sono state ottenute prestazioni migliori su un vocabolario di solo cifre (Chan et al. 2001,2003), si tratta di una precisione media per riconoscimento di parola del 93%. Si è ottenuto un riconoscimento del 92% con l'utilizzo di 6 parole-comando(Jorgensen et al.,2003). Recenti studi stanno cercando di superare i limiti principali propri di questi sistemi EMG. Ne diamo conto in modo schematico perchè la produzione è vasta: si va dalla scelta del vocabolario alle unità acustiche(Walliczek et al, 2006;. Schultz e Wand, 2010), dall'interfaccia (Maier-Hein et al, 2005) ai modelli articolatori integrativi, dagli studi sulla posizione degli elettrodi(Jou et al, 2007;. Schultz e Wand, 2010) al preprocessing (Maier-Hein et al, 2005)(Jorgensen e Binsted, 2005; Jou et al, 2006). C'è chi affronta le problematiche legate all'uso degli utenti con disabilità cercando di verificare le possibili variabili che influenzano la qualità del segnale (Wand e Schultz, 2009), ma anche ponendo attenzione alle caratteristiche di fruizione di queste apparecchi (Manabe et al., 2003, Manabe e Zhang, 2004).

Se da un lato la maggior parte dei primi studi è stato effettuato utilizzando piccolo numero di parole pronunciate sempre e soltanto in laboratorio (Chan et al, 2001.; Jorgensen et al, 2003; Maier-Hein et al, 2005), in uno studio recente, effettuato su un vocabolario di 100 parole, è stato dimostrato che è possibile ottenere un riconoscimento intorno al 70% (Jouet al., 2006). Infine, il lavoro più vasto è stato condotto su un data-base di registrazioni EMG ottenuto sottoponendo a lettura di frasi silenziosa e non, 78 parlanti (Wand e Schultz ,2009, 2010) L'EMG-based è utilizzato per il riconoscimento vocale in ambienti con forti componenti rumorose, il caso più noto è sicuramente il suo impiego per le missioni spaziali. Lo studio effettuato per la NASA (Betts et al., 2006) raggiunge il 74% di riconoscimento di 15 parole, nello specifico si tratta di un sistema che opera in tempo reale e che è stato applicato a soggetti esposti ad un livello di 95 dB di rumore.

2.2.5. Interpretazione dei segnali elettroencefalografici (EEG)

Suppes et al.(1997) furono i primi a dimostrare che le parole isolate possono essere riconosciute sulla base di registrazioni EEG e MEG (magnetoencefalografia).

Utilizzare un sistema BCI richiede di norma che gli utenti siano in grado di modulare le proprie onde cerebrali affinché il segnale diventi un'istruzione eseguibile dal dispositivo (Nijholt et al., 2008). Questa capacità deve essere acquisita e questo addestramento richiede tempo e impegno da parte dell'utente (Neuper et al., 2003).

Vediamo ora qui di seguito in maniera più approfondita gli studi dai quali questa tesi trae ispirazione. Nei primi due casi si utilizza l'EEG e nell'ultimo elettrodi intracorticali.

Più specificatamente questi studi sono una sotto sezione del *silent speech* denominato *unspoken speech*.

2.2.5.1. Wester e Schultz (2006)

Si tratta di un lavoro di tesi per il quale sono state effettuate delle registrazioni con EEG sistema internazionale 10-20 a 16 canali con elettrodi posti nelle aree di Broca e Wernicke (fig.43). I soggetti che hanno partecipato sono 6, di nazionalità tedesca, e non sono parlanti nativi d'inglese, sono tutti laureati. Il protocollo prevedeva che fossero ripetute delle liste di parole divise in *corpora*:

C.Digit: one, two, three, four, five, six, seven, eight, nine, zero

C.Digit5: one, two, three, four, five

C.Lettura: good afternoon ladies and gentlemen, welcome to the interact center my name is marek wester, let me introduce our new prototype, thank you for your attention, any questions

C. Telefonico: yes, no, accept, deny, wait

C.Player: start, back, next, louder, turn down

GRE: brittle, cordial, diffidence, regicide, profundity, presage, nonplused, insipid, fluster, tepid)

C.Alpha: alpha, bravo, charlie, delta, echo. Usate dall' International Civil Aviation Organization (ICAO)

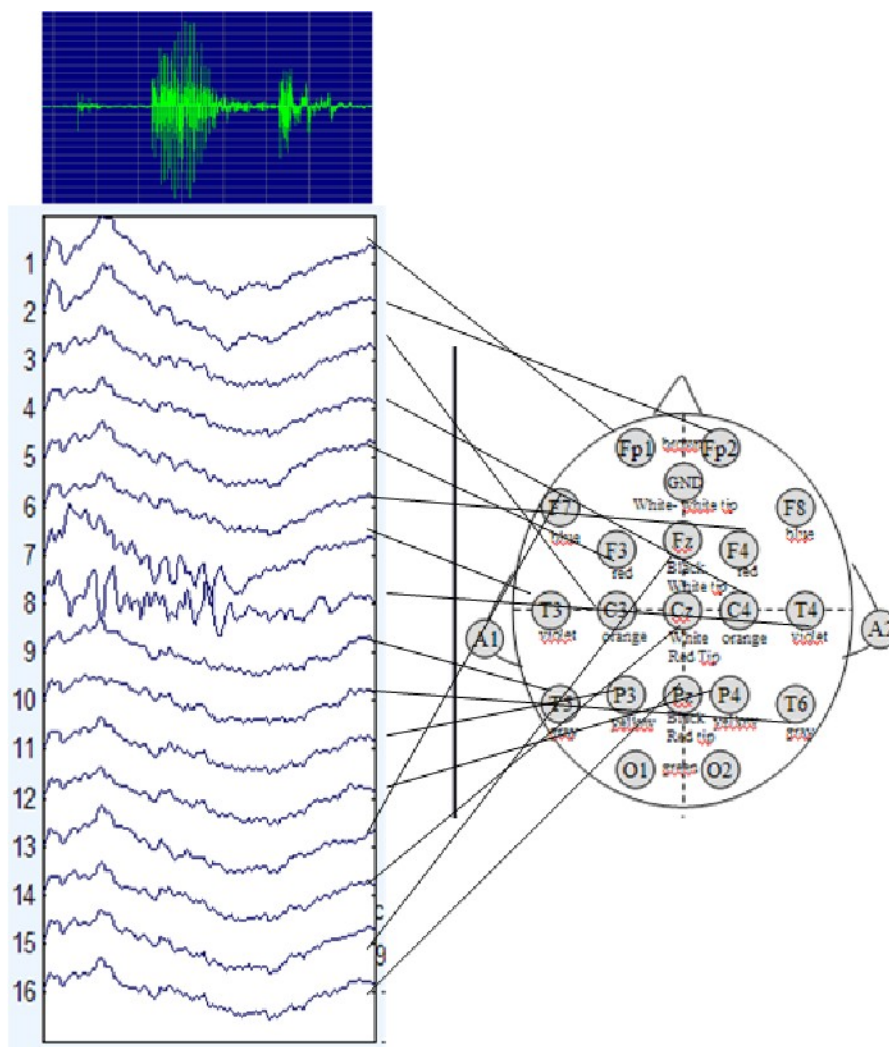


Figura 43: schermata che riproduce l'acquisizione

Le modalità di produzione del parlato sono state così divise: *normal speech*, *whispered speech*, *silent speech*, *mumbled speech* e *unspoken speech*.

La somministrazione degli stimoli è stata guidata da un 'istruttore e ha seguito lo schema in figura 44.

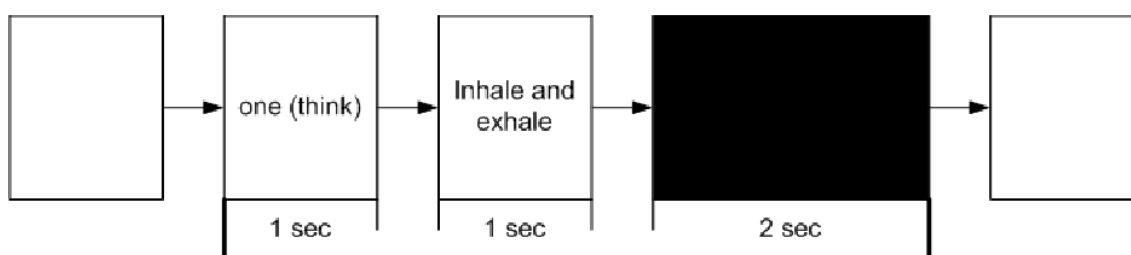


Figura 44:schema dei tempi di esecuzione del trial

Per il riconoscimento dell'*unspoken speech*, l'analisi e l'estrazione dei coefficienti sono state eseguite utilizzando *Short Time Fourier* (STF). I risultati relativi all'accuratezza delle parole per l'*unspoken speech* vanno da un 29% ad un massimo del 45%, si tratta di valori molto bassi per cui il riconoscimento è insoddisfacente. Ma sono utili i dati relativi agli elettrodi utilizzati per l'estrazione delle features infatti come si può vedere nelle figure gli elettrodi centro-temporali forniscono dati per il 40% quelli frontali intorno al 30% ma la combinazione apporta un contributo intorno al 45% per raggiungere il 47.24% aggiungendo anche quelli che vanno da T5 a T6 (fig 45a, fig 45b).

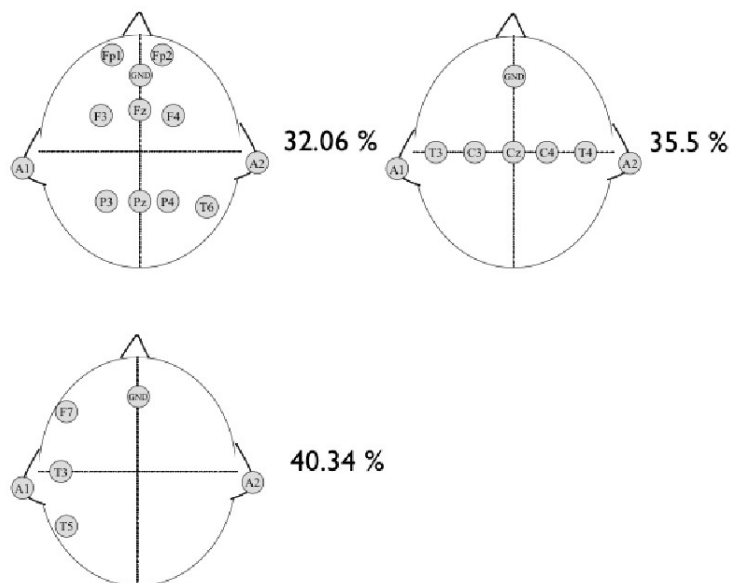


Figura 45a: risultati con miglioramento in percentuale modificando gli elettrodi

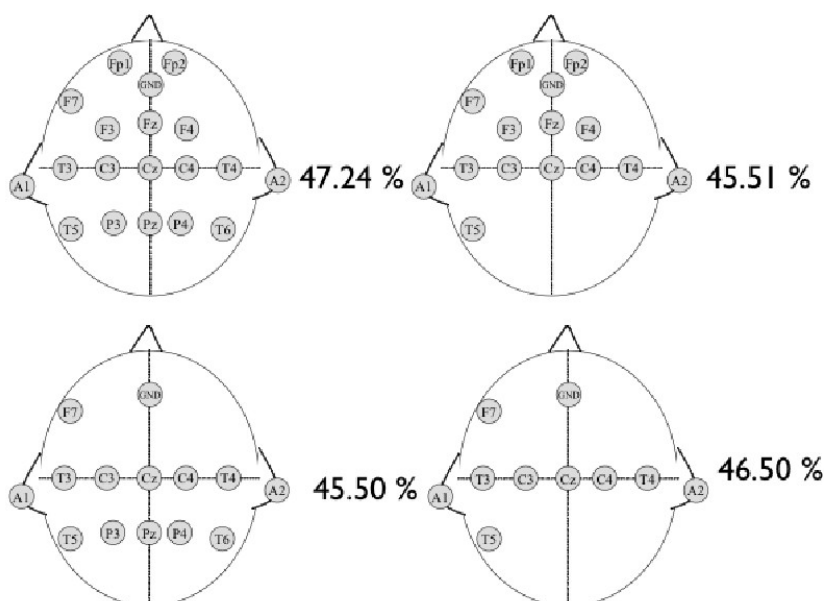


Figura 45b: risultati con miglioramento in percentuale modificando gli elettrodi

2.2.5.2. DaSalla et al (2009)

Riportiamo lo Studio del gruppo di Da Salla e del suo gruppo perché è stata per molti tratti un riferimento iniziale sia per le metodiche relative all'elicitazione dei dati sia per quelle di trattamento del segnale (DaSalla et al. 2009).

Nel suo studio ha proposto uno schema di controllo per un SSI BCI con attività neurale associata all'*imagery*: Il soggetto immagina di produrre delle vocali (/a/ e /u/).

I soggetti (2 maschi e 1 femmina) hanno eseguito 50 ripetizioni per ogni vocale, è stato inserito anche un task di controllo /+/ in cui il soggetto non deve pensare alla produzione linguistica. Il protocollo di somministrazione era il seguente. Per discriminare tra le attività, gli autori hanno riprogettato i più comuni (CSP) filtri spaziali. Questa tipologia può procedere ad una discriminazione solo tra due classi. I Risultati: i modelli spaziali hanno mostrato attivazioni per lo più simmetrica e centrata in corrispondenza della regione corteccia motoria, in particolare le posizioni di CZ e Fz del sistema internazionale 10-20.

Successivamente al filtraggio con CSP è stata addestrata un SVM non lineare con le 20 prove precedentemente selezionate per ciascun task, le rimanenti 30 prove per ogni task sono state sottoposte a classificazione.

L'attivazione che viene classificata in questo studio è intorno a 350ms, quindi un potenziale ritardato di P300 con un voltaggio molto alto, intorno ai 10 microvolt. I risultati relativi al test di riconoscimento sono in media intorno al $78 \pm 5,7$ e $68 \pm 68 \pm 12$ di valore percentuale ma per alcuni soggetti è di circa l'80%. I risultati sembrano ottimi per il riconoscimento di un segnale proveniente da segnale EEG durante *unspoken speech*, ma gli autori potrebbero non avere valutato alcuni fattori che potrebbero avere condotto ad una valutazione non realistica dei dati ottenuti. Vediamo una ad una le evidenze:

latenza a 350ms rientra nel *range* della P300, che ricordiamo essere un potenziale endogeno con caratteristiche ben specifiche di elaborazione interna di un segnale target, che può essere manipolato volontariamente e nei compiti di BCI il soggetto può arrivare persino a forzare il sistema di riconoscimento. Per essere un potenziale premotorio questa latenza ci sembra troppo lunga; infatti le prime attivazioni volontarie si possono già vedere tra i 50 e 100 ms e per quelle involontarie intorno ai 30ms. A sostegno del fatto che si possa trattare di una P300 di tipo BCI-*simple* c'è il dato del voltaggio; infatti il picco raggiunge i 10microvolt di ampiezza, valore tipico della componente. Il dato è in contrasto con i valori in ampiezza dei potenziali premotori che hanno un voltaggio molto più basso, intorno a 2-3microvolt. Il test a due mette a confronto i target tra di loro e poi ciascuno con il non-target: il risultato migliore si ottiene dal confronto tra target e non-target proprio come per il si-

stema BCI-*simple* in cui viene suggerito un solo stimolo interessante fra tanti, in questo caso se ne suggeriscono 2 facenti parte alla stessa categoria ma diversi molto dal *control*. Per questo riteniamo che i risultati del confronto siano meno significativi tra i due *target* rispetto che tra i *target* e il *control*, perché infatti il soggetto porrà molta attenzione ai primi due e “nulla” all'altro.

2.2.5.3. DIVA

I sistemi SSI basati su metodiche invasive di impianto di elettrodi a livello intra-corticale e tecniche di decodifica di tipo complesso si prefiggono l'obiettivo di ripristinare la comunicazione per le persone completamente paralizzate (locked-in).

Per la comunicazione scritta si utilizzano tastiere virtuali (Kennedy et al., 2000), per la comunicazione vocale il più promettente è il Modello DIVA (Kennedy, 2006; Brumberg et al, 2007, 2008;. Guenther et al, 2008;. Bartels et al.,2008).

Il DIVA (*Directions Into Velocities of Articulators* (F.H. Guenther, 2006)) nasce come un modello artificiale a reti neurali per la produzione del parlato. La sua particolarità sta nell'essere una macchina capace di apprendere come un bambino nella sua fase di apprendimento della lingua madre. Da un lato avremo il computer che simula l'adulto dando i *target*, dall'altro avremo il bambino, cioè la macchina. Quando un bambino impara a produrre i suoni che costituiranno il parlato, trae le informazioni da due sistemi neurali.

Il primo è un sistema di controllo di tipo *feedforward*: consiste in un programma motorio comprendente l'elenco dei suoni le istruzioni per produrli compresi luogo modo e velocità di esecuzione. Ad ognuna di queste indicazioni fa riferimento un sistema neurale preposto che in fase di produzione darà le istruzioni per la corretta esecuzione. Il secondo sistema di controllo che guida le produzioni del bambino è di tipo *feedback*: informazioni orosenso-rie, uditive (come il suono dei fonemi concatenati in sillabe).

Le prime fasi di apprendimento saranno guidate dal sistema di *feedback* infatti il sistema *feedforward* non sarà ancora del tutto formato, perché i movimenti associati al conseguente audio non sono ancora stati mappati e memorizzati, ma saranno essi a guidare le future produzioni. Nella prima fase i due sistemi si sviluppano contemporaneamente, si parte da un *babbling* per arrivare al *target* corretto.

Una volta appreso la macchina sarà in grado di produrre da sola il parlato. Grazie alla sua aderenza al sistema neuro motorio umano è stato possibile studiare e comprendere dall'interno alcuni fenomeni caratteristici del parlato come la coarticolazione, gli effetti della velocità di produzione, ecc. Il modello, di cui riportiamo uno schema in figura 46, tenta una

sovrapposizione perfetta con le strutture neurologiche dedicate alla produzione del linguaggio, questo lavoro intrapreso nei primi anni del 2000 ha condotto a dei risultati mirabili quando per la prima volta il modello è stato utilizzato per permettere la comunicazione ad un giovane americano *locked-in* (fig.47).

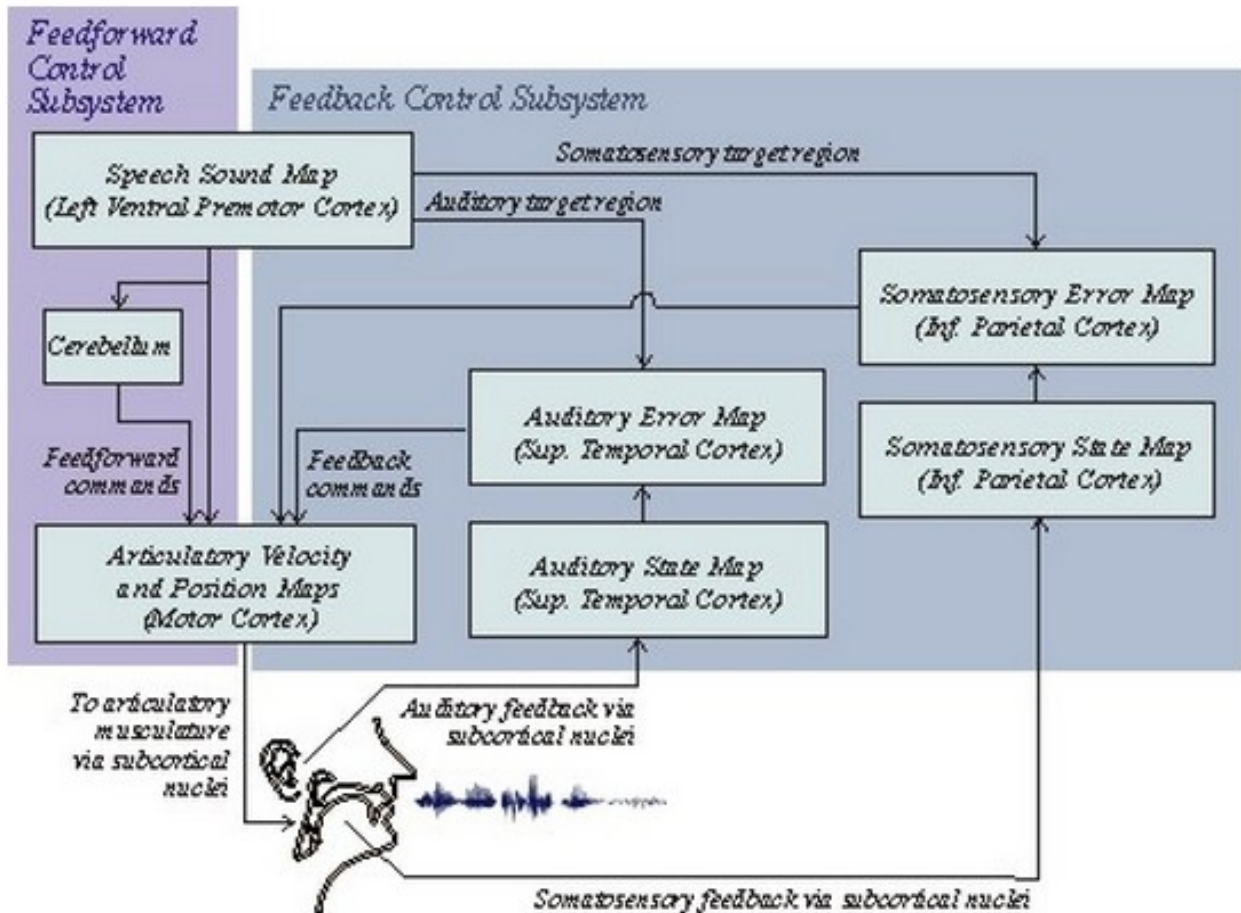


Figura 46: schema di funzionamento del DIVA

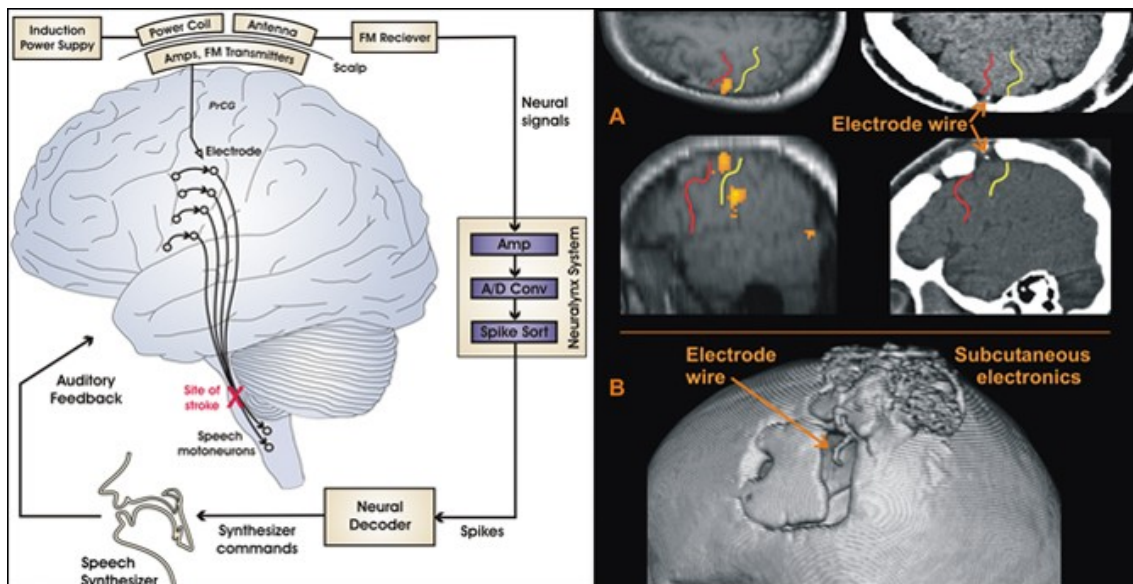


Figura 47: impianto per sistema DIVA

2.2.5.4. I problemi comuni

Posizionamento del sensore e robustezza – In tutte le tecnologie presentate, i sensori utilizzati devono essere accuratamente posizionati per ottenere la migliore risposta. i cambiamenti nella posizione del sensore all'inizio di ogni nuova acquisizione può creare problemi nell'elaborazione dei segnali e quindi nella classificazione. Anche le cuffie premontate non garantiscono un segnale costante.

Indipendenza dello Speaker - Un altro problema collegato è l'indipendenza del parlante. Infatti in questi studi le caratteristiche anatomiche e fisiologiche che vanno dai suoi muscoli articolatori fino alla densità delle connessioni sinaptiche se da un lato vengono utilizzate per raggiungere la sintesi dall'altro diventano di ostacolo ad essa a causa della forte variabilità tra soggetti.

Dizionari – Ottenere un parlato continuo su sistemi portatili è sicuramente la tappa finale, ma è del tutto ragionevole che le prime macchine renderanno possibile la comunicazione a partire da un numero limitato di unità, ma scelte in modo coprire quanto più possibile le aree semantiche della comunicazione reale di base.

In figura 48 riportiamo il confronto tra le diverse tecnologie utilizzate nell'ambito del BCI

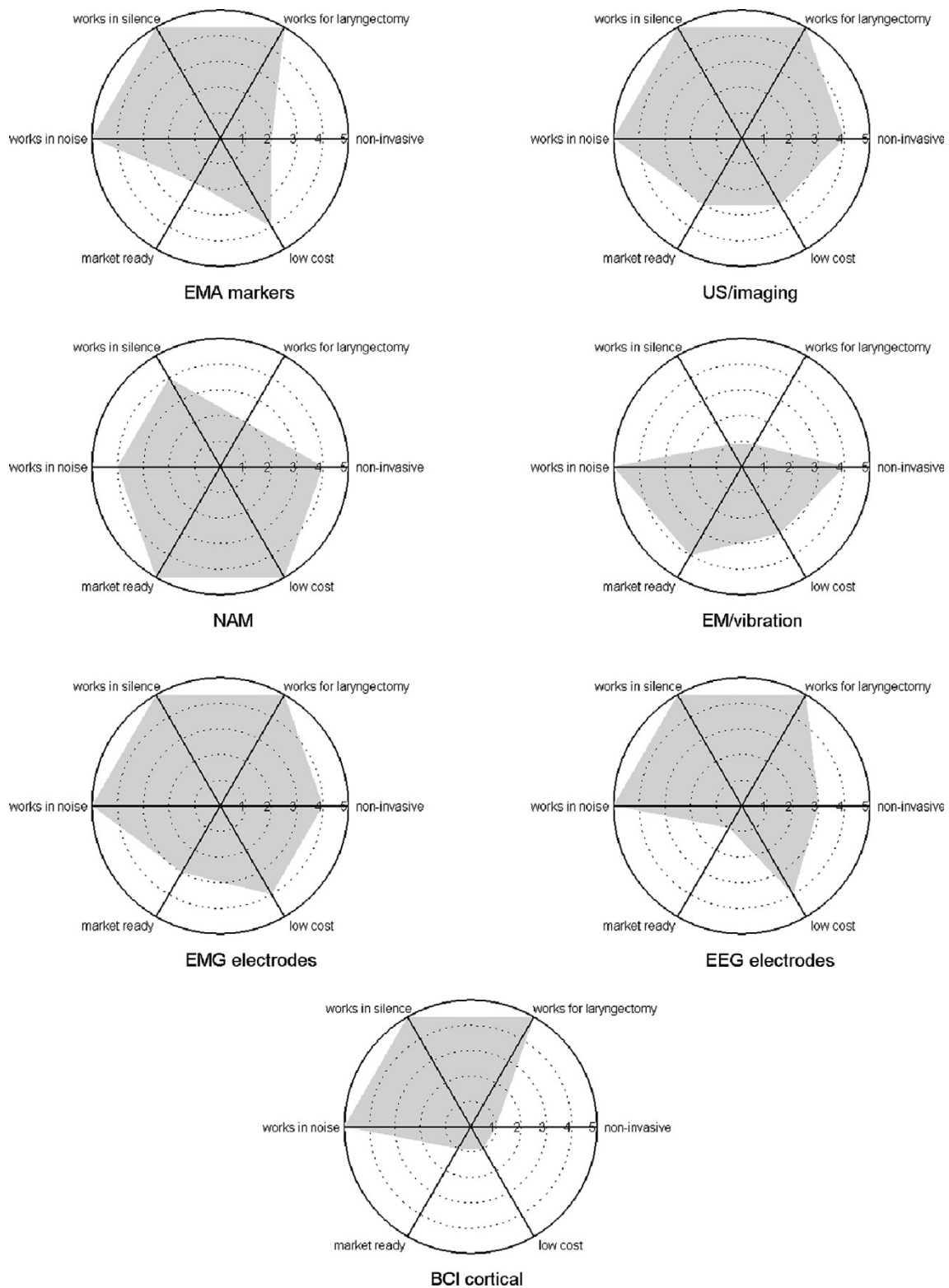


Figura 48: confronto tra tecniche BCI

3. LA RICERCA: TRE PROTOCOLLI SPERIMENTALI

3.1 Il primo protocollo

Si tratta della prima raccolta di dati, da noi effettuata, di rilevazione di segnale encefalografico mediante tecnologia EEG. Abbiamo quindi scelto di seguire un protocollo adottato per il medesimo fine dal gruppo di DaSalla (2009) al fine di avere una possibilità di comparazione nei risultati. Tuttavia sono state introdotte alcune variazioni. Il protocollo iniziale prevedeva in un primo momento che il numero dei soggetti fosse limitato a due di sesso maschile di circa 30 anni di età. Nella prima sessione, gli stimoli di tipo vocalico i fonemi /a/, /i/, /u/ sono stati presentati in sequenza su uno schermo a fondo neutro sotto forma di grafemi. La vocale sarebbe apparsa scritta sullo schermo per 2000ms. Successivamente appare lo schermo nero e in questi 2000ms il soggetto deve ripetere mentalmente la pronuncia della vocale. L'inizio e la fine della produzione saranno segnalati dal soggetto all'esaminatore con il *blink* degli occhi. Nella sessione successiva si richiede al soggetto di articolare il movimento di produzione della vocale, che apparirà sullo schermo con la medesima modalità descritta per la sessione precedente, senza emettere suono. L'inizio e la fine della produzione saranno segnalati dal soggetto all'esaminatore con il *blink* degli occhi. Si era prevista una terza sessione in cui si richiedeva di eseguire compiti in maniera alternata. Le ripetizioni per ciascuna vocale erano stabilite per un numero di 9. Nella figura 49 è rappresentata la temporizzazione.

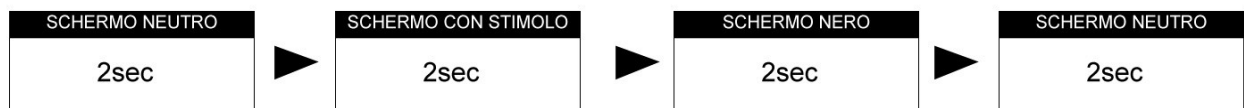


Figura 49: temporizzazione degli stimoli

A fronte delle conoscenze acquisite ci siamo resi conto che avremmo dovuto modificare il protocollo tenendo conto della strumentazione. Più in chiaro, gli esperimenti da noi condotti dovevano rispondere a diverse esigenze nella fase di analisi. Infatti abbiamo da un lato proposto ai soggetti un *trial* di esperimenti in cui si richiedeva di pensare di pronunciare la vocale rappresentata dall'icona, nel secondo *trial* si è richiesto a i soggetti non solo di pensare ma di articolare, ma senza fonazione sempre lo stesso gruppo di vocali. Abbiamo ipotizzato di poter operare una sottrazione dal primo *trial* del secondo. Ciò avrebbe potuto permetterci di individuare in maniera più chiara quello che potrebbe apparire dalla sola analisi dei tracciati del primo tipo. Il protocollo definitivo ha previsto che alle tre vocali scelte /a/ /i/ /u/ fosse inserito uno stimolo di controllo /+/ (in questo caso al soggetto era richie-

sto di “non pensare”) per permettere in fase di analisi di poter effettuare una distinzione tra pensiero di parlato e “altro”. Le vocali sono state scelte per la loro posizione estrema nel triangolo vocalico, cioè per le caratteristiche articolatorie. Inoltre se inizialmente avevamo pensato di somministrare gli stimoli tramite uno *slideshow* in ppt e inserire il segnale per la sincronizzazione nel compito (il *blink* come effettuato da Western, 2009), successivamente si è scelto di adottare il software Presentation (già in uso presso il CRIL) al fine di ottenere la randomizzazione degli stimoli e una temporizzazione automatica del segnale EEG. Ogni trial inizia con un pre-stimolo di 1 sec. in cui il soggetto deve fissare lo schermo nero con un focalizzatore. Successivamente appare lo stimolo per 2 sec. nei quali il soggetto deve svolgere il compito. Lo stimolo è la stilizzazione della forma della bocca nell'atto di produrre la vocale corrispondente: bocca aperta /a/, labbra protuse /u/ e infine labbra appena aperte e in posizione orizzontale /i/ (fig.50). Quindi appare una schermata nera ad indicare la fine del *trial*.

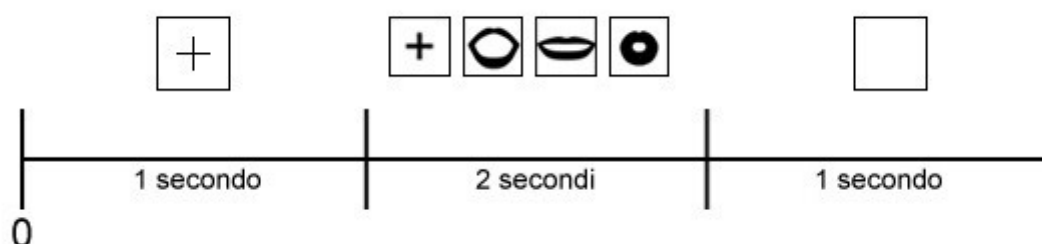


Figura 50: temporizzazione degli stimoli con relative figure

3.1.1. I soggetti

I soggetti partecipanti all'esperimento sono 9 tutti destrimani con età compresa tra i 25 e i 33 anni. Tutti sono madre lingua italiana 8 di questi sono originari e risiedono nel Salento solo uno dei soggetti è del Lazio. Il grado di istruzione è di tipo universitaria o post universitaria. Per motivi statistici abbiamo aumentato il numero di soggetti a 9 e deciso di introdurre un 33% di soggetti femminili. Tutti dichiarano di non aver problemi o di non aver avuto problemi di carattere neurologico, rilasciando una dichiarazione di consenso informato per questa sperimentazione (app.).

3.1.2. La raccolta del *corpus*

I segnali sono stati raccolti presso il laboratorio del Cril di Lecce nei mesi di giugno e luglio del 2010. In totale sono state registrate 50 ripetizioni per ciascuno dei 4 compiti per un totale di 200 epoche a soggetto. Abbiamo così ottenuto 50 campioni per ogni tipo di stimolo per entrambe le sessioni (pensato articolato e solo pensato). Per la registrazione abbiamo utilizzato il software Brain Vision Recorder, con amplificatore Brain Amp. Abbiamo scelto di usare una cuffia a 64 canali di cui riportiamo i settaggi di registrazione di seguito (figg. 51A-e). Hardware e software sono della Brain Products. L'aver scelto di usare tutti e 64 i canali della cuffia è stato dettato dalla volontà di avere una migliore risoluzione spaziale seppur solo al livello della superficie, anche se questo ha comportato un allungarsi dei tempi sia per il montaggio degli elettrodi sulla cuffia sia per il settaggio delle impedenze dei singoli elettrodi una volta fatta indossare la cuffia al soggetto: si stima circa un ora in media. Le registrazioni hanno avuto una durata totale di 30 minuti per ciascun soggetto. Gli elettrodi di riferimento in questo caso sono: C3, C4, Cz e Fz.

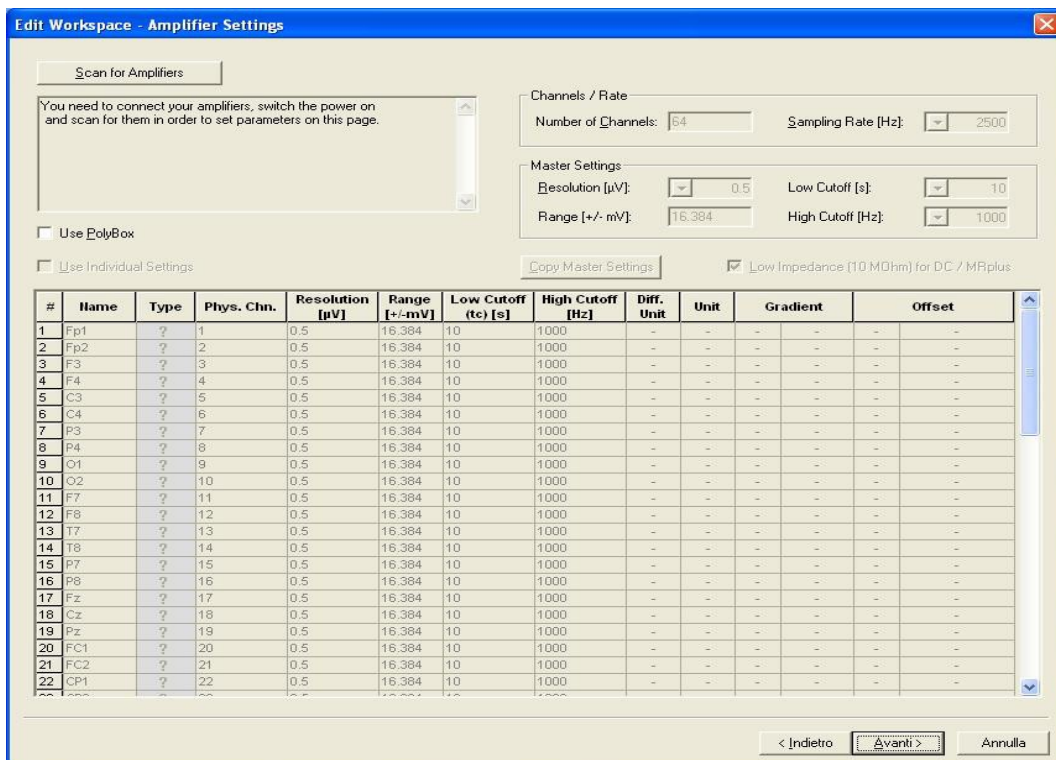


Figura 51a: schermata di settaggio amplificazione

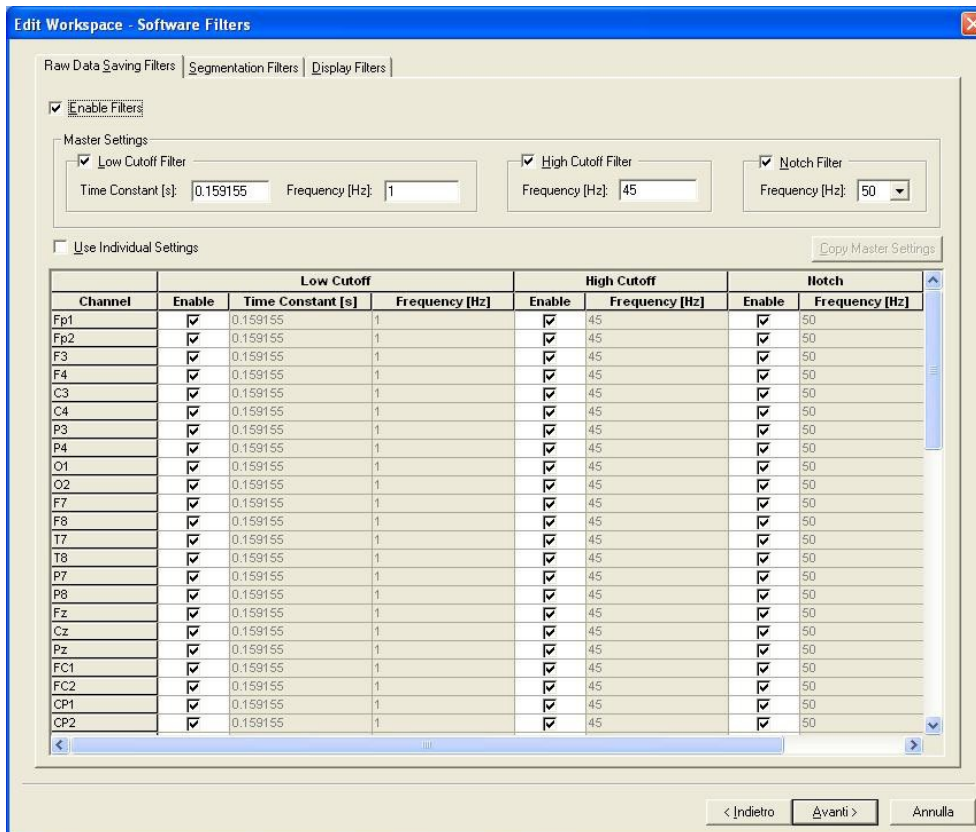


Figura 51b: schermata di settaggio filtri

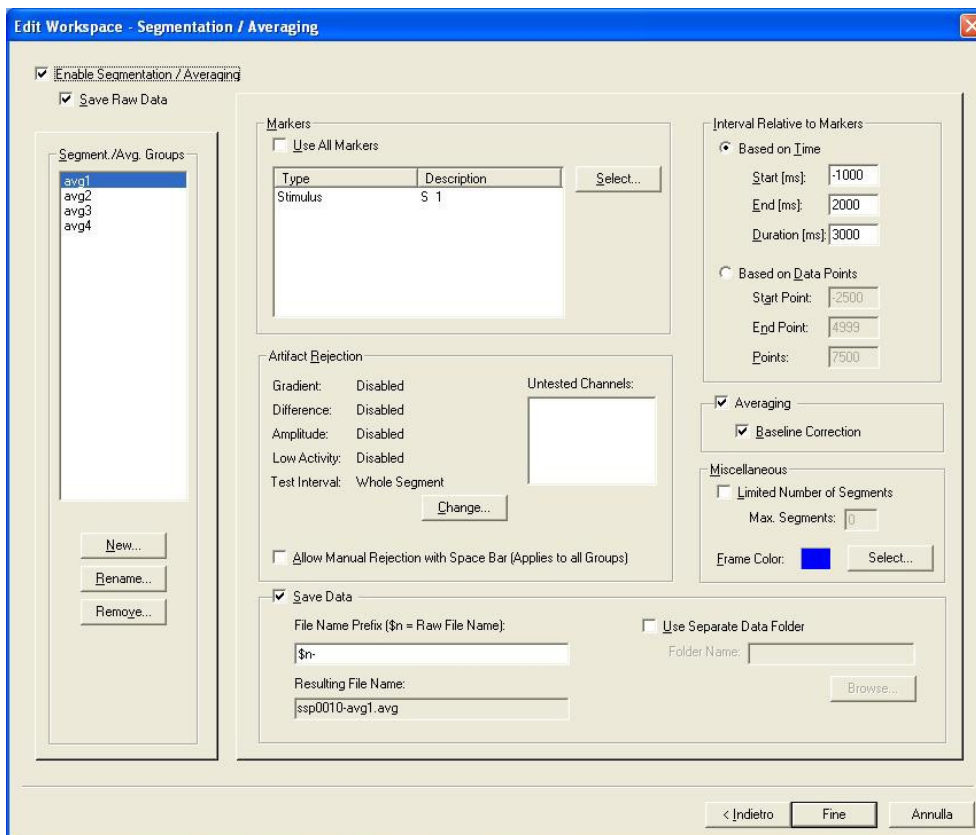


Figura 51c: schermata di settaggio acquisizione online dell'average

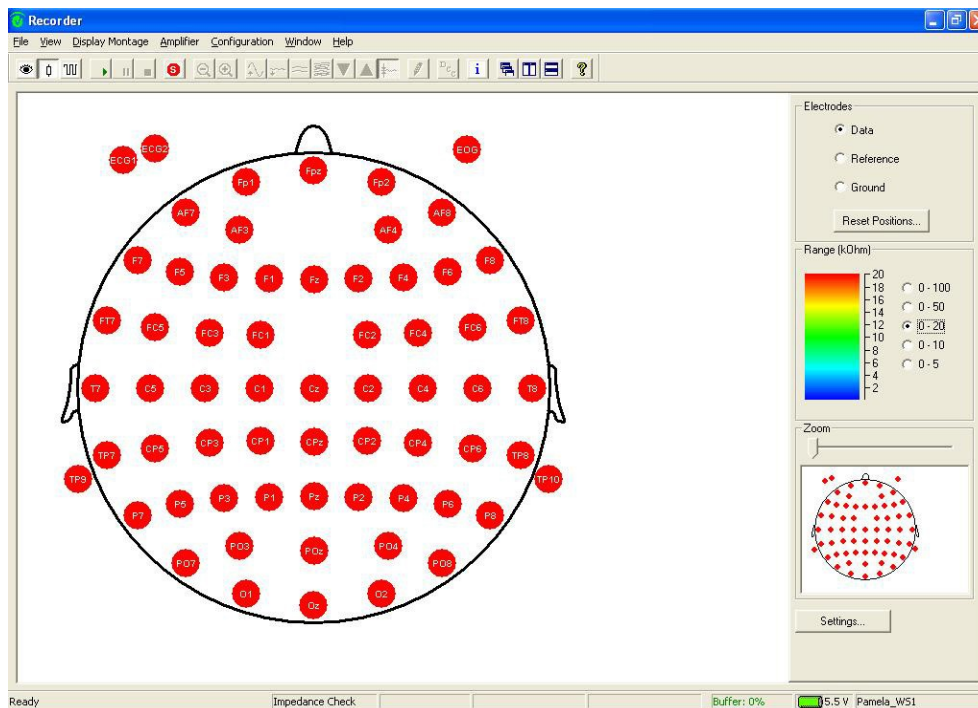


Figura 51d: schermata iniziale per il check dei canali di registrazione

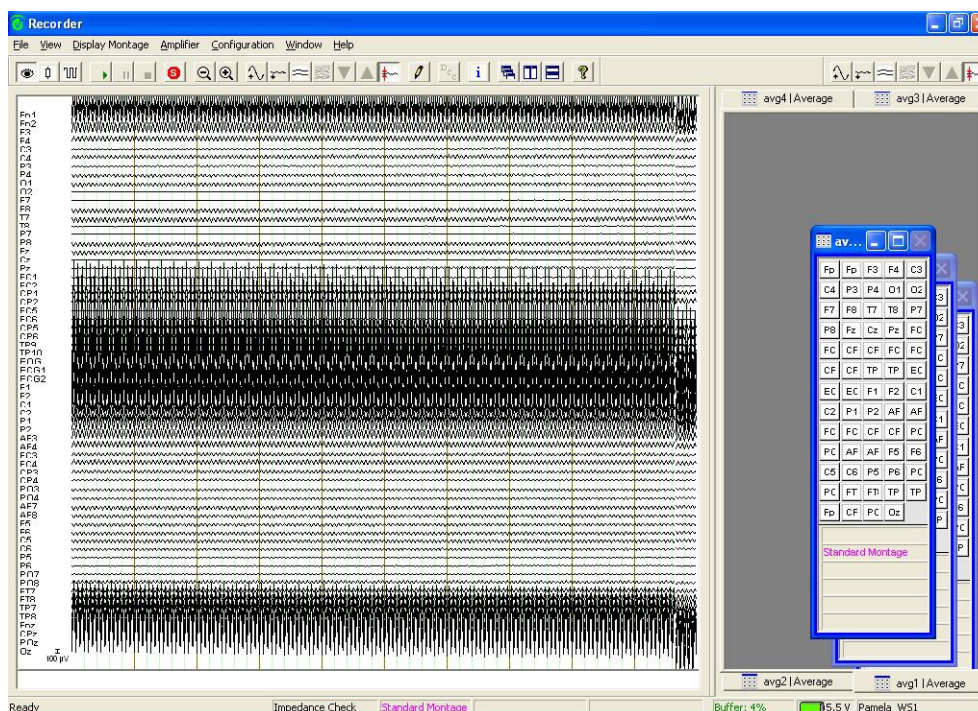


Figura 51e: schermata al momento dell'accensione dei canali con cuffia non montata

3.1.3. Gli strumenti: *Presentation*

Si tratta di un ambiente per la scrittura di script per la presentazioni di stimoli, sviluppato per gli esperimenti nell'ambito delle neuro-scienze dall'egg all'fMRI. Ha la possibilità di comunicare, attraverso le porte parallele, registrando così l'input dei trigger.

E' uno strumento che permette di sviluppare protocolli complessi con randomizzazione degli stimoli e con molteplici possibilità di somministrazione degli stessi (audio e video) E la possibilità di usare molte periferiche di risposta (microfono per l'acquisizione della voce) anche in contemporanea. Questa sua flessibilità comporta una difficoltà nella programmazione. Infatti l'interfaccia come si può vedere dalle figure non è propriamente *user friendly* e molto complesso realizzare lo script. Altri problemi riscontrati sono la difficoltà di trovare istruzioni base anche su i forum; infatti trattandosi di un programma di nicchia si trova scarsa o nulla documentazione. Abbiamo quindi creato lo script partendo da un esempio molto semplice e introducendo le modifiche necessarie per adattarlo alle nostre esigenze. Lo script utilizzato per somministrare gli stimoli è il seguente:

```

scenario = "silentspeech_m00";
default_font_size = 40;
write_codes = true;
pulse_width = 20;
default_trial_duration = 2000;

#articolato no emissione
begin;

picture {} default;

trial {
  trial_duration = 1000;

  picture {
    text { caption = "+"; };
    x = 0; y = 0;
  };
  } instruction_trial1;
trial {
  trial_duration = 1000;

  picture {
    text { caption = "respira,fermo,concentrati"; };
    x = 0; y = 0;
  };
} instruction_trial;

trial {
  trial_duration = 1500;
  picture {
    text { caption = " "; };
    x = 0; y = 0;
  };
} resting_trial;
array {
  LOOP $i 4;
  $k = '$i + 1';
  trial {
    picture {
      bitmap { filename = "mvoc$k.jpg"; };

```

```

x = 0; y = 0;
};
code = "mvoc$k";
port_code = $k
};
ENDLOOP;
} main_trials;

begin_pcl;

sub

  show_batch( int mvoc1, int mvoc4 )
begin
  loop until
    mvoc1 > mvoc4
  begin
    main_trials[mvoc1].present();
    mvoc1 = mvoc1 + 1
  end
end;

main_trials.shuffle();

# 10 ripetizioni
instruction_trial.present();
instruction_trial1.present();
resting_trial.present();

show_batch( 1, 1 );
resting_trial.present();
instruction_trial1.present();
show_batch( 2, 2 );
resting_trial.present();
instruction_trial1.present();
show_batch( 3, 3 );
resting_trial.present();
instruction_trial1.present();
show_batch( 4, 4 );
resting_trial.present();

```

3.1.4. Gli strumenti: *Brain Vision*

Questo software è dedicato alla registrazione dei segnali elettroencefalografici, acquisisce un numero di canali pari a quelli dell'amplificatore collegato. Ha delle interfacce chiare ed è molto agevole da utilizzare. Si parte con il riconoscimento dell'amplificatore quindi con lo stabilire il tipo di montaggio di cuffia (come in tutti i software di questo genere), si selezionano il numero e il tipo di canali da attivare e la risoluzione in Hz. A questo punto possiamo scegliere di applicare dei filtri in ingresso, soluzione da noi adottata in questo primo protocollo. E' possibile selezionare dei filtri di rimozione in entrata degli artefatti sia oculari che muscolari. Inoltre per i potenziali evocati è utile utilizzare la funzione di *average* automatico. Infatti con questa funzione già durante la fase di registrazione si creano dei file paralleli tanti quanti sono i tipi di stimoli, ad esempio nel nostro caso 4, ciò è possibile grazie ai *marker* (o *trigger*) con cui avremo codificato ogni singolo stimolo. Purtroppo nel nostro caso per motivi di formato, non sono stati utilizzati. Ma potrebbero essere utilizzati in seguito adottando un altro software per l'analisi.

3.2. Il secondo protocollo

Per questo protocollo ci siamo concentrati ancora di più sui concetti di teoria motoria e di neuroni specchio, per cui dall'ascolto alla decodifica si arriva attraverso una "riproduzione" interna dei gesti articolatori. L'idea è stata di sottoporre i soggetti allo stimolo audio e poi al compito di produzione e al pensiero di produzione. Secondo quanto detto sopra dovremmo individuare una regione comune e anche delle attivazioni sovrapponibili nello svolgimento tutti e 3 i compiti. Il *trial* comincia con lo stimolo audio in cuffia per la durata di 250ms che riproduce una delle 3 vocali (/a/, /i/, /u/) oppure lo stimolo di controllo (una sinusoide a 440 Hz) e poi appare una schermata verde con una durata di 1s, prontamente il soggetto deve produrre la vocale, internamente nel caso della prima sessione oppure vocalizzandola nel caso della seconda sessione. Segue il simbolo [+] con una durata di 1s, all'apparire di questo il soggetto interrompe il compito. Gli stimoli sono stati proposti in maniera randomica ciascuno è proposto 100 volte; infatti si è deciso di raddoppiare il numero di ripetizioni per stimolo al fine di abbattere in maniera più efficace il rumore di fondo. La Il suono emesso è registrato in contemporanea alla registrazione eeg.

La riduzione dei tempi soprattutto per la produzione del compito è stata dettata dalla necessità di arginare il più possibile il problema del *jittering*, ma anche per non affaticare troppo i soggetti riducendo così la durata complessiva delle sedute di registrazione.

3.2.1. I soggetti

A fronte dell'esperienza fatta nella precedente campagna di registrazione di dati e della successiva fase di analisi che ha portato alla luce il problema della grande variabilità in termini di risposte neurofisiologiche tra soggetto e soggetto, per questa campagna di registrazione abbiamo ridotto il numero di soggetti al numero di 3 solo uomini di età compresa tra i 26 e i 34 anni. L'istruzione è sempre di tipo universitario o post universitario, si tratta di soggetti di madrelingua italiana destrimani: solo uno non ha partecipato al precedente esperimento. Si tratta di volontari, tutti rilasciano una dichiarazione in cui sostengono di non aver problemi o di non aver avuto problemi di carattere neurologico e firmano una dichiarazione di consenso informato per questa sperimentazione (app.).

3.2.2. La raccolta del *corpus*

I segnali sono stati raccolti presso il laboratorio del Cril di Lecce nel mese luglio del 2011 impiegando una sola settimana. In totale sono state registrate 100 ripetizioni per ciascuno dei 4 compiti per un totale di 800 epoche a soggetto. Abbiamo così ottenuto 100 campioni per ogni tipo di stimolo per entrambi i tipi di *trial*. Anche in questa occasione per la registrazione abbiamo utilizzato il software *Brain Vision Recorder*, con amplificatore *Brain Amp* abbiamo scelto di usare una cuffia a 64 canali, i settaggi di registrazione sono gli stessi del precedente esperimento (fig.51). *Hardware* e *software* sono della *Brain Products*. Nuovamente si è scelto di registrare a 64 canali e questa volta è stato dettato dalla volontà di poter utilizzare in fase di filtraggio l'ICA. Le registrazioni hanno avuto una durata totale di circa 25 minuti per ciascun soggetto. Gli elettrodi di riferimento sono aumentati rispetto al protocollo 1°; infatti si è deciso di prestare particolare attenzione a quelli posizionati lungo tutta la linea che va da un estremo dell'orecchio all'altro, ma lo vedremo in maniera più approfondita in fase di analisi. Alcuni elettrodi a causa dell'usura, ricordiamo che la pasta conduttrice è di tipo abrasivo hanno perso il contatto o hanno introdotto rumore, sul canale in fase di registrazione. Questo è stato ovviato sostituendo l'elettrodo sul campo oppure in un caso rimuovendo dalla registrazione l'intero canale.

3.2.3. Gli strumenti: *Brain Vision*

Per la registrazione abbiamo nuovamente utilizzato la strumentazione di base del CRIL adottata precedentemente. I filtri non sono stati inseriti in fase di registrazione. Per le schermate con le tarature utilizzate per la registrazione rimandiamo alle figure 51a-e.

3.2.4. *Gli strumenti: Eprime*

Per questa serie di esperimenti, per la somministrazione degli stimoli abbiamo utilizzato un software diverso, ciò è stato dettato dalla disponibilità in termini di licenza del software acquistata dal CRIL. Questo software per la presentazione e sincronizzazione, funziona sia tramite interfaccia grafica, sia attraverso programmazione. Ciò ci ha permesso di poter scrivere il nostro esperimento, pur non avendo mai visto prima il software, in pochi giorni. E' molto rapido e flessibile, permette di inserire contemporaneamente per ogni stimolo figure audio oltre alla possibilità di registrare la traccia audio e video, in simultanea e in sincrono con la registrazione EEG (figg.52 a-d).



Figura 52a: soggetto al momento del montaggio della cuffia

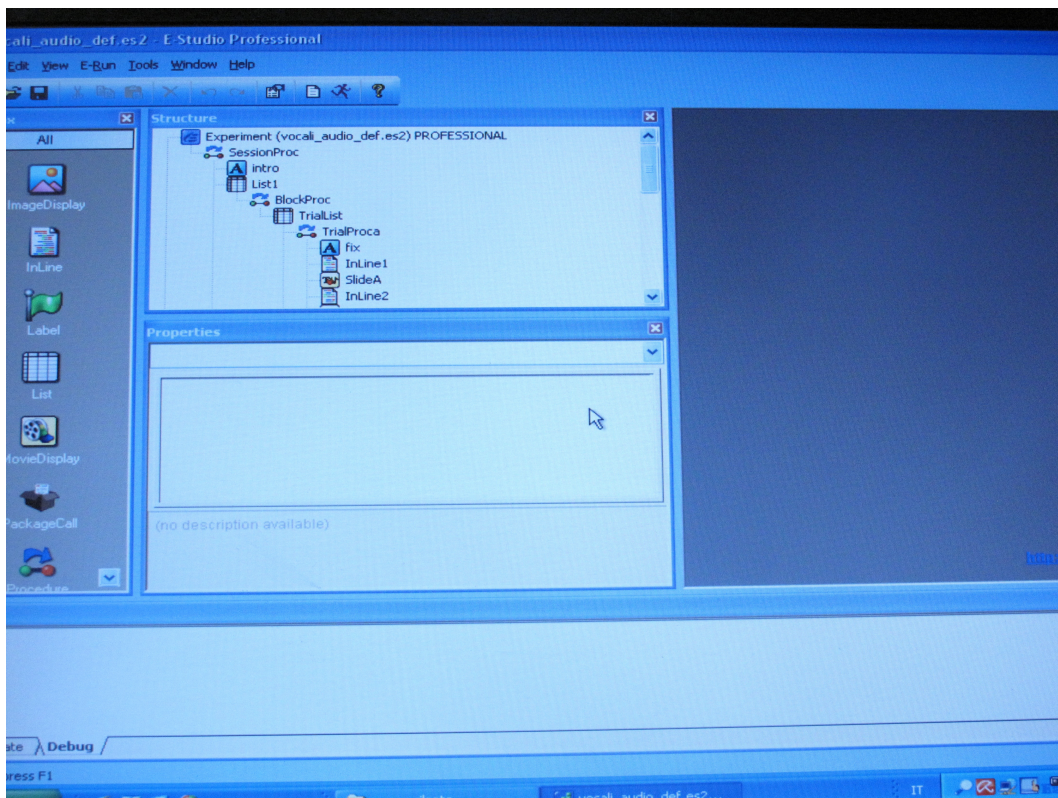


Figura 52b: schermata riassuntiva del protocollo di stimolazione

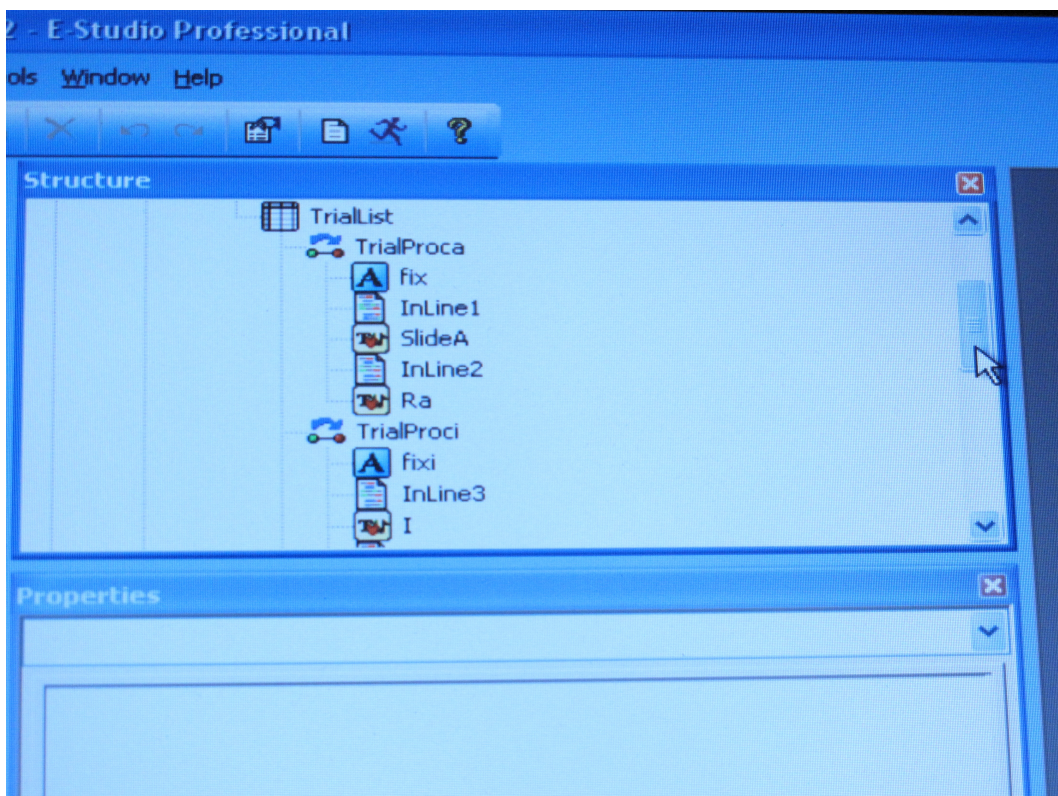


Figura 52c: schermata di dettaglio del contenuto del protocollo

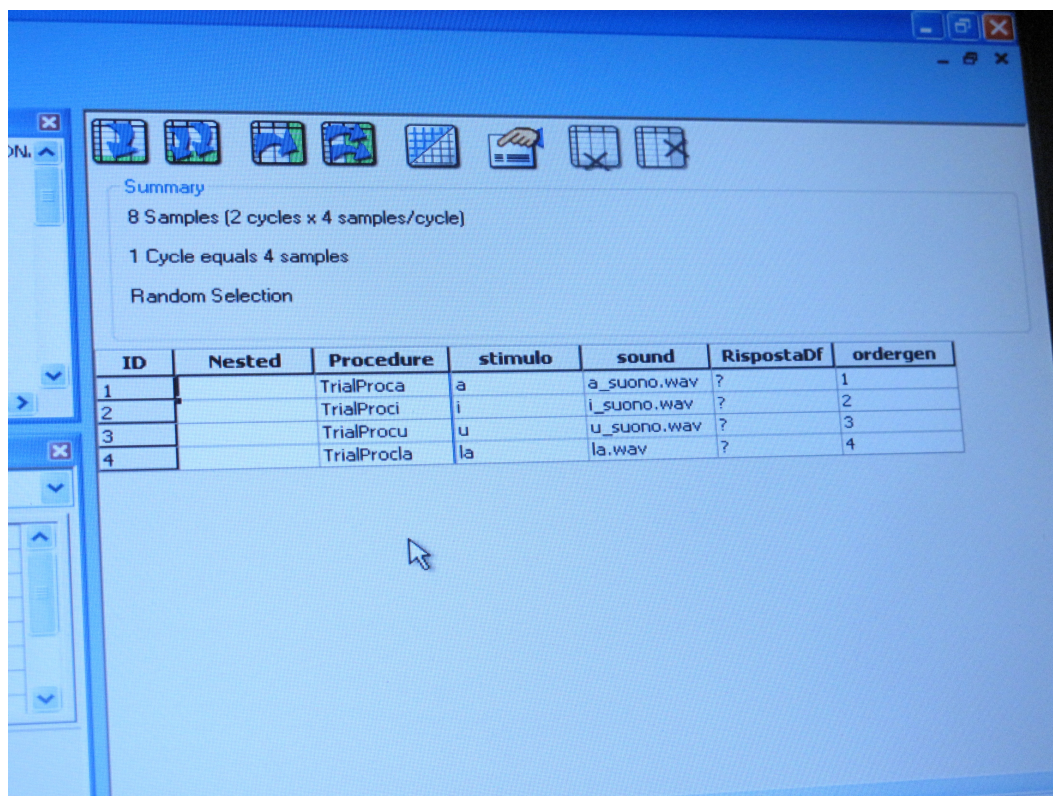


Figura 52d: schermata di dettaglio di selezione dei file di stimolo e impostazioni

3.3. Il terzo protocollo

Questo protocollo è nato dall'esperienza di tirocinio presso U.O di Neurologia presso l'Ospedale di Belcolle di Viterbo. La possibilità di poter usufruire non solo delle apparecchiature ma anche delle competenze dei tecnici e dei medici del reparto ha reso questa esperienza di alto valore formativo. L'assistere alle numerose registrazioni sia con EEG che con Pev, come anche l'assistere alle refertazioni ci ha permesso di acquisire competenze che solo studi specifici offrono. Per motivi dipendenti dalla struttura e dalle apparecchiature abbiamo modificato il protocollo rendendolo ancora più snello e simile alle procedure che sono utilizzate per la registrazione dei potenziali evocati in ambito diagnostico. Sono stati utilizzati 5 elettrodi sullo scalpo Cz, C3, C4, T3, T4 e 2 elettrodi per l'EOG con 2 referenze. Il protocollo viene eseguito ad occhi chiusi, ciò permette di ridurre gli artefatti, il soggetto è sottoposto ad una stimolazione luminosa che dà il *timing* della produzione del *target* (le 3 vocali /a/, /i/, /u/). Il target di controllo è registrato con il soggetto sottoposto allo stimolo luminoso, con gli occhi chiusi e rilassato. Ogni target è stato ripetuto 100 volte. Lo stimolo è somministrato ogni 1000 millisecondi. La prima sessione prevede che il soggetto ripeta internamente ciascuna vocale, la seconda sessione prevede che il soggetto ripeta vocalizzando la vocale riducendo al minimo i movimenti. Le vocali non sono rando-

mizzate.

Come si noterà questo protocollo si discosta per la procedura dai precedenti, è molto rapido e richiede molta concentrazione da parte del soggetto; infatti il compito viene eseguito in una stanza non insonorizzata quindi in una situazione molto naturale in cui esistono stimoli esterni che il soggetto deve porre in secondo piano. La differenza più rilevante è data dalla scelta dei soggetti, che abbiamo ridotto ad uno solo, ma sottoponendolo alla sperimentazione in giornate differenti per un totale di 3 esperimenti.

3.3.1. I soggetti

Il soggetto è di sesso femminile di 30 anni di età, destrimana di madre lingua italiana. Si tratta di un soggetto con istruzione di tipo post universitario è un soggetto volontario che ha rilasciato una dichiarazione in cui sostiene di non aver problemi o di non aver avuto problemi di carattere neurologico. Ha firmato una dichiarazione di consenso informato per questa sperimentazione (app.).

3.3.2. La raccolta del *corpus*

Gli esperimenti sono stati condotti durante il periodo di tirocinio presso l'U.O. di Neurologia e Neurochirurgia dell'ospedale di Belcolle. Abbiamo acquisito un soggetto in 3 giorni diversi, somministrando 2 *trial* con 4 *task* ripetuti 100 volte ognuno. Ogni esperimento è durato 30 minuti compreso il montaggio e settaggio della cuffia ad opera dei tecnici.

La registrazione appare continua in basso appare la linea dei *marker* degli stimoli flash. La registrazione si interrompe ogni 100 ripetizioni, ma è registrata su un'unica traccia. I file generati sono di due tipi: il formato "*.doc" dove sono inseriti i dati del soggetto e il file con estensione "*.eeg" contenente la registrazione in formato digitale.

3.3.3. Gli strumenti: ATES

Il software per le registrazioni è Neurotravel della Ates, si tratta di un prodotto specifico per le indagini cliniche con parametri pre-impostati che agevolano il compito del tecnico ma che sono più limitanti, in un ambito di ricerca. La cuffia utilizzata è di tipo regolabile con elettrodi ponte imbevuti di soluzione salina per i canali di interesse e per la terra mentre per gli EOG e le referenze si è adottato gli elettrodi a coppetta con pasta conduttrice adesiva. Per i parametri di registrazione riportiamo anche le schermate dei settaggi con commenti in nota di figura. La registrazione è stata effettuata con un campionamento a 250Hz e un *notch* per la frequenza di rete, la sensibilità è di 5 microvolt.

Abbiamo registrato 5 canali sullo scalpo Cz, C3, T3, C4, T4 con referenza sul lobo sinistro e monitorato i movimenti oculari tramite EOG con due elettrodi più referenza propria (figg. 53a-f). La terra è stata posizionata in Fpz, (l'elettrodo nominato è Oz).



Figura 53a: il soggetto con la cuffia montata

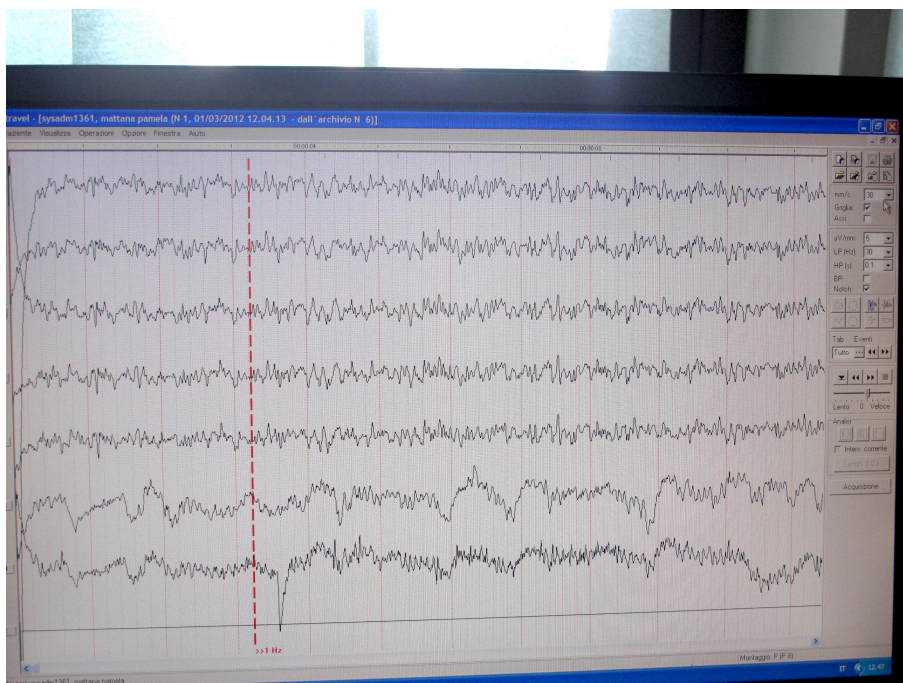


Figura 53b: schermata durante la registrazione dell'esperimento

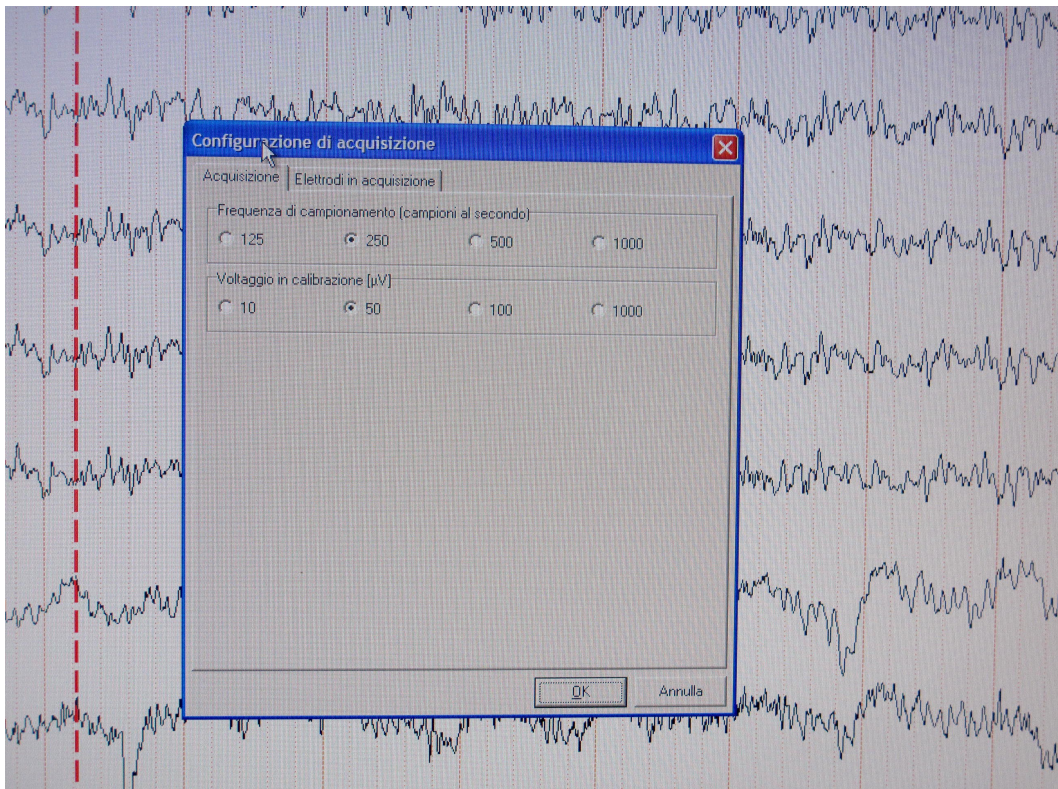


Figura 53c: schermata con impostazioni di registrazione

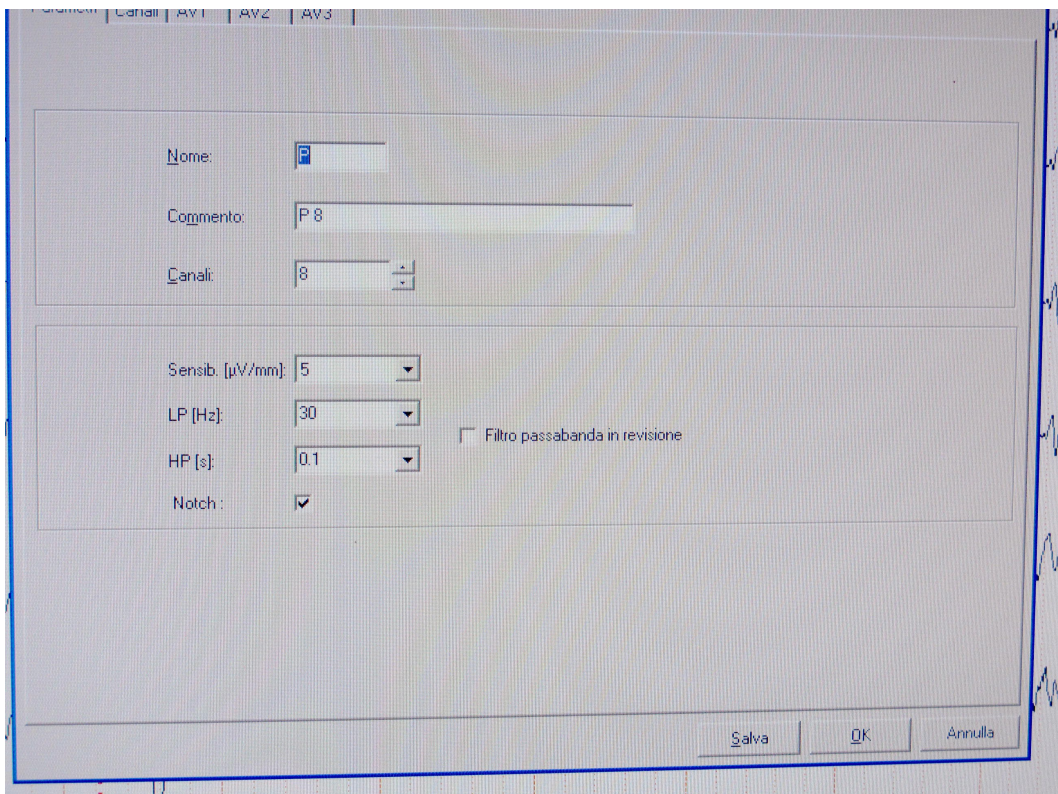


Figura 53d: schermata con impostazioni di filtraggio

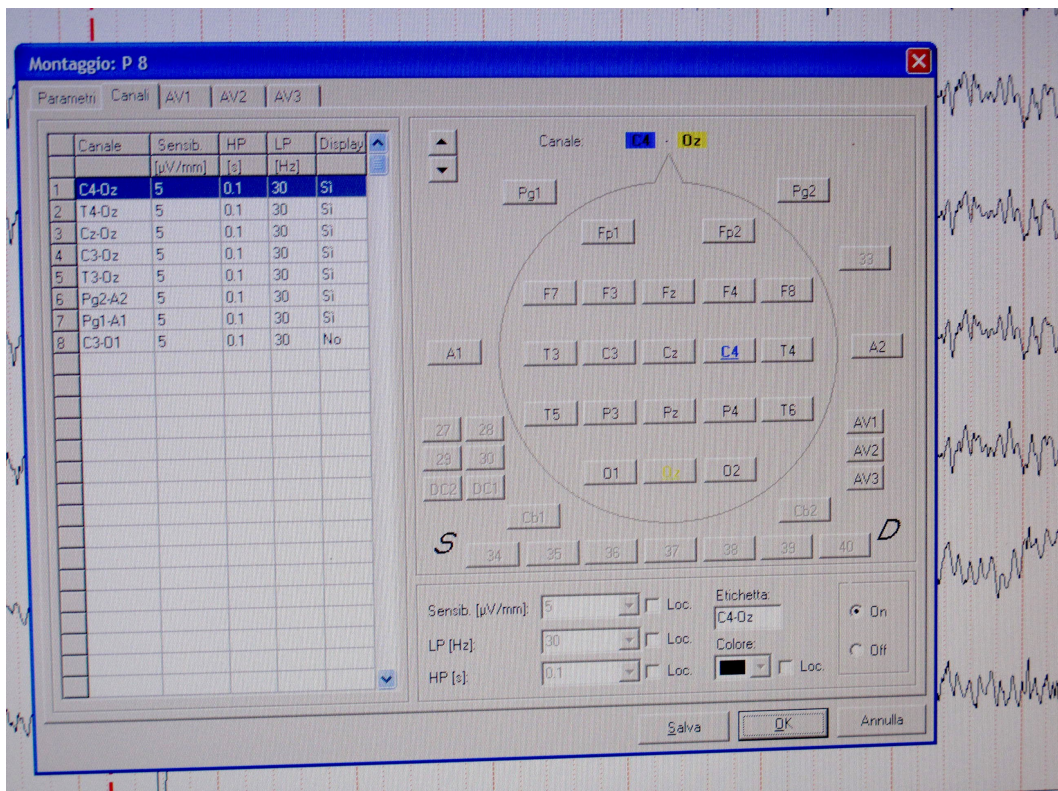


Figura 53e: schermata con impostazioni di montaggio e referenze

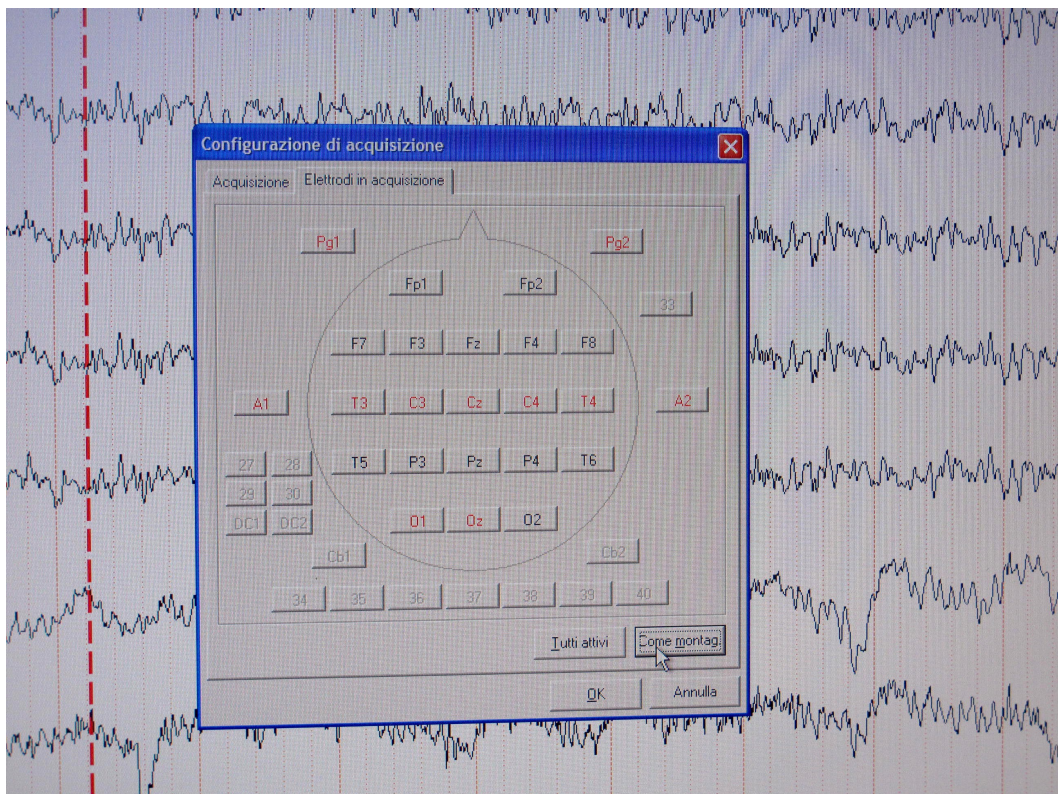


Figura 53f: schermata per la visualizzazione degli elettrodi

3.4.I. materiali

La parte più consistente di questo lavoro, sia in termini di tempo che di quantità numerica, sono di tipo digitale e renderli in formato cartaceo sarebbe inutile. Si tratta di file di diversa natura che abbiamo inserito in un CD allegato che deve essere inteso come parte integrante del lavoro di tesi. Di seguito abbiamo scelto di inserire l'elenco di questi dati e la loro breve descrizione.

Primo esperimento:

Le registrazioni sono divise in due cartelle una per la modalità articolato e una per la modalità non articolato. In ciascuna cartella ci sono 9 file con estensione "*.eeg", con estensione "*.vhdr" e "*.vmrk"

Secondo esperimento:

Le registrazioni sono divise in due cartelle una per la modalità articolato e una per la modalità non articolato. In ciascuna cartella ci sono 3 file con estensione "*.eeg", con estensione "*.vhdr" e "*.vmrk". In questo caso abbiamo anche la registrazione della produzione audio, organizzati in una cartella "audio" contenente 3 sottocartelle, una per soggetto dove sono contenuti i file in formato "*.wav" delle singole produzioni.

Terzo esperimento:

Le registrazioni hanno prodotto 2 file con formato ".eeg" (cioè il formato proprietario Ates) per l'analisi sono stati convertiti in due formati ambedue leggibili con EEGLAB, si tratta di "*.EDF" ed "*.EDF+". Il secondo formato permette di visualizzare anche le informazioni aggiuntive ad esempio i trigger e le annotazioni messe dall'operatore nel corso della registrazione.*

4. DAL TRATTAMENTO DEI *CORPORA* AI DATI

4.1 Gli strumenti: EEG Lab

EEGLAB è un *toolbox free* di MATLAB distribuito da GNU e GPL, scaricabile dal *link* <http://sccn.ucsd.edu/eeglab> . Ricordiamo che per alcune funzioni base come ad esempio la visione dei canali e alcuni *plugin* EEGLAB può funzionare senza MATLAB che necessita di una licenza a pagamento usando un emulatore. La sua funzione è quella di elaborare i segnali provenienti principalmente da EEG, magnetoencefalografia (MEG), ma offre l'opportunità di elaborare anche altri segnali neurofisiologici. La prima versione di questo elaboratore nasce nel 2004 da Scott Makeig e Arnaud Delorme e fino ad oggi, proprio per la sua natura libera è tra i più utilizzati nei laboratori di ricerca, inoltre dalla sua nascita e sviluppo hanno contribuito a vario titolo circa 200 ricercatori.

EEGLAB offre sia la possibilità di visionare la traccia grafica che di modificarla scartando porzioni di segnale; è possibile applicare filtri diversi per la rimozione degli artefatti, inoltre i vari *plugin* permettono di effettuare anche analisi complesse come l'ICA e diverse modalità di visualizzazione dei dati. EEGLAB consente agli utenti di importare i dati elettrofisiologici in formati digitale effettuando *preprocessing*. Vi sono inoltre una serie di comandi appositi per lo studio degli Erp.

Vediamo nello specifico come si presenta e alcune funzioni adottate da noi ai fini della ricerca. Una volta installato EEGLAB all'interno di MATLAB è possibile aprirlo direttamente digitando l'istruzione all'interno della finestra di comando a questo punto si aprirà la schermata di accesso dalla quale possiamo accedere all'apertura dei file sorgente oppure dei *data-set* cioè dei file già lavorati. Nel caso dei file sorgente è possibile scegliere di aprire solo alcuni canali e compiere operazioni di *downsampling*, è possibile effettuare, anzi è consigliabile, una *channel location* e anche avviare l'ICA per poter lavorare poi successivamente alle fasi di filtraggio. Queste operazioni richiedono un consistente tempo macchina, quindi è opportuno lavorare con solo gli elettrodi necessari.

E' possibile seguire le fasi di elaborazione sul *workspace* di MATLAB.

Una volta caricato il file di nostro interesse è possibile visionarlo con il comando *scroll-channel* ed è quindi possibile selezionare sia la finestra temporale in cui muoversi sia il numero di canali da visionare e che scorrono sulla verticale. Questa opzione è molto utile quando si hanno molti canali e registrazioni molto lunghe.

Questo *Toolbox* è in grado di inserire i *marker* corrispondenti agli stimoli che verranno visualizzati come delle linee verticali in corrispondenza dello start dello stimolo e avranno il codice da noi assegnato in fase di registrazione, corrispondente allo stimolo stesso. A questo punto è possibile per noi estrarre le epoche: bisogna selezionare lo stimolo e il *range*

in ms quindi nominare il file. Avremo così un file per ogni tipo di stimolo contenete tutte le ripetizioni. Sarà possibile effettuare l'*average* esaminandolo nel tempo oppure in frequenza, ma anche attraverso mappe tridimensionali. Attraverso la simulazione video è anche possibile ricostruire il movimento nel tempo delle attivazioni sullo scalpo.

Per chi fosse interessato all'utilizzo di questo toolbox rimandiamo al link <http://sccn.ucsd.edu/wiki> si tratta del tutorial, ricordando che è uno strumento *in fieri* per cui è molto utile anche andare a reperire, sui siti collegati e nel forum dedicato, le informazioni aggiuntive.

4.2. Preprocessing: rimozione degli artefatti e filtraggi

Sotto il termine *preprocessing* sono raccolte tutte quelle procedure di elaborazione del segnale che hanno come fine il miglioramento del rapporto segnale rumore (SNR); si tratta di un aspetto essenziale nell'estrazione di un Erp. Infatti solo dopo l'uscita del segnale dalla fase di *preprocessing*, posso essere estratte le *features* che verranno utilizzate poi in fase di classificazione.

Al fine di rilevare la componente Erp di nostro interesse dal segnale EEG si utilizzano differenti tecniche che pervengono tutte però dallo stesso modello di riferimento per cui il segnale registrato Y è la somma del potenziale evocato U e di tutte le altre componenti EEG X . Quindi $Y=U+X$. Dove X è l'attività asincrona di fondo del segnale elettroencefalografico. Per arrivare ad Y cioè ad estrarre l'Erp dal fondo esistono differenti tecniche che agiscono sui parametri sia della frequenza che del tempo. La metodica di *preprocessing* a cui è stato sottoposto il nostro materiale ha comportato la revisione di tutto il materiale registrato, il *software* utilizzato in questo caso è EEGLAB. Sono stati scartati i segnali di prova e le false partenze quindi sono stati rinominati i file così da poter creare un corpus. Si è provveduto alla riduzione della banda portando il segnale da un iniziale 2058 Hz a i 250 Hz . Una riduzione così drastica si è resa necessaria al fine di poter gestire il trattamento dei segnali senza appesantire l'hardware.

4.2.1. Il problema degli artefatti

Con il termine artefatti vengono denotati una serie di segnali che disturbano la rilevazione del segnale Erp. Gli artefatti compongono il segnale elettrofisiologico, e li consideriamo rumore, solo in rapporto al nostro obiettivo. Nella figura 54 sono rappresentati artefatti fisiologici di segnale non encefalico. Le soluzioni posso essere molteplici ciascuna però comporta dei rischi per questo è importante scegliere la soluzione in base al tipo di ricerca e valutando volta per volta i rapporto costi benefici. Esistono soluzioni come lo scartare ma-

nualmente le singole epoche con artefatti che comportano la diminuzione della riduzione del corpus di dati e quindi della perdita di informazioni. E' possibile utilizzare un metodo automatizzato che scarti le epoche sulla base di una soglia di ampiezza, ma questo può compromettere i risultati. E' opportuno adottare soluzioni che siano un compromesso o che con particolari accorgimenti e tecniche non compromettono il segnale utile. Seguendo le indicazioni della letteratura contemporanea per i nostri *copora 1°*, *2°* si è proceduto con il filtraggio *off-line* degli artefatti relativi al *blink* oculare scegliendo il filtro tipo *Blind Source Separation*. Il filtro interviene facendo la stima delle componenti la cui combinazione lineare è l'EEG registrato.

Il numero di tali componenti è legato al numero di canali di registrazione. Una volta individuata la componente di artefatto, essa può essere rimossa e il segnale può essere ricostruito senza artefatto (BSS, fig.54)

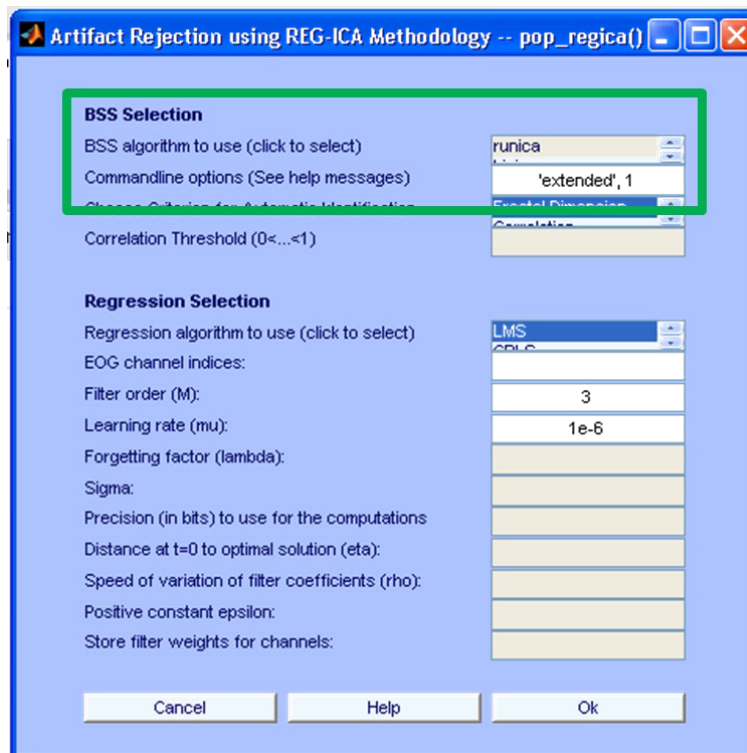


Figura 54: schermata di impostazione per il filtro BSS

E' stato applicato un filtro passa-basso a 45Hz e una *baseline-correction* di *default* come consigliato dai progettisti del software. Il problema della rimozione degli artefatti oggi, è di norma risolto utilizzando l'ICA.

4.2.2. ICA, independent Component Analysis

La formulazione matematica del problema della separazione cieca delle sorgenti dei segnali utilizzando l'equazione ICA è la seguente:

$$x_i = \sum_{j=1}^m a_{ij}s_j, \quad i=1\dots n$$

Le variabili osservate x_i sono modellate come somma lineare di variabili nascoste.

È una forma di analisi fattoriale. La matrice a_{ij} è una costante s_j sono i fattori nascosti random chiamati "componenti-independenti" oppure "segnali originali". Questa formulazione ha lo scopo di calcolare sia a_{ij} che s_j conoscendo solo x_e

L'ICA può essere utilizzata solo se le componenti rispondono a determinate caratteristiche; infatti le componenti s_j devono essere statisticamente indipendenti, e non-gaussiane. Inoltre è opportuno ma non necessario che il numero delle componenti indipendenti sia uguale al numero delle variabili osservate.

Si tratta di un metodo che utilizza come base di partenza il modello del cranio con le sue relative capacità conduttrici. Il modello adottato per risolvere il problema bioelettrico, è composto da una serie di sfere concentriche alle quali si assegna conduttività e suscettibilità diverse. Si differenziano così le varie parti, dal cervello alla cute. Più è accurata e realistica la riproduzione del modello migliore sarà il risultato di localizzazione, ma i modelli complessi e accurati richiedono una notevole potenza di calcolo, mentre modelli meno sofisticati sono meno accurati anche se permettono di compiere i calcoli più rapidamente. I modelli della sorgente e del volume conduttore, nel caso della massima aderenza all'anatomia, sono ottenuti attraverso i dati della risonanza magnetica. A questo punto ad ogni strato si attribuiscono valori realistici della conduttività elettrica e della permeabilità magnetica. Quindi si effettua il posizionamento dei generatori degli ERP/F (questa operazione viene effettuata su tutte le componenti), ciò permette di localizzare le aree corticali attive e anche come e quando le diverse aree corticali si attivano.

Torniamo quindi al problema inverso di cui abbiamo detto l'ICA essere una delle soluzioni maggiormente adottate; ad esempio nel caso di una tipica distribuzione elettrica prodotta dall'attività di un'area corticale all'interno di un solco, essa può essere descritta da un singolo dipolo tangente, ma anche da due dipoli disposti radialmente e opposti come orientamento. Come viene risolto questo problema? Spesso si cerca di identificare a priori il numero di sorgenti "sospette", per poi procedere. La stima è fondata sul criterio matematico

dei minimi quadrati, che ricava i parametri del dipolo come la spiegazione migliore alle misure ottenute. Vengono stimati i parametri che riducono al minimo la differenza tra le misurazioni effettuate ed il modello. In pratica, si producono le stime dei parametri del dipolo, si risolve il problema diretto e si calcola la differenza tra la distribuzione calcolata e quella reale. La soluzione del problema inverso, nel nostro caso la sorgente del dipolo stimato, sarà data da un numero minimo di parametri.

4.2.3. Epoching

Le successive fasi sono state effettuate su epoche di 3000 ms, -1000ms e +2000ms considerando come 0 la comparsa dello stimolo nel caso degli esperimenti 1 e 2. nel caso del 3° esperimento le epoche sono di un secondo -200ms e +800ms sempre considerando come 0 la comparsa dello stimolo.

4.2.4. Averaging

Come già abbiamo fatto notare, tra le caratteristiche salienti di un segnale di tipo Erp c'è da segnalare che le variazioni in ampiezza prodotte sono estremamente piccole (intorno a 1-3 microvolt) rispetto all'attività di fondo cerebrale. Risulta quindi difficile identificare queste risposte. Ricordiamo che gli ERP vengono definiti come quelle variazioni dei tracciati EEG e MEG che producono sia un allineamento della fase del segnale che un aumento della potenza a differenza delle risposte "indotte" definite come variazioni della potenza media del segnale sempre riconducibili ad evento. Per far "emergere" le risposte evocate dall'attività di fondo si ricorre all'*averaging*. Lo stimolo (nel nostro caso anche task mentali) si ripete numerose volte: il numero di ripetizioni sarà dipendente dall'ampiezza della risposta. I tracciati sono suddivisi in epoche sincronizzate rispetto all'evento, le epoche relative allo stesso evento vengono mediate tra loro cioè l'attività evocata dall'evento si somma algebricamente rispetto all'attività di fondo la quale essendo casuale rispetto all'evento, tende a ridursi o annullarsi quindi: la risposta media sarà l'Erp (fig 55)

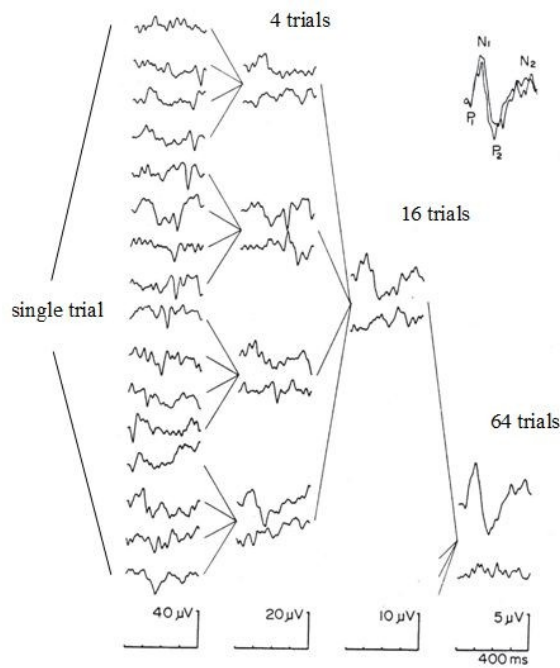


Figura 55: esempio di averaging

Un problema peculiare del segnale sottoposto a *averaging* è il *jittering* (fig. 56).

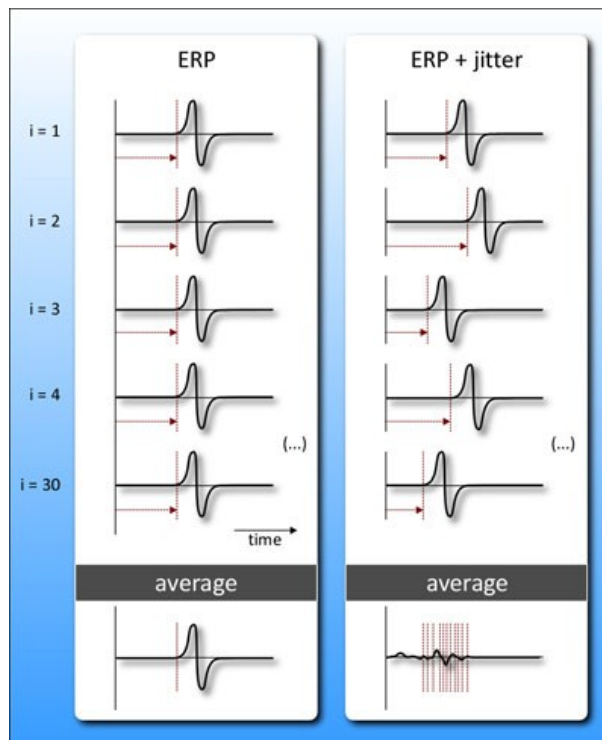


Figura 56: esempio dimostrativo di jittering nell'erp

Infatti se da un lato l'alto numero di ripetizioni diminuisce il rapporto segnale/rumore, dal-

l'altro più è alto il numero di ripetizioni e più è possibile che le risposte non siano allineate in latenza e che la componente risulti appiattita o distorta dal *jittering*.

Per ovviare a questo problemi ultimamente si è pensato di utilizzare un metodo di comparazione che prenda in considerazione singolarmente le ripetizioni. Le singole *trial* vengono esaminate e successivamente sommate una ad una, questo metodo prende appunto il nome di *singles trial*. Per la stima *single-trial* dei potenziali evocati negli ultimi anni sono state proposte numerose tecniche. Elenchiamo in questa sezione alcuni degli approcci più rilevanti per avere un'idea generale sullo stato dell'arte nell'ambito della stima dei potenziali evocati. Le tecniche bayesiane: queste tecniche si basano sulla possibilità di effettuare la stima del vettore dei parametri incognito utilizzando non solo un insieme di dati misurati, ma anche sfruttando tutte le possibili informazioni disponibili a priori sul segnale stesso. Il problema fondamentale è però stabilire se si debba dare più peso al *prior* o al *posterior*. Si ricorre a tal fine ai cosiddetti metodi di regolarizzazione, algoritmi che permettono di stabilire un compromesso ottimale tra *prior* e *posterior* (P.A.Karjalainen, 1997). Karjalainen ha proposto un metodo per la regolarizzazione applicata ai sottospazi per la stima dei PE *single-trial*. Questa tecnica applicata a dati simulati e a dati reali ha portato a risultati positivi. Brainstem Auditory Evoked Potentials (BAEP) è stato introdotto nel 1996 da Nati: si tratta di un sistema fisiologico modellato su un modello lineare e la cui stima è considerata come un problema inverso. L'approccio empirico di Bayes è stato adottato da Friston (Friston et al.,2006) nelle risposte evocate ed indotte sia per dati da magnetoencefalografia che da elettroencefalografia (MEG/EEG). Il metodo è basato sulla stima della componente di covarianza attraverso la *restricted maximum likelihood* (ReML).

Nel 2007 Georgiadis (Georgiadis et al., 2007) ha valutato le informazioni a priori sulle proprietà dei PE. Nelle situazioni di fluttuazione dinamica tra uno stimolo e il successivo, l'informazione a priori può essere utilizzata attraverso metodi di stima come il filtro di Kalman. Fuori dall'ambito bayesiano ricordiamo:l'RBF Neural Network (Merzagora et al.,2004) utilizza reti neurali a tre strati con funzione di attivazione a simmetria radiale per stimare i singoli potenziali. Si tratta di un approccio che permette di ottenere segnali regolari con un tempo di calcolo moderato. L'*iterative independent Component Analysis* di (Zouridakis et al.,2007) è un metodo che processa un data set multivariato permettendo di estrarre anche solo una componente. Questa tecnica sfrutta l'indipendenza statistica rispetto ai segnali misurati che non lo sono. Si procede risalendo alle sorgenti del PE, eliminando il contributo delle sorgenti nel rumore EEG. Altre tecniche che utilizzano i parametri tempo-frequenza come la *Continuous Wavelet Transform* (Mouraux et al.,2008) che permette di

identificare la locazione temporale e l'ampiezza delle componenti principali. L'idea di partenza della WT è analoga quella che sottostà l'analisi in frequenza di Fourier: scomporre il segnale in analisi in termini di funzioni elementari, mentre nell'analisi di Fourier tali funzioni sono rappresentate dalle sinusoidi, nella WT sono delle vere e proprie funzioni "wavelets". Sono di natura oscillatorie e limitate nel tempo, così da permettere la rappresentazione di quei segnali costituiti da componenti oscillatorie variabili per i parametri di ampiezza e/o frequenza. Infatti le "wavelets" si ottengono da una funzione base per traslazione o cambiamento di scala nel tempo. Esistono altri metodi tra cui il filtraggio spazio temporale per la stima single-trial della P300 (Li et al.,2009), il Selective Eigen-Rate Principal Components (SER) di (Andrew et al.,2005) Infine ricordiamo il CSP che vedremo nello specifico qui di seguito.

4.2.5. COMMON SPATIAL PATTERNS

Questa tecnica verrà descritta in maggior dettaglio perchè nell'ambito della nostra ricerca è stata utilizzata per confrontare i nostri dati con quelli di uno studio preso a riferimento: si tratta del lavoro già citato di Da Salla et al dal titolo appunto "*Single-trial classification of vowel speech imagery using common spatial patterns*".

La tecnica CSP (Common Spatial Patterns) permette di creare dei filtri in grado di massimizzare la varianza di segnali in una data condizione riducendo al minimo la varianza di segnali in un'altra condizione. Vedi formula sottostante

$$\bar{C}_g = \frac{1}{n} \sum_{i=1}^n \frac{E_g^i (E_g^i)^T}{\text{trace}(E_g^i (E_g^i)^T)}$$

dove con *trace* si indica la somma di tutti gli elementi presenti sulla diagonale principale della matrice

$$(E_g^i (E_g^i)^T)$$

g rappresenta il tipo di task considerato mentre i indica le epoche;

E_g^i è una matrice in cui le righe rappresentano i canali e le colonne i rispettivi campioni.

\bar{C}_g può essere considerata una matrice di covarianza mediata su n, dove n rappresenta il numero di esempi. Dato che si effettuano prove di confronto tra due singoli task, il risultato dell'operazione precedente sono 2 matrici \bar{C}_1 e \bar{C}_2 la loro somma è la matrice

C_c che può essere anche espressa utilizzando i suoi autovettori:

$$C_c = \bar{C}_1 + \bar{C}_2$$

Si utilizza ora una trasformazione lineare chiamata *whitening transformation*

$$W = \sqrt{\lambda_c^{-1}} V_c^T;$$

che equalizza la varianza nell'autospazio.

La *whitening transformation* è applicata a \bar{C}_g

Viene poi definita la matrice di proiezione P:

$$P = (U^T W)^T$$

La seguente relazione permette l'ottimizzazione per discriminare i 2 *task*.

$$Z_g^i = P E_g^i;$$

4.3 *Pattern recognition*: estrazione delle features

Il *pattern recognition (PR)* viene tradotto spesso come riconoscimento di forme anche se in realtà il termine forma non si sovrappone perfettamente al significato molteplice di *pattern* infatti è preferibile mantenere questo termine non tradotto. Per *pattern* si intende un qualcosa che è riconducibile attraverso il riconoscimento di caratteristiche prototipiche, ad una classe ad esempio i volti, il carattere grafico scritto dalla mano, ma anche un suono o un evento più complesso come i movimenti della crosta terrestre o i segnali elettroencefalografici di cui ci occupiamo. Potremmo quindi dire che si tratta di elementi che si possono classificare ma le cui caratteristiche, cioè quelle prototipiche della classe possono risultare vaghe. Il PR è un'operazione naturale nell'essere umano ma anche negli animali, il problema sorge quando tentiamo di programmare una macchina perché svolga compiti percettivi e cognitivi al fine di riconoscere i *pattern*. Si tratta infatti di un'area specifica dell'apprendimento automatico. Consiste nell'analizzare e identificar dei *pattern* all'interno di un insieme

di dati, queste operazioni conducono quindi ad una classificazione secondo appunto dei valori comuni ai dati. Si procede infatti con una descrizione delle caratteristiche o *features*, in cui si assegna un valore di tipo numerico anche per gli aspetti concettuali della realtà osservata. La seconda fase prevede la delimitazione dello spazio della classe che dovrà essere il più possibile congruente a quello delle caratteristiche.

Infine si sceglierà un algoritmo che esegua in automatico la classificazione.

Qualora il *pattern* fosse specificato in modo rigido *a priori* parleremmo di *pattern matching*; mentre è proprio questa peculiarità di costruire la regola dai dati che fa sì che il PR risulti essere il metodo più studiato e applicato nelle ricerche nei settori più vari dalla psicologia e scienze cognitive all'economia e all'informatica.

DOMINIO	APPLICAZIONE	INPUT	OUTPUT
Biblioteconomia e archivistica	Cataloghi online	Documento di testo	categoria
Analisi documenti	OCR	Immagini di documenti	caratteri
Sistemi biometrici	Riconoscimento identità	Iride, voce, impronta digitale	autorizzazione
BCI	Interazione con la macchina tramite segnali profondi	Segnali eeg, eog, emg	Comunicazione, o movimento tramite periferiche

Tabella 3: ambiti di applicazione del pattern recognition

Nel caso specifico dei segnali elettroencefalografici una volta superata la fase di filtraggio dei segnali registrati si può procedere direttamente ad identificare i *target* e *non-target*.

Questo processo di analisi è il "*feature extraction*". Con per cui si procede con il ricondurre le singole ripetizioni degli stimoli a valori numerici (come abbiamo già detto sopra) quindi a *features*. Le dimensioni ridotte permettono al classificatore di lavorare sui vari domini da quello del tempo a quello della frequenza. Vedremo nei paragrafi seguenti e nel capitolo sul trattamento dei dati quali sono gli algoritmi adottati nel nostro lavoro.

Abbiamo detto che il PR è uno dei sistemi di apprendimento automatico ma in queste pagine riporteremo anche un altro esempio, si tratta del *Machine Learning* si tratta di una macchina che a partire da una stessa sorgente di dati costruisce la sua esperienza e im-

para e migliora le sue prestazioni. Lo strumento comune adottato nel PR e nel ML è la statistica.

4.4. L' SVM

Sotto la dicitura SVM (*Support Vector Machines*), sono raggruppati insieme una serie di metodi di apprendimento supervisionato per la regressione e la classificazione di pattern. Ma che cos'è l'apprendimento supervisionato?

Per apprendimento supervisionato intendiamo una caratteristica tecnica di apprendimento automatico che ha lo scopo di istruire un sistema informatico, quindi una macchina, affinché possa eseguire dei compiti in automatico. Per realizzare un SVM è necessario seguire una procedura ben definita che prevede dei passaggi sequenziali vediamo in ordine gli *step*. Dato un insieme di dati in ingresso, questi dati sono definiti insieme I .

Dato l'insieme dei dati in uscita definito come insieme O , che sta per output, per i quali possiamo avere valori continui, oppure una etichetta numerica.

Si definisce una funzione h la quale associa ad ogni dato in ingresso (I) la sua risposta corretta (O). Come è chiaro dalla forma, si tratta di un presupposto comune a tutti gli algoritmi di apprendimento supervisionato da cui deriva che con un numero adeguato di esempi, l'algoritmo sarà in grado di creare una funzione h_1 tale che questa ultima funzione sarà in grado di approssimare la funzione h . E' appunto utilizzando questo procedimento che la macchina viene addestrata a risolvere il problema questo *step* è definito fase di **Train**.

La fase di **Test** prevede che alla funzione h_1 siano proposti dei dati mai analizzati, se la vicinanza tra le risposte di h_1 e h sarà adeguata avremo quindi dei risultati simili e potremo procedere con la fase di utilizzo. Le macchine con SVM sono spesso utilizzate come tecnica alternativa e contrapposta alle tecniche classiche di addestramento con reti neurali. Infatti molti degli algoritmi d'addestramento di reti neurali lavorano linearmente, hanno come presupposto la similarità tra input ed output. Nella realtà dei dati empirici soprattutto di natura biologica quindi neurofisiologica questo non si verifica. E' facile comprendere come le reti neurali quindi l'elaborazione attraverso questa tipologia di algoritmi dipenda strettamente dai dati in ingresso, un numero ridotto di dati implica una scarsa esperienza della macchina, nel caso in cui i dati siano numerosi ciò lo renderebbe lento perché avremo un algoritmo molto più esteso e complicato. Inoltre questi algoritmi sono molto sensibili alla variazione quindi potrebbe discriminare in maniera non corretta i dati dando risultati falsati. La tecnica di addestramento dell'SVM ottempera ad entrambe le richieste, poiché si tratta di un algoritmo tale da poter rappresentare proprio quelle funzioni non lineari complesse

come i nostri dati. I parametri propri di questa rete derivano dalla soluzione di un problema detto di programmazione quadratica convessa che prevede un unico minimo globale con vincoli di uguaglianza o di tipo box. Il parametro ha un vincolo per cui deve essere mantenuto all'interno di un intervallo. Le SVM appartengono alla famiglia di classificatori con discriminativo, esse cercano di modellare in maniera diretta la relazione che intercorre tra dati in entrata e quelli in uscita, in questo modo si tenta di minimizzare la *loss function* cioè una funzione di perdita. Il procedimento prevede che le coordinate degli oggetti da classificare vengano mappate in uno spazio detto *feature space* utilizzando funzioni non lineari, chiamate *feature function* (Φ). Il *feature space* è uno spazio tridimensionale in cui le due classi possono essere separate con un classificatore lineare (nel video è rappresentato dal piano, anche in fig. 57). Cioè si passa da uno spazio bidimensionale ad uno tridimensionale per separare i dati e una volta separati con l'algoritmo classificatore scelto, lo spazio diviene nuovamente bidimensionale (cfr. <http://www.youtube.com/watch?v=3liCbRZPrZA>). La funzione Φ combina quindi lo spazio input contenente le caratteristiche originali degli oggetti nello spazio delle *features* che potrebbe anche essere infinito. Poiché siamo in uno spazio multidimensionale per trovare l'iperpiano di separazione vengono usate delle funzioni dette **kernel** quindi identificheremo la funzione Φ a partire da una combinazione di funzioni di *kernel*. Con funzione *kernel* e una funzione si intende una funzione corrispondente ad un prodotto scalare in uno spazio esteso.

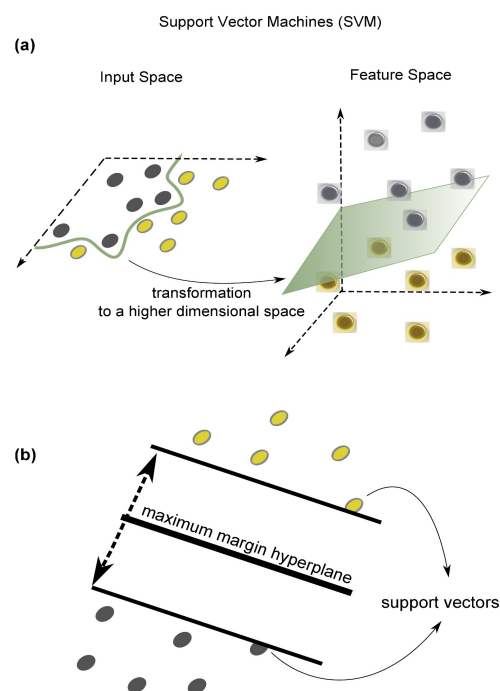


Figura 57 : procedimento di mappatura nello spazio.

La scelta deve essere adeguata la problema infatti un numero eccessivo di parametri per il modello statistico di campione creerebbe sì un modello perfetto per la *training set*, ma con l'introduzione dei nuovi dati incorreremmo nell'*overfitting* (eccessivo adattamento).

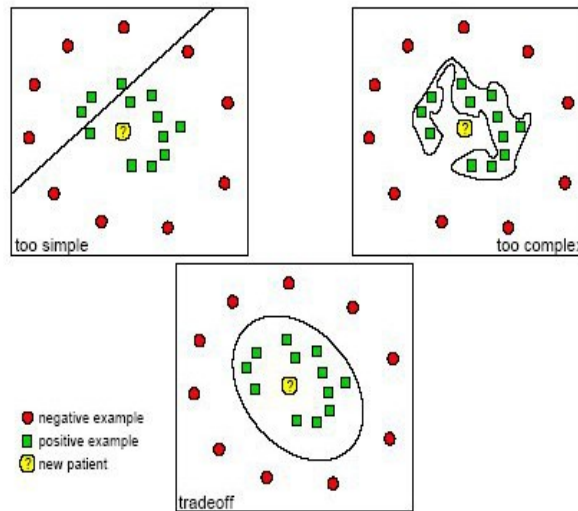


Figura 59: sono rappresentati possibili errori di classificazione dall'inclusione di dati errati all'esclusione di dati corretti

L'errore di validazione aumenta mentre sui dati di training diminuisce. A questo punto si possono differenziare due tipi di casi in base alla separabilità dei *dataset*. Avremo una **SVM lineare** (fig. 59) nel caso in cui i *dataset* siano separabili con un leggero rumore, quindi è posto il margine m che fornisce il *discrimen* tra i dati, in questo caso le cose funzionano come una separazione lineare binaria in cui il margine dell'iperpiano di separazione funge da distanza minima fra le due classi.

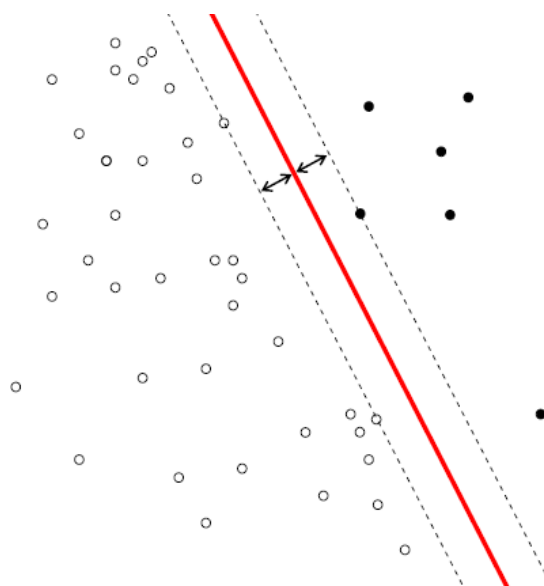


Figura 59: esempio di separazione lineare con SVM

Nel caso i dataset siano separabili con un rumore tale da risultare eccessivo, potremo proiettare il nostro problema in uno spazio di dimensioni maggiori cioè effettuare una proiezione in uno spazio tale per cui i dati risultino separabili.

Lo spazio di dimensioni maggiori, come abbiamo già detto, causa rallentamenti nel calcolo infatti l'algoritmo di apprendimento dovrà lavorare con vettori di grandi dimensioni. Le funzioni di *kernel* specifiche per la soluzione di questi problemi permettono di costruire un classificatore lineare partendo uno **non-lineare**. Per chiarezza rimandiamo alla figura sottostante.

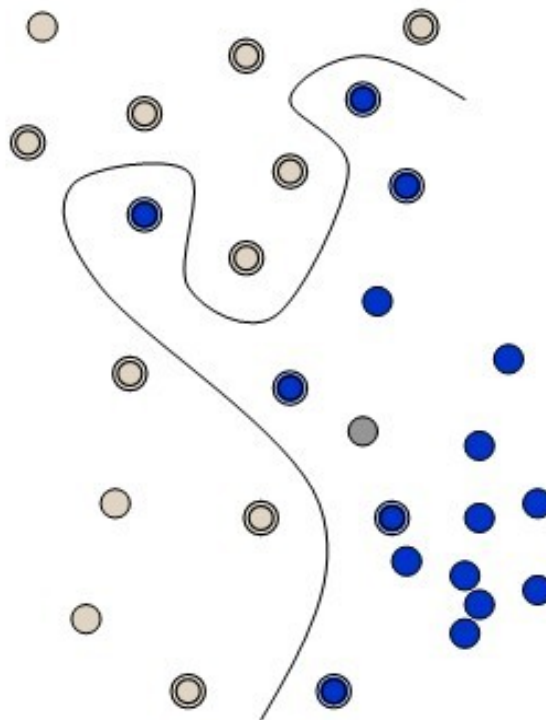


Figura 60: esempio di separazione non- lineare con SVM.

5. DAI DATI AI RISULTATI

5.1. Introduzione

Nei precedenti capitoli abbiamo illustrato gli esperimenti effettuati e le modalità utilizzate per estrarre i segnali del rumore. Nel presente capitolo verrà affrontato il problema della classificazione ovvero del riconoscimento della vocale pensata o articolata senza emettere suono, sulla base dei segnali EEG. Nel primo esperimento sono state utilizzate diverse modalità di classificazione, partendo dalla semplice discriminazione (es. [a] vs [i]) per giungere alla classificazione dell'elemento indipendentemente dalle possibili alternative ad esempio [a] vs. i restanti 3 [i,u,+]. Nel paragrafo 5.2. illustreremo le prime fasi di preparazione sul primo *corpus* che hanno visto l'utilizzo del parametro frequenza come principale fattore di discriminazione e il successivo utilizzo della media mobile a 3. Successivamente presenteremo l'utilizzo della media mobile a 9, quindi nel 5.3. mostreremo l'introduzione del metodo CSP che utilizza il parametro tempo e che ci ha permesso di confrontare i nostri dati con quelli della letteratura di riferimento. Infine nel paragrafo 5.4. presenteremo un'ulteriore elaborazione dei dati nel dominio del tempo utilizzando due tipi di algoritmi: il primo detto misto prevede l'uso di un vettore derivato dalla media di dati e il secondo detto singolo poiché prevede un vettore da dato singolo. Quindi seguono i risultati finali ottenuti con questi due metodi per ciascuno dei 3 esperimenti.

Nel paragrafo 5.5. abbiamo aggiunto i risultati relativi al test effettuato sul secondo esperimento sulle risposte dei soggetti alla prima fase di percezione - elaborazione dello stimolo audio. Tutti codici degli algoritmi e le relative matrici e funzioni sono riportate nell'appendice.

5.2. La prima fase: elaborazione nel dominio della frequenza (primo esperimento)

I dati prodotti dalle acquisizioni del primo esperimento sono stati divisi in insiemi di dati, uno per il caso articolato e l'altro per quello non articolato, ognuno composto da 36 file Matlab relativi ai 4 diversi compiti. In questo modo ad esempio, il file 1_2.m rappresenta l'acquisizione relativa al soggetto 1 in condizione di compito mentale relativo alla [a]; allo stesso modo il file 4_4.m si riferisce al soggetto 4 che pensa alla [u].

Ognuno di questi file è organizzato sotto forma di matrice di dimensione 750 x 50; le 750 righe sono riconducibili al campionamento dei 3 s che compongono ogni prova per la frequenza di campionamento di 250 Hz; le 50 colonne sono invece relative alle 50 diverse prove eseguite dal soggetto in questione.

Per il riconoscimento si è utilizzato l'approccio delle SVM (Support Vector Machine).

Al fine di creare un set (insieme coerente) di dati necessario all'addestramento del SVM

(fase di train) e uno necessario alla verifica della sua funzionalità (fase di test) si è suddiviso ognuno dei 4 canali (Ch1÷ Ch4) che compongono l'insieme di dati di un soggetto, in due canali distinti: **Chtrain** composto dalle prime 25 prove e **Chtest** composto dalle restanti 25 prove. Al fine di ottenere dei risultati consistenti con quelli già presenti in letteratura, le acquisizioni contenute nell'insieme di dati sono state sottoposte ad una operazione di media matematica. Data l'esiguità dei dati in possesso, si è deciso di adottare la media mobile di tipo 3 a 3 sui campioni per evitare che il numero di esempi diminuisca e attenuare inoltre eventuali esempi spuri che possano assumere una direzione nettamente diversa dalla direzione seguita dalla grande maggioranza degli esempi. Al fine di mantenere esempi sufficienti è stato scelto di operare la media mobile 3 a 3. Il procedimento seguito è riassunto nella Tabella 1 prendendo come esempio il soggetto 1 nell'articolato (cfr. app.).

Come risultato di questa operazione si ottengono delle sotto-matrici (indicate con w_1 , a_1 , i_1 ed u_1 per i dati di train e w_{1t} , a_{1t} , i_{1t} ed u_{1t} per i dati riferiti al test) di dimensioni 750×23 . Per ogni soggetto quindi si deve ripetere lo script 8 volte al fine di ottenere le 4 sotto-matrici per l'addestramento e le 4 per il test (prova).

I segnali EEG presentano andamenti nel tempo di difficile caratterizzazione. Al fine di poter impiegare dei modelli più semplici che meglio si prestino ad una elaborazione mediante SVM, si è passati dal dominio del tempo a quello della frequenza utilizzando un apposito codice Matlab (LF.m), che accetta in ingresso una matrice 750×23 (es. w_1) e ne restituisce una di dimensioni 12×23 . Le operazioni del passaggio in frequenza sono riassunte nella Tabella 2 in appendice. Successivamente si è proceduto alla costruzione delle matrici per l'addestramento e la verifica. Con le sotto-matrici nel dominio della frequenza ottenute mediante l'uso dello **script LF.m**, si procede alla creazione delle matrici necessarie all'addestramento e al test dell'SVM che dovrà poi fornire i risultati circa l'accuratezza nel riconoscimento dei 4 compiti mentali. Allo scopo si sono realizzate, per nostra comodità, 8 sotto-matrici ausiliarie (nulla, a, i, u, nullat, at, it, ut), che poi vengono unite per formare le matrici `in_train` e `in_val` necessarie all'SVM. Per la creazione delle matrici `out_train` e `out_val`, si è fatto ricorso ad un apposito *script*. Anche in questo caso si fa ricorso all'uso di apposite sotto-matrici, una per ogni compito. La matrice `out_val` è ottenuta in modo analogo alla precedente. Da notare come le 4 matrici ottenute (`in_train`, `in_val`, `out_train`, `out_val`) siano poi trasposte; questo è da ricondursi alla modalità di funzionamento dell'SVM utilizzata in questo lavoro. Le matrici `out_train` e `out_val` sono ottenute associando una riga composta da 23 simboli "1" alla relativa classe (nulla, a, i ed u), mentre le restanti 3 righe sono composte da 23 simboli "-1".

Le 4 matrici, in_train e in_val di dimensioni 92x48, e out_train e out_val di dimensioni 92x4 devono essere elaborate dall'SVM per la classificazione. Come anticipato, per la classificazione dei 4 diversi compiti mentali si utilizza un classificatore binario SVM (Support Vector Machine). Inizialmente tale SVM è configurato in modalità lineare. Lo script che implementa tale SVM ha il compito di elaborare le matrici in_train, in_val, out_train e out_val fornendo in uscita il valore dell'accuratezza per ognuno dei 4 compiti (null, vocale a, i ed u). A titolo di esempio si riportano i grafici, uno per il caso articolato e l'altro per il caso non articolato, relativi al soggetto 1.

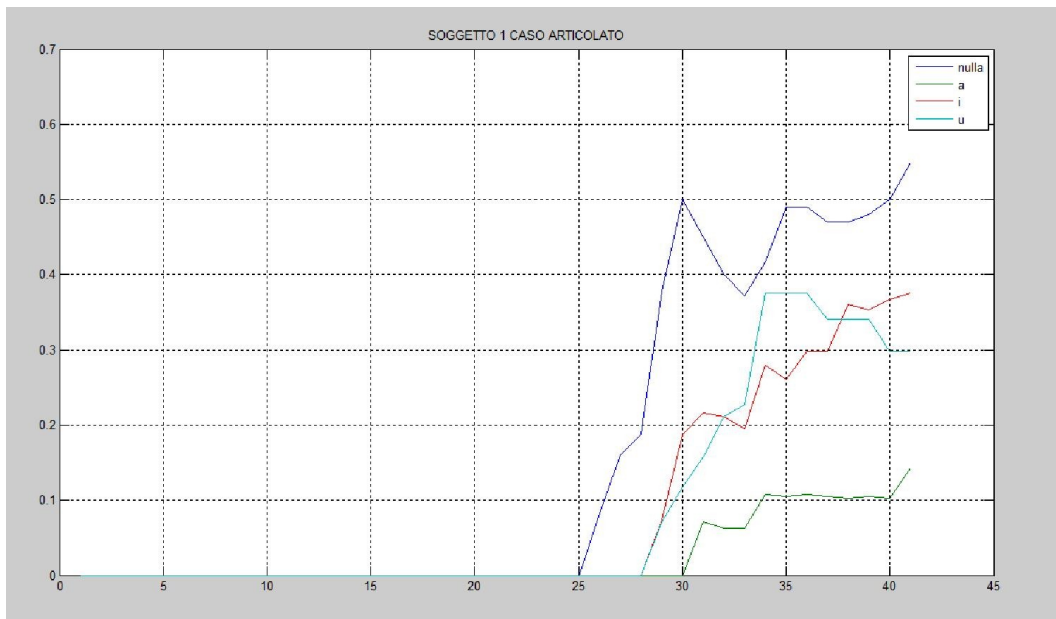


Figura 61: risultati relativi al soggetto 1, caso articolato – SVM lineare

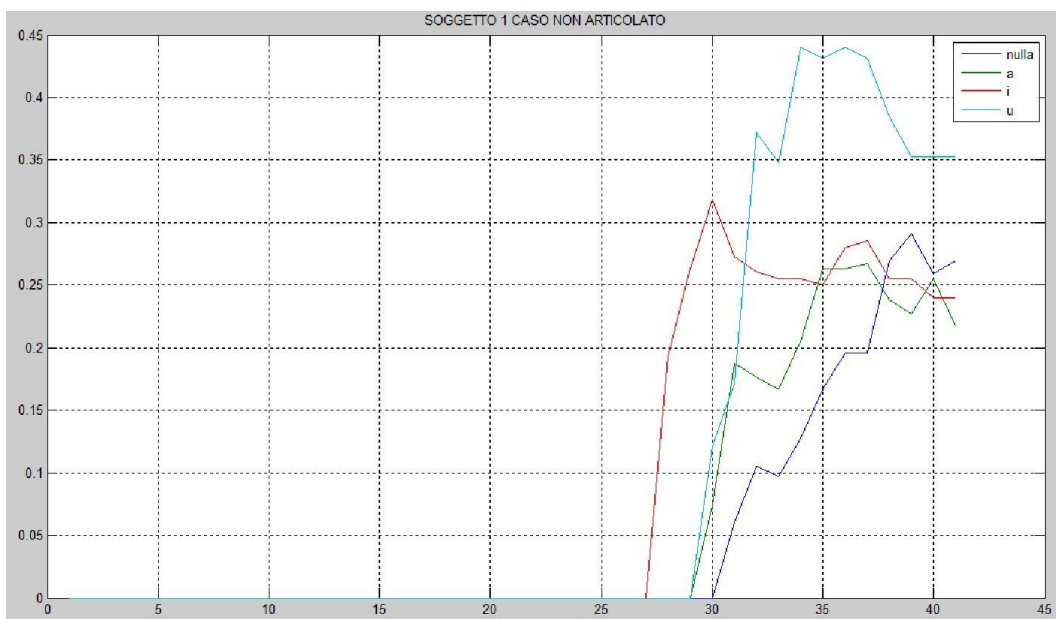


Figura 62: risultati relativi al soggetto 1, caso non articolato - SVM lineare.

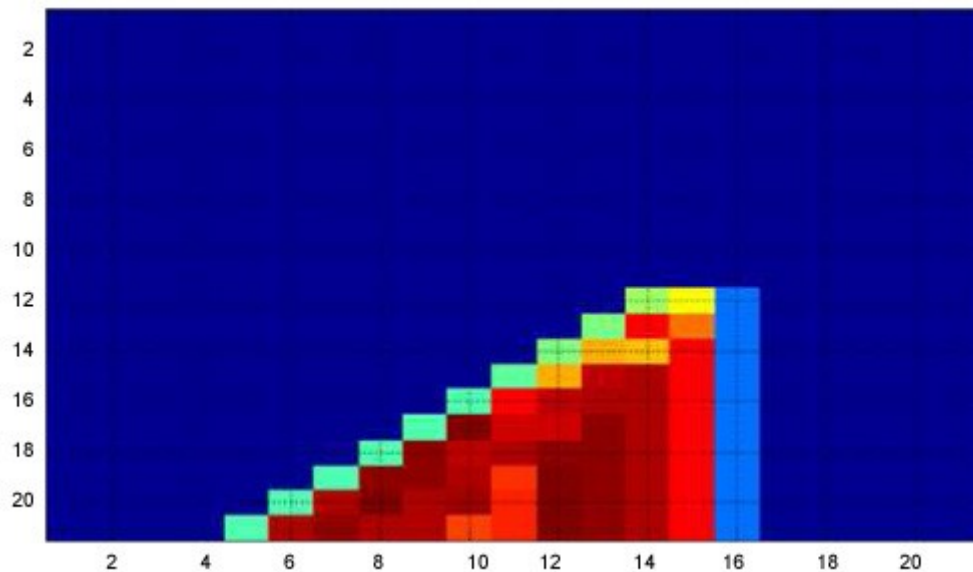


Figura 63: risultato relativo al compito della [a], soggetto 1, caso non articolato - SVM RBF

Al fine di ottenere una migliore accuratezza dei risultati, si è provveduto all'uso dell'SVM con kernel non lineare di tipo RBF (Radial Basis Function). Anche in questo caso, a titolo di esempio si mostra il risultato relativo al compito "A" del soggetto 1 caso non articolato. Poiché la media a 3 non ha dato i risultati sperati, per migliorarne il riconoscimento si è scelto di procedere con un nuovo approccio basato su una media mobile effettuata su 9 campioni per volta. Si cerca di investigare se con questa scelta volta a ridurre il rumore di fondo si ottenga un andamento più simile tra le varie prove e quindi si possa migliorare la classificazione. La costruzione delle sotto-matrici per il train e il test (addestramento e prova), analoga a quanto fatto per il caso di media mobile a 3 campioni. I risultati tuttavia si sono rivelati insoddisfacenti. Si è perciò deciso di inserire una maggiore specificità nella finestra di analisi introducendo il codice denominato P300. Con questo codice si effettua la media mobile su 3 elementi relativamente alle acquisizioni giudicate migliori. Nelle elaborazioni precedenti si effettuava la FFT prendendo in considerazione tutto il tempo, compreso l'intervallo in cui non veniva inviato alcun segnale visivo all'utente. Tale parte d'attesa quindi viene intesa come rumore, per eliminarlo, in questa elaborazione si utilizza una finestra di analisi appunto che contenga la P300 (da 0ms a 600ms). Per determinare l'intervallo dei campioni relativo alla P300 si deve impostare e risolvere la seguente proporzione: fissato un offset 250:600 ms \Rightarrow 1000:150=600:X \Rightarrow X=401 allora l'intervallo è il seguente: 251:401ms

I risultati relativi alle elaborazioni utilizzando l'analisi in frequenza (FFT) sono riportate nelle tabelle 9, 10 e 11 per il confronto CSP-FFT.

5.3. Un'ulteriore metodo di classificazione: *Common Spatial Patterns* (primo esperimento)

Anche al fine di poter effettuare un confronto dei risultati raggiunti con quelli disponibili in letteratura, quelli ottenuti da Da Salla (2009), si è ritenuto opportuno analizzare i dati relativi al primo esperimento, con la classificazione effettuata utilizzando il metodo CSP (Common Spatial Patterns).

La trattazione matematica seguita in questo lavoro fa riferimento al già citato articolo di Da Salla in cui l'intervallo di tempo corrispondente alla comparsa della P300 è di 500ms. L'analisi seguente è incentrata su 4 soggetti, nello specifico soggetto 1 e 5 caso articolato e 3 e 7 caso non articolato.

Questa scelta è dettata dal fatto che, ad una prima analisi, si poteva distinguere facilmente l'andamento del compito. Il metodo CSP è stato implementato con la funzione *salla.m*, appositamente creata. La classificazione viene effettuata confrontando i vari target due a due, ricorrendo così all'uso di una singola SVM. Tale funzione restituisce due matrici che, opportunamente manipolate, sono utilizzate dall'SVM per la classificazione dei target mentali. La classificazione è ottenuta confrontando due compiti alla volta.

La funzione *salla.m* restituisce due matrici; prendendo come esempio il confronto tra i compito [a] ed [i], le matrici prodotte sono E_a (dimensioni 504x50) ed E_i (dimensioni 504x50). Il fatto che la comparsa della P300 sia da individuarsi in un intervallo di 500ms ha fatto sì che i campioni nel tempo da analizzare fossero compresi nell'intervallo 251÷376 (campioni). Fatto questo si procede alla creazione delle matrici di addestramento e test che verranno utilizzate dall'SVM; per la matrice di addestramento vengono presi i primi 25 campioni della matrice E_a e i primi 25 campioni della E_i che sono poi unite in un'unica matrice di dimensioni 504x50. Per costruire la matrice di test si prendono i restanti 25 campioni delle matrici E_a ed E_i . Le matrici *out_val* e *out_train*, necessarie all'SVM, sono composte da un unico vettore di dimensioni 1x50 in cui i primi 25 elementi sono tutti "1" e i restanti 25 sono tutti "-1".

Per una migliore valutazione si effettua un ulteriore confronto apportando una modifica al codice *salla.m* denominandolo *salla1.m* il quale prevede un particolare ordinamento dei vettori. L'elaborazione dei dati porta, relativamente all'accuratezza nel riconoscimento, ai seguenti risultati:

	SOGGETTO	/a/:/i/	/a/:/u/	/a/:/nulla/	/i/:/u/	/i/:/nulla/	/u/:/nulla/
Salla.m	1 Art.	0.6222	0.6522	0.7241	0.6667	0.7037	0.6429
	5 Art.	0.7241	0.6522	0.7241	0.5417	0.6122	0.4255
	3 Non Art.	0.5556	0.4400	0.4706	0.6429	0.7170	0.5217
	7 Non Art.	0.3500	0.46891	0.5778	0.6154	0.6296	0.6809

	SOGGETTO	/a/:/i/	/a/:/u/	/a/:/nulla/	/i/:/u/	/i/:/nulla/	/u/:/nulla/
Salla1.m	1 Art.	0.5333	0.6522	0.6786	0.7059	0.7037	0.6667
	5 Art.	0.6538	0.6522	0.6909	0.5652	0.6250	0.4091
	3 Non Art.	0.5660	0.4400	0.5000	0.64815	0.6545	0.5106
	7 Non Art.	0.333	0.4615	0.5238	0.6667	0.6534	0.7308

Tabelle 4 - 5: confronti risultati Salla.m – Salla1.m

La serie di prove e quindi l'attenta valutazione ha evidenziato che la modificata al codice non ha prodotto risultati significativi.

Tutte le prove effettuate hanno visto la suddivisione degli insiemi di dati per l' addestramento e il test dell'SVM la prima metà dei valori per l' addestramento e la seconda per i test dell'SVM.

Al fine di poter valutare quanto questa scelta abbia o meno influenzato i risultati fin qui proposti si è adottata una nuova strategia selezionando i valori dell' insieme di dati in modo del tutto casuale dove ogni vettore rappresentativo dei valori del insieme di dati viene preso una sola volta per ogni ciclo. Lo script **SVM random** costruisce una matrice formata dai vettori-colonna del insieme di dati scelti in modo casuale. da qui costruisce la matrice di addestramento e di test che verrà elaborata dall'SVM lineare. Per avere una valutazione accurata si è ripetuto tale procedimento per 30 volte.

In questo modo l'SVM lineare elabora le matrici di addestramento e di test per un numero di 30 volte, facendo in modo che ad ogni ciclo le due matrici siano sempre formate da vettori-colonna scelti in modo casuale e quindi sempre diversi per tutte le 30 prove. Il risultato è calcolato come la media di queste 30 prove riportando l' errore relativo commesso.

I risultati descritti rientrano nella casistica di segnali che appartengono al dominio del tempo; come precedentemente accennato i segnali presentano andamenti nel tempo di difficile caratterizzazione. Al fine di poter impiegare dei modelli più semplici, adatti ad una elaborazione mediante SVM, si è proceduto al-passaggio dal dominio del tempo a quello della frequenza. A tal fine il parametro C utilizzato nell'SVM viene opportunamente modificato in base all'utilizzo dei dati nel tempo e in frequenza; nello specifico:

Elaborazione dati in frequenza: C=3:1:6

Elaborazione dati nel tempo: C=3:1:3. Per un confronto più approfondito si è proceduto al confronto mediante SVM_random utilizzando segnali in frequenza; allo scopo si è implementato lo script **LF1.m** anche qui si è considerato che il P300 sia di 500 ms questo ha fatto sì che i campioni da analizzare nel tempo e trasformare in frequenza siano compresi nell'intervallo 251÷376. Al punto 11 in appendice si riporta un esempio di matrice di confronto. In questo modo i 4 canali ch1 che costituiscono l'insieme di dati (dataset) di uno stesso soggetto vengono trasformati in frequenza mediante la funzione appositamente creata LF.m; Una volta ottenuti i 4 canali con i dati in frequenza si va a realizzare la matrice $E_a - E_i$ tali matrici vengono poi elaborate dallo script **SVM_random**. Qui di seguito nelle tabelle 9, 10, 11, riportiamo i risultati relativi ai test con algoritmo CSP che utilizza come parametro principale il tempo e l'algoritmo FFT (par 5.1) che utilizza la frequenza. Si tratta di dati relativi esclusivamente al primo esperimento e ai soggetti che hanno ottenuto i risultati migliori nei test.

compito	A-I[%]	A-U[%]	A-+[%]	I-U[%]	I-+[%]	U-+[%]
CSP	67.71	66.41	75.05	67.58	70.68	64.73
	±5.30	±2.61	±4.06	±1.9	± 1.49	±1.49
FFT	66.90	63.99	66.32	58.53	60.12	63.90
	±3.99	±4.91	±4.24	±7.51	±5.97	± 6.87

Tabella 6: confronto CSP-FFT - soggetto 1, caso articolato

compito	A-I[%]	A-U[%]	A-+[%]	I-U[%]	I-+[%]	U-+[%]
CSP	72.75±	65.65	75.81	57.63	62.12	58.53
	1.04	± 1.58	± 4.30	± 4.98	± 1.99	± 5.91
FFT	60.03	48.38	55.92	60.83	58.45	52.63
	± 7.03	± 9.45	± 6.08	± 5.08	± 7.79	± 7.81

Tabella 7: confronto CSP-FFT - soggetto 5, caso articolato

compito	A-I[%]	A-U[%]	A-+[%]	I-U[%]	I-+[%]	U-+[%]
CSP	55.62	49.14	59.01	64.40	71.98	64.01
	±0.39	±4.86	±6.16	±0.67	±0.87	±4.97
FFT	59.84	52.29	52.76	56.30	59.97	47.36
	± 6.18	± 8.63	± 9.60	± 8.95	± 7.57	± 7.26

Tabella 8: confronto CSP-FFT - soggetto 3, caso non articolato

compito	A-I[%]	A-U[%]	A-+[%]	I-U[%]	I-+[%]	U-+[%]
CSP	48.45 ± 8.44	53.59 ± 5.64	62.80 ± 5.57	63.25 ± 2.72	64.23 ± 2.01	69.35 ± 2.46
FFT	60.07 ± 8.08	65.18 ± 4.83	57.74 ± 8.71	49.30 ± 7.18	61.44 ± 7.65	56.47 ± 8.75

Tabella 9: confronto CSP-FFT - soggetto 7, caso non articolato

5.4. Ulteriori elaborazioni nel dominio del tempo

A fronte dei risultati ottenuti sostituendo al kernel RBF il kernel quadratico con risultati migliori nella classificazione, si è quindi proceduto alla riclassificazione dei dati con il nuovo modello. Avremo due algoritmi per i quali riportiamo la parte centrale a seguire e le relative funzioni in appendice. Il primo algoritmo si distingue dal secondo per l'adozione della media mobile, vedremo nei grafici dei risultati quali differenze otteniamo dal punto di vista dei valori di riconoscimento dei compiti. Il primo algoritmo `mixture.m` utilizza come vettore un insieme di dati mediati (da qui in poi chiamato **a media di dati**), prevede la possibilità di selezionare i canali da rimuovere e che quindi non vengono sottoposti a test come nel nostro caso i canali EOG. E' possibile scegliere l'intervallo temporale prestando attenzione allo 0 inteso in questo caso come l'inizio dell'epoca e non come il momento coincidente l'apparire dello stimolo. Successivamente è possibile selezionare ulteriormente il numero dei canali su cui effettuare la prova. la prova è effettuata su una media mobile effettuata con un addestramento composto dal 25% dei dati e il restante 75% è utilizzato per la prova. La funzione è denominata `mean_mixture` (app.). All'interno è inserita la funzione per il CSP, quindi l'SVM relativa appare sotto la forma `s_v_m.m` (app.).

La seconda elaborazione a dato singolo permette di lavorare anche esso nel dominio del tempo, è possibile selezionare i canali su i quali effettuare la prova inserendo tra parentesi quelli da non sottoporre al calcolo e quelli da testare. E' quindi possibile indicare l'intervallo temporale in cui operare il confronto, ricordiamo che l'unità di misura per il tempo è il ms e come nel `mixture.m` bisogna tenere presente lo slittamento dello 0 come inizio della selezione per cui può coincidere spesso con il -1 dell'epoca originaria.

5.5. I risultati

Riportiamo quindi i grafici con i risultati relativi ai test effettuati su tutti gli esperimenti di cui riportiamo i dati grezzi nelle tabelle in appendice. I test sono stati effettuati sui dati di ciascun esperimento in un primo momento eseguendo il confronto tra due compiti singoli e

nella seconda e ultima fase tra un singolo e i restanti 3.

Ciò è stato dettato dall'esigenza di calare la macchina nella realtà applicativa cioè del riconoscimento di un compito tra molti come può essere appunto il riconoscimento della produzione di un fono rispetto agli altri conosciuti dalla macchina.

Quindi avremo [a] vs.[iu+], [i] vs [au+], [u] vs [ai+] e infine il segnale di controllo [+] vs [aiu] Inoltre come è evidente abbiamo un sostanziale miglioramento nella soluzione del compito infatti mentre nel confronto a due si riscontravano valori di poco superiori al caso (50%), per la prova “uno contro tutti” contrariamente a quanto ci si aspetterebbe, la situazione si rivela più incoraggiante.

5.5.1. Il primo esperimento: sessione 1 e 2

Per il primo esperimento il segnale è stato registrato utilizzando 64 canali ma in fase di analisi sono stati utilizzati solamente i canali: Fz Cz C3 C4 C5 C6, l'intervallo temporale si concentra intorno ai 250 ms dall'apparire dello stimolo. Il protocollo prevedeva la partecipazione di 9 soggetti divisi in 6 uomini e 3 donne. La somministrazione dello stimolo era eseguita attraverso la presentazione sullo schermo del computer di immagini stilizzate di bocche che riproducevano la posizione comune per la produzione delle vocali selezionate [a], [i],[u] ed un compito di riposo come stimolo di controllo [+].Nella prima sessione era richiesto di pensare di produrre la vocale corrispondente senza muovere l'apparato fonatorio, nella seconda era richiesto di produrre la vocale senza però emettere suono

5.5.1.1. Valori di riconoscimento per la macchina con vettore a media di dati

I Valori medi di riconoscimento con i dati mediati sono del 74% per il compito [a], del 77% per il compito [i], per il compito [u] del 75% e per il compito [+] l'80%. Il valore minimo di riconoscimento si ha nel caso “03I vs 03AU+” con il 60,3% (compito [i]) mentre il valore maggiore si raggiunge con “06U vs 06AI+” (compito [u]) con un riconoscimento del 90,6%.Il soggetto 6 ottiene i migliori risultati per il riconoscimento tra i 9 soggetti ma anche i soggetti 1, 4 e 8 ottengono un riconoscimento intorno l'80%. I risultati relativi al test sulla sessione che prevedeva l'articolato senza fonazione sono:per il compito [a] il 78%, per il compito [i] l'81%, per il compito [u] il 77% e per il compito [+] il 79%.

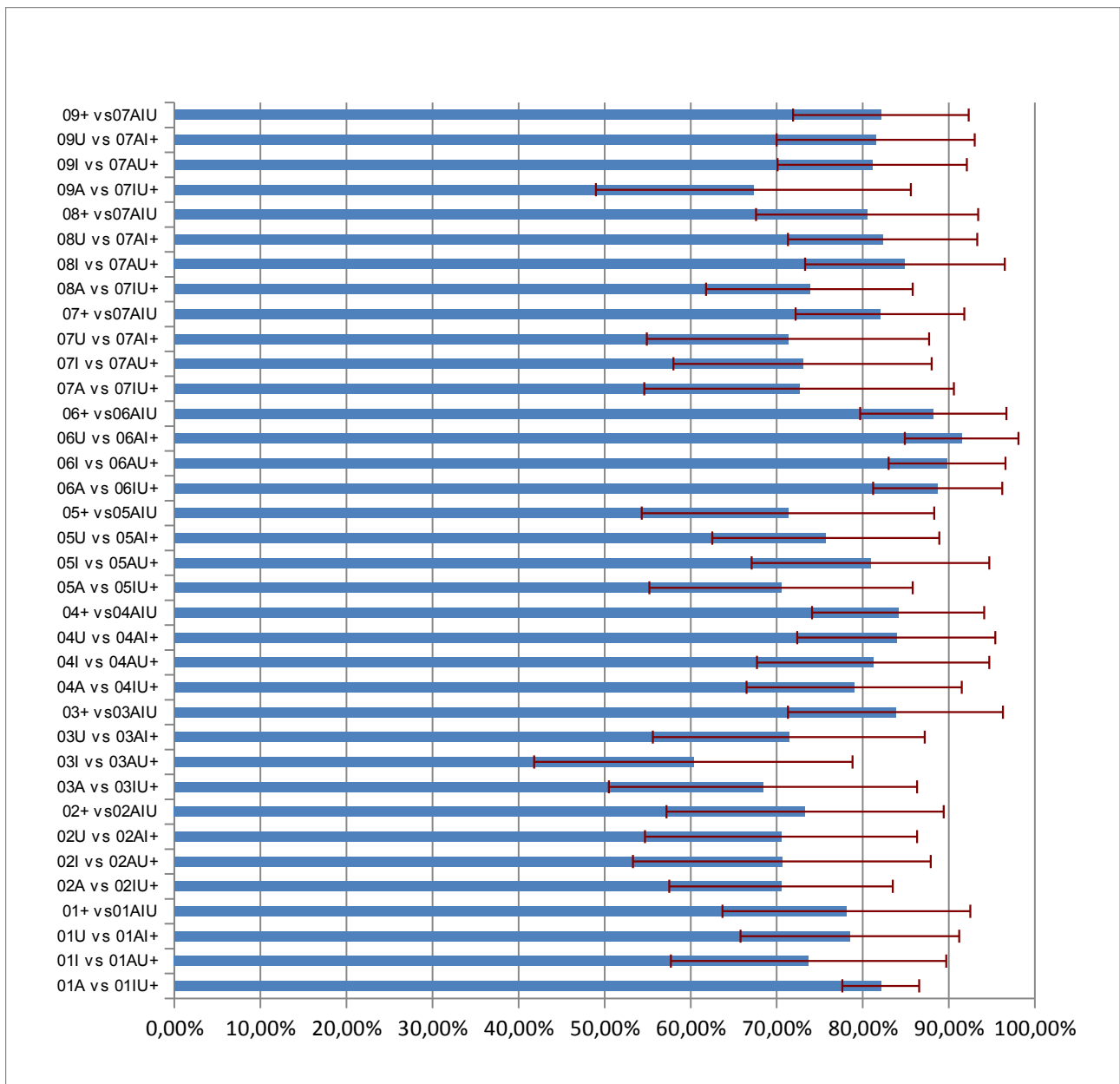


Figura 64: primo esperimento, sessione silente, risultati dei test su tutti i soggetti con metodo a media di dati.

5.5.1.2. Valori di riconoscimento per la macchina con algoritmo a vettore di dati singoli.

Fornisce valori medi di riconoscimento per [a] del 69%, per [i] del 67%, per [u] del 69% e del 67% per il compito [+]; il valore massimo del 89,9% nel soggetto N°6 lo si ottiene per il riconoscimento di [a] mentre il minimo del 57,7% per il riconoscimento di [i] nel soggetto N°3. Nella sessione di articolato senza fonazione i risultati sono: per [a] il 67%, per [i] il 66%, per [u] il 68% e il 67% per il compito [+] e con un massimo per [u] del 68% e un minimo per [i] di 66%.

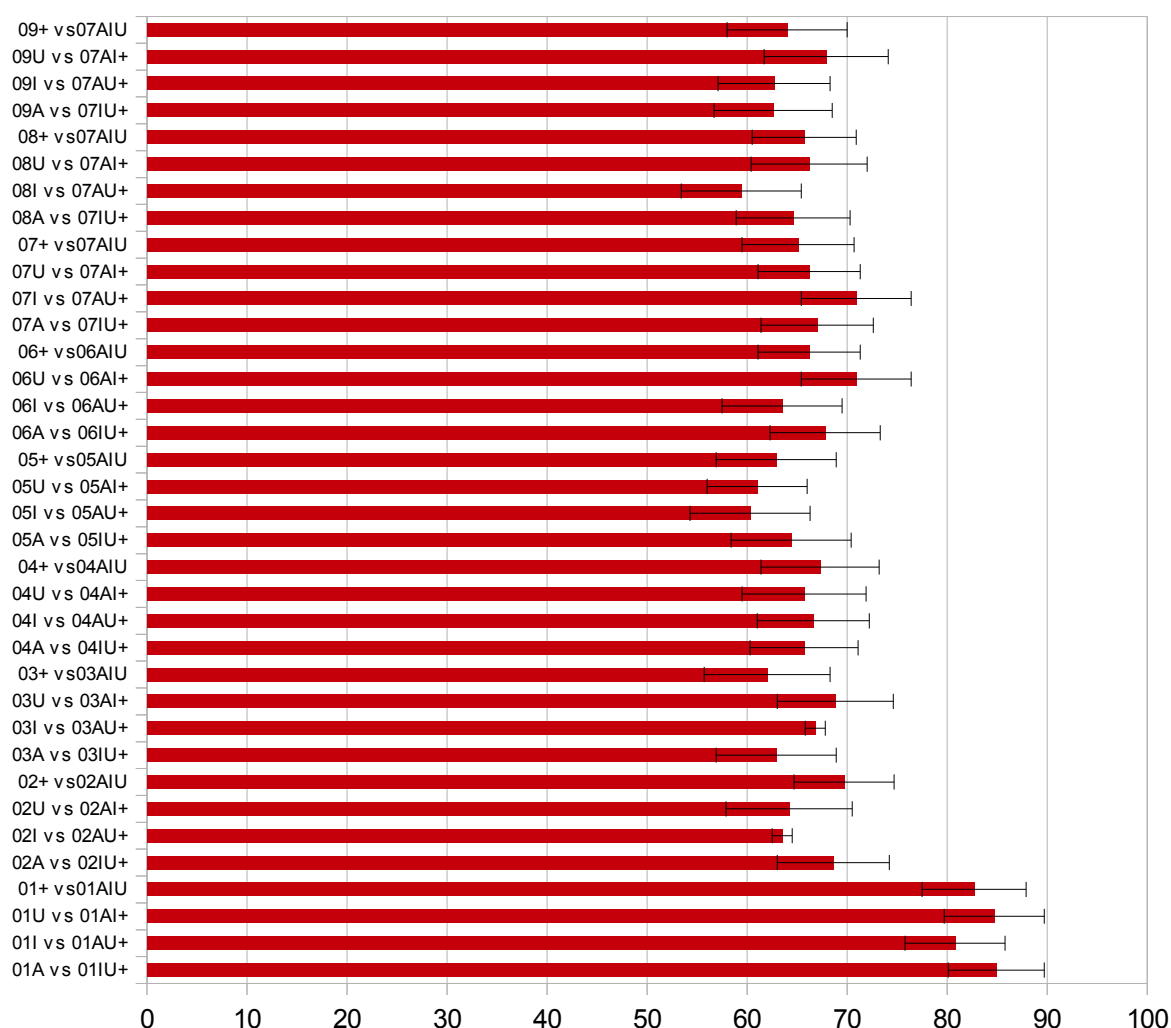


Figura 65: primo esperimento, sessione articolata, risultati dei test su tutti i soggetti con metodo a dati singoli.

5.4.1.3 Riflessioni sui dati del primo esperimento

Una prima impressione ci porta a ritenere il metodo a media di dati più efficace ma se ci soffermiamo ad osservare la sua deviazione standard si aggira intorno al 13% mentre il metodo a dati singoli è sotto il 6%. Inoltre il numero di soggetti porta all'interno dei dati un allargamento della forbice tra risultati minimi e risultati massimi. Nel caso della media di dati, per la prima sessione per [a] si va da un minimo del 67% per il soggetto n°9, ad un massimo del 89% per il soggetto n°6. Per [i] si va da un minimo del 60% per il soggetto n°3, ad un massimo del 89% per il soggetto n°6. Per [u] si va da un minimo del 70% per il soggetto n°2, ad un massimo del 91% per il soggetto n°6.

Nel caso dell'uso del metodo con dati singoli, per la prima sessione per [a] si va da un minimo del 62% per il soggetto n°2, ad un massimo del 89% per il soggetto n°6. Per [i] si va da un minimo del 58% per il soggetto n°3, ad un massimo del 85% per il soggetto n°6. Per [u] si va da un minimo del 64% per il soggetto n°1, ad un massimo del 87% per il soggetto n°6. Anche nella seconda sessione che prevedeva l'articolazione senza fonazione, sono presenti le divergenze tra soggetti.

Abbiamo perciò per il metodo a media di dati: [a] riconosciuto al 67% per il soggetto n°5 e al 91% per il soggetto n°1, [i] riconosciuto al 90% per il soggetto n°1 e al 67% per il n°5 infine [u] riconosciuto al 91% dal soggetto n°1 e al 70% del soggetto n°5.

Così per il metodo con dati singoli: [a] riconosciuto al 85% per il soggetto n°1 e al 63% per il soggetto n°3, [i] riconosciuto al 81% per il soggetto n°1 e al 60% per il n°5 infine [u] riconosciuto al 85% dal soggetto n°1 e al 61% dal soggetto n°5.

Da questi valori sembrerebbe che il numero e la varietà dei soggetti in cui erano comprese 3 donne abbia portato a dei risultati con divergenze anche importanti dal punto di vista del valore percentuale.

5.5.2. Il secondo esperimento: sessione 1 e 2

Per l'analisi dai dati del il secondo esperimento si è utilizzato il CSP-SVM con il kernel quadratico in quanto sembra dare i migliori risultati di classificazione. Il segnale è stato registrato utilizzando 64 canali, in fase di analisi sono stati selezionati i seguenti: Fz, Cz, C3, C4, C5, C6 e gli EOG. L'intervallo temporale è centrato intorno ai 250 ms dallo stimolo. Il protocollo prevedeva la partecipazione di 3 soggetti di sesso maschile. La differenza più importante rispetto al precedente è che la somministrazione dello stimolo era eseguita attraverso il segnale audio relativo alle vocali selezionate [a], [i], [u] ed uno stimolo a 440 Hz come stimolo di controllo [+]. Nella prima sessione era richiesto di pensare di produrre la vocale corrispondente senza muovere l'apparato fonatorio, nella seconda era richiesto di produrre la vocale non emettendo il suono.

Nel secondo esperimento come si può notare dall'esame dei grafici sottostanti, per ambedue le sessione abbiamo un avvicinamento della percentuale di riconoscimento derivante dai due algoritmi. Nel caso del metodo a media di dati abbiamo il 77% per il metodo con dati singoli il 74% ma la distanza tra le relative deviazioni standard si fa maggiore arrivando fino al 15% per il metodo a media di dati e scendendo al 4% per il metodo con dati singoli. Nella sessione che prevedeva lo svolgimento silenzioso, con metodo a media di dati si ottengono i seguenti risultati: per [a] il 79%, per [i] il 78%, per [u] il 77% e per il compito

di controllo il 72%. Con metodo a dati singoli abbiamo: per [a] il 73%, per [i] il 74%, per [u] il 75% e per il compito di controllo il 76%.

Nella sessione che prevedeva lo svolgimento con fonazione, con il metodo a media di dati si ottengono i seguenti risultati: per [a] il 77%, per [i] il 71%, per [u] il 73% e per il compito di controllo il 71%. Con il metodo con dati singoli abbiamo: per [a] il 70%, per [i] il 71%, per [u] il 68% e per il compito di controllo il 79%.

Più nel dettaglio per il silente spicca il secondo soggetto denominato M per i valori di riconoscimento di oltre l'80% con l'uso del metodo con dati singoli, e abbiamo lo stesso soggetto al 94% per il riconoscimento del compito [i]. Per quanto concerne la seconda sessione nella quale è prevista l'articolazione con fonazione abbiamo il 3° soggetto che si stacca dalla media nel metodo a media di dati di 7 punti percentuali e nel metodo con dati singoli del 2,5%.

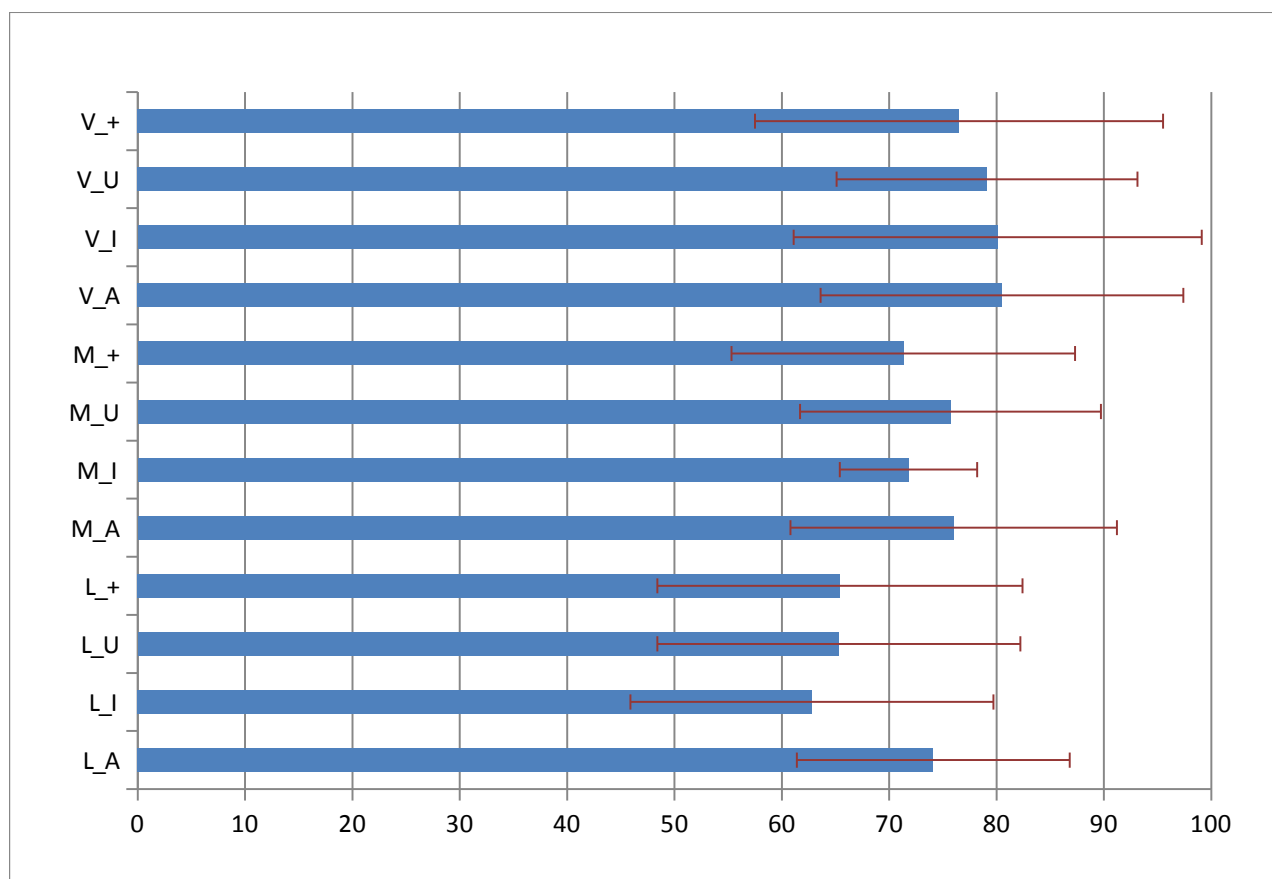


Figura 66: secondo esperimento, sessione articolata, risultati su tutti i soggetti con metodo a media di dati.

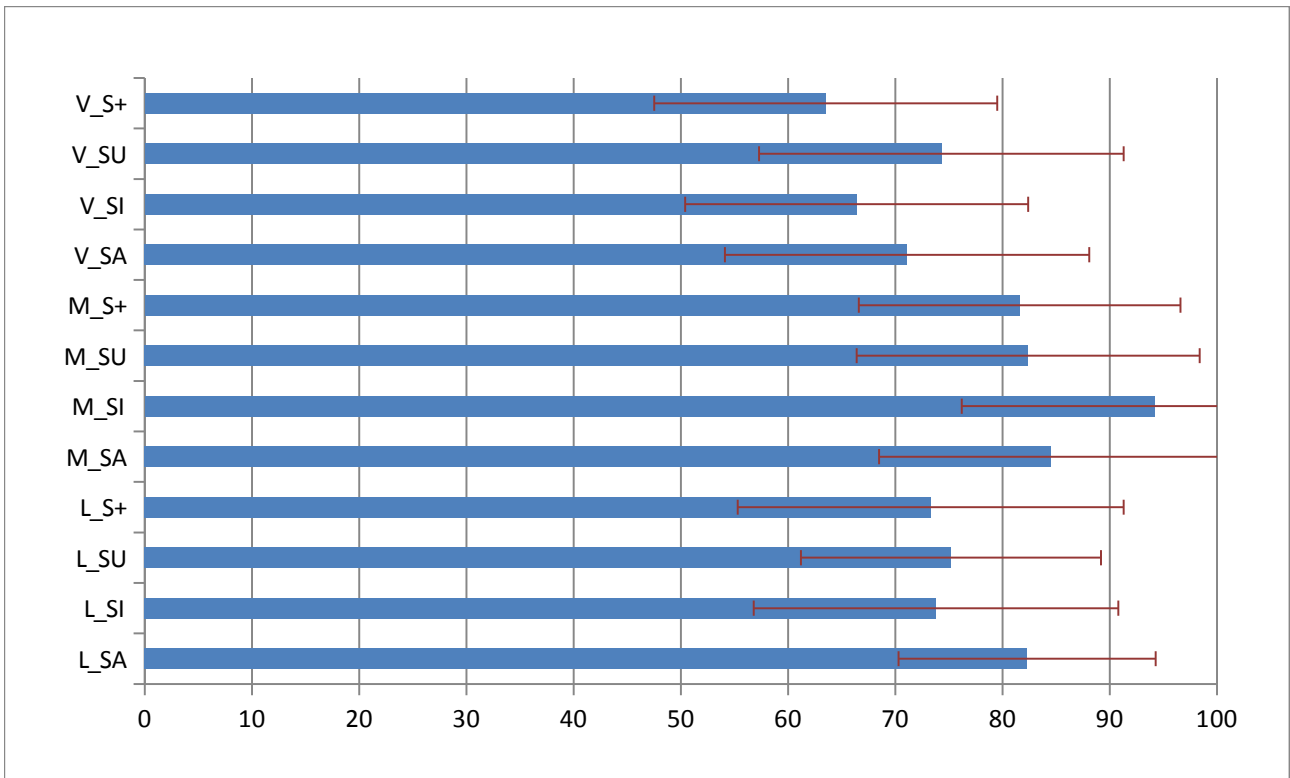


Figura 67: secondo esperimento, sessione silente, risultati su tutti i soggetti con metodo a dati.

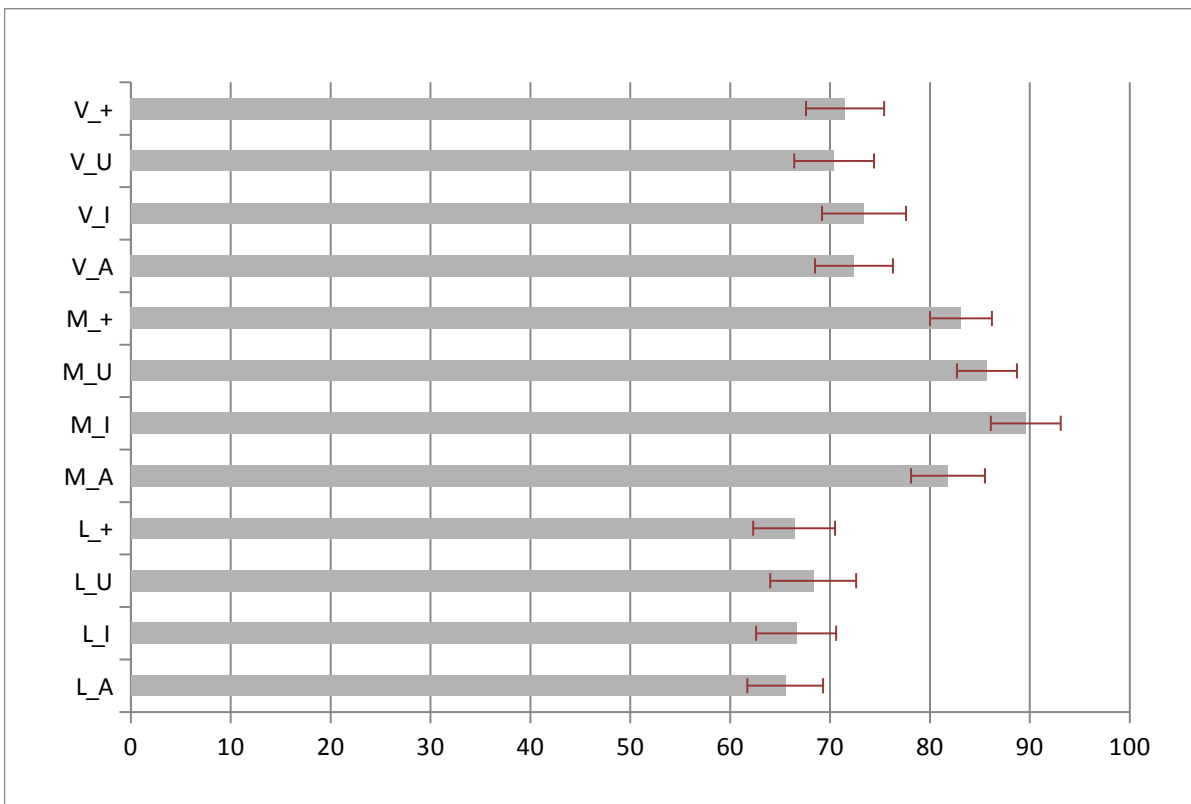


Figura 68: secondo esperimento, sessione articolata, risultati su tutti i soggetti con metodo a media di dati.

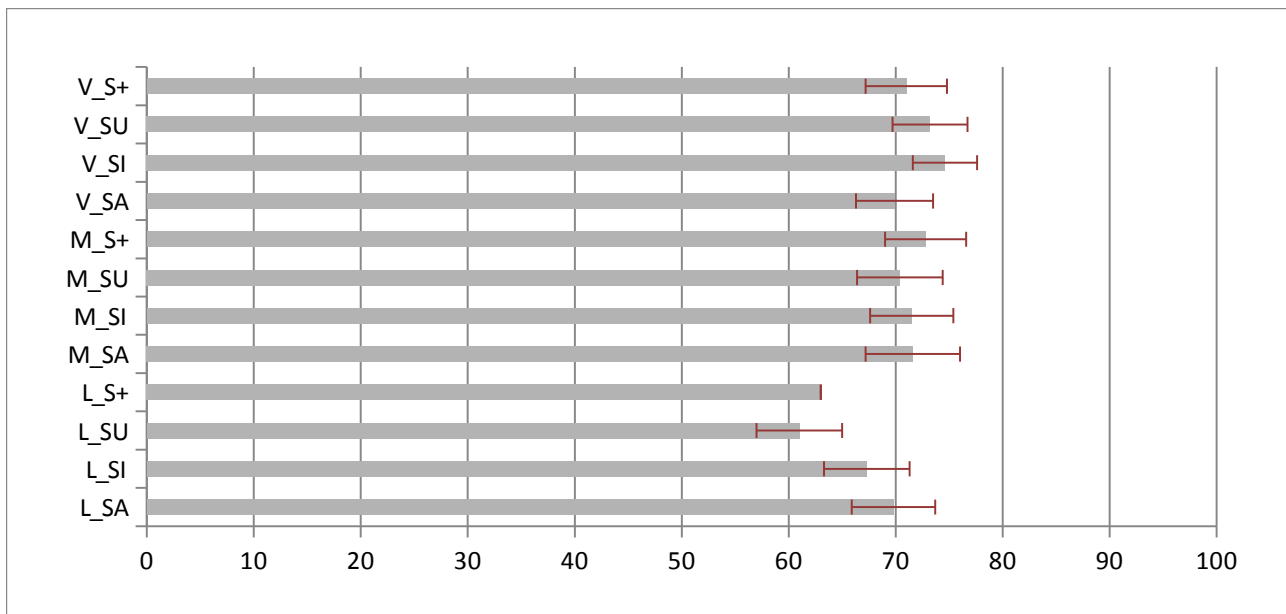


Figura 69: secondo esperimento, sessione articolata, risultati su tutti i soggetti con metodo a dati singoli.

5.5.2.1. Riflessioni sui dati del 2° esperimento

Esaminando tutti i risultati relativi ai singoli 3 soggetti dell'esperimento si nota che le percentuali sono omogenee, notiamo nella massa uniforme due soggetti che si distinguono: il soggetto 6 per il non articolato che ottiene dei risultati attorno al 90% sia con il metodo a media di dati che con il metodo a dati singoli che in questo caso hanno deviazioni standard vicine (7% media di dati e 4,5% dati singoli). Per l'articolato abbiamo il soggetto 1 che per il vettore a media di dati si aggira attorno al 90% di riconoscimento con una deviazione standard di circa il 6% mentre per il metodo a dati singoli l'83% con una deviazione standard del 5%. Più in particolare abbiamo per la prima sessione silente con algoritmo vettore a media di dati: [a] 84% per il soggetto M, [i] 94% per il soggetto M, [u] 82% per il soggetto M. Mentre con algoritmo metodo a dati singoli: [a] 82% per il soggetto M per il soggetto V 71%, [i] 90% per il soggetto M e 76% per il soggetto V, [u] 86% per il soggetto M e 74% per il soggetto V. Per la seconda sessione con vocalizzazione con algoritmo metodo a media di dati: [a] 81% per il soggetto V e 74% per il soggetto L, [i] 81% per il soggetto V e 63% per il soggetto L, [u] 81% per il soggetto V e 65% per il soggetto L. Per la seconda sessione con vocalizzazione con algoritmo metodo a dati singoli: [a] 72% per il soggetto M e 70% sia per L che per V, [i] 70% per M e 75% per V, infine [u] 71% per M e 73% per V. Essendo il numero dei partecipanti ridotto e forse grazie anche alla migliore conoscenza dei compiti, i risultati appaiono più uniformi e quindi di più facile interpretazione.

5.5.3. Il terzo esperimento: sessione unica

Al terzo esperimento ha partecipato un solo soggetto e gli elettrodi utilizzati sono stati: Fz, Cz, C3, C4, C5, C6, più ovviamente gli EOG e le referenze. Lo stimolo in questo protocollo è molto semplificato, si tratta di un flash che si ripete costante a distanza di 1sec all'apparire del quale il soggetto pensa di produrre il compito. Si sono effettuate tre sessioni di 100 ripetizioni in maniera silenziosa per ciascun compito, ma data la natura dello stimolo in questo esperimento non abbiamo il compito [+] e ovviamente non c'è stata possibilità di inserire casualità. Il tempo di analisi è limitato ad una finestra compresa tra -500 e 500 dallo stimolo.

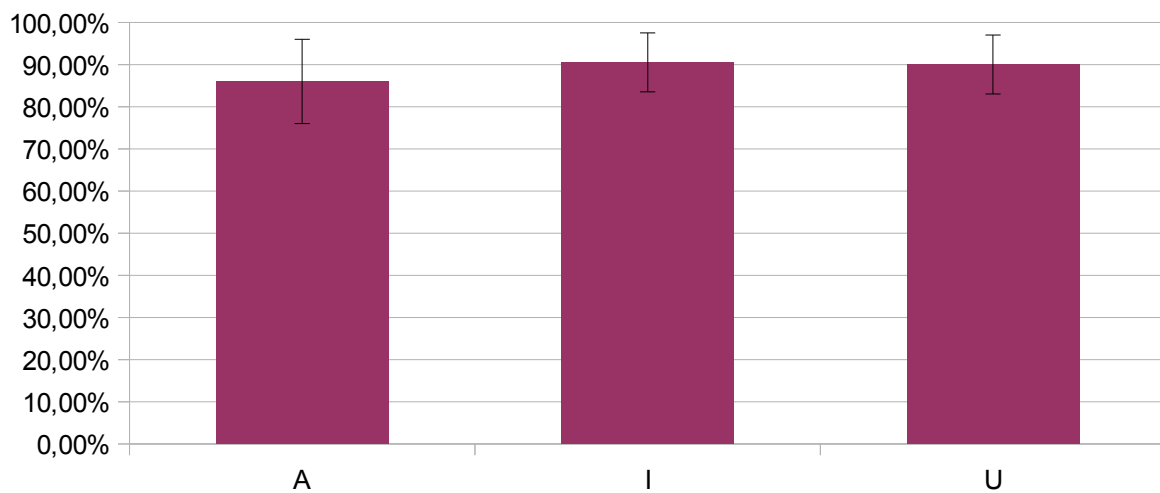


Figura 70: terzo esperimento, sessione unica, risultati con metodo a media di dati.

Come per l'esperimento precedente si è utilizzato il CSP con SVM a kernel quadratico. Si sono ottenuti i risultati migliori sia per il metodo a dati singoli che per il metodo a media di dati con oltre il 90% di riconoscimento per ogni compito. Nel dettaglio abbiamo il metodo a media di dati: per [a] 86%, per [i] 90%, per [u] 90%. Il metodo con dati singoli: per [a] il 95%, per [i] il 95%, per [u] il 96%. Il miglior risultato può essere originato da molteplici cause tra cui l'esperienza e professionalità di chi ha montato la cuffia, la motivazione del soggetto sperimentate e la semplicità del compito, che non richiedeva al soggetto di coinvolgere altre aree per esaminare i compiti. Anche il fatto che non fosse presente alcuna randomizzazione può aver facilitato lo sperimentatore.

5.5.4. Test risposte audio

Infine è stato effettuato uno studio sulle risposte allo stimolo acustico presente nel secondo esperimento. Si è cercato in questo caso di individuare e discriminare il segnale in ingrosso, come per quello in produzione. Per la classificazione del segnale sulla parte audio sono stati selezionati canali diversi della stessa cuffia a 64 elettrodi. Si tratta dei canali corrispondenti a T7, T8, Tp7, TP8, TP9 e Tp10. La latenza temporale è fissata a 50ms dallo stimolo acustico.

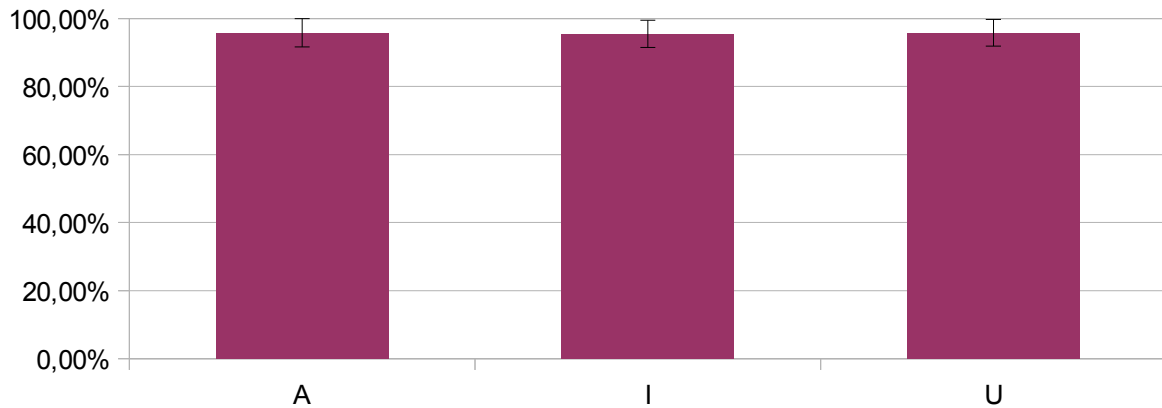


Figura 71: terzo esperimento, sessione unica, risultati con metodo a dati singoli.

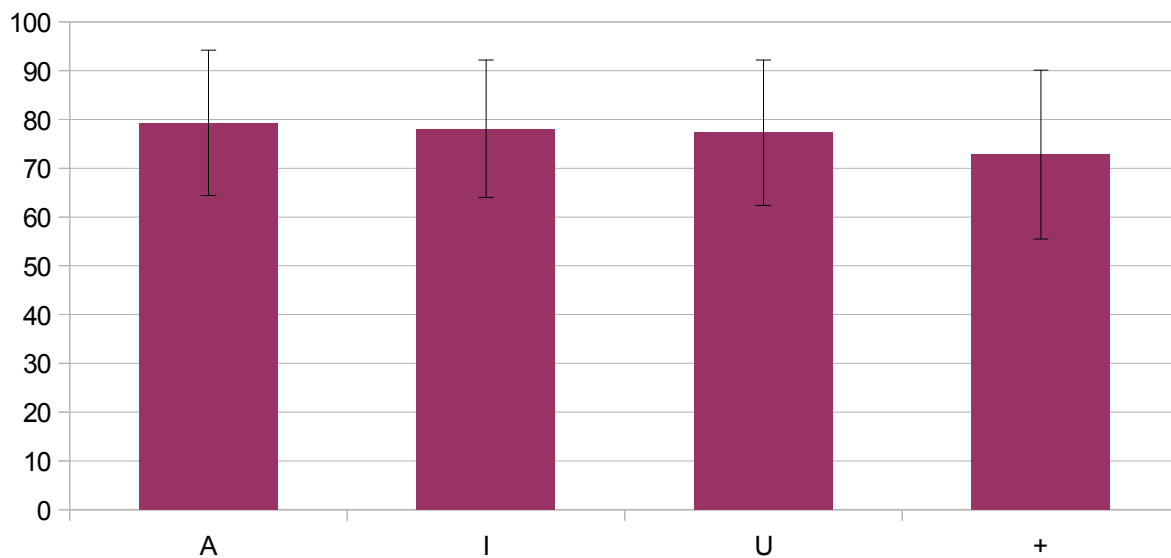


Figura 72: terzo esperimento, sessione unica, risultati con metodo a media di dati.

5.6. Riflessioni conclusive sui risultati

Nell'analisi dei risultati si sono potute osservare importanti variazioni tra soggetti: esse

sono insite nel tipo di sperimentazione, dipendenti oltre che dal numero più o meno ampio di soggetti, anche da fattori quali le caratteristiche biologiche, di stato e di umore che caratterizzano e rendono differente il segnale fra i soggetti ma anche tra un soggetto e se stesso in momenti differenti. Dal confronto effettuato sui risultati ottenuti in ciascuno dei 3 esperimenti risulta evidente come la scelta di un minor numero di essi la scelta di un solo genere sessuale (nel 2° esperimento tutti maschi) così come la minore differenza anagrafica, comporta un'omogeneità nei dati che diventa miglioramento qualitativo nel momento in cui si scelgono non solo un numero limitato di soggetti ma ciascun soggetto effettua la fase di addestramento, come nel caso dell'unico soggetto del 3° esperimento. I risultati derivanti dalle prime fasi di test riportate nei paragrafi 5.1 e 5.2 sono molto al di sotto dei valori successivamente ottenuti con i test riportati nel paragrafo 5.3, la principale differenza non sta nei codici che nel caso di quelli utilizzati per l'analisi nel tempo sono già in forma definitiva nel paragrafo relativo (5.2), ma bensì nella scelta di testare un compito contro i restanti compiti (es. [a] vs [i] [u] [+]). Infatti anche le forme finali dei test con dati multipli e con dati singoli sono state utilizzate in prove di test a due (es. [a] vs [i]), ma queste hanno prodotto risultati simili e di poco superiori a quelli presentati nelle tabelle 9, 10, 11. Quindi si è scelto di non riportarli poiché si è ritenuto non aggiungessero ulteriori informazioni.

5.7. Analisi neurofisiologiche dei segnali: mappe di attivazione e Average

Abbiamo deciso di inserire questo paragrafo in coda perché si tratta di analisi che hanno preceduto il lavoro di elaborazione finora riportato, ma che può essere trattato in maniera scissa per la sua natura indipendente legata alla fisiologia. Vedremo attraverso le mappe di attivazione al variare nel tempo dell'area che si attivano nello svolgere i compiti. Con l'average invece è possibile ottenere la visualizzazione delle differenze medie dei compiti, sia per i singoli soggetti sia effettuando una media tra tutti i soggetti detta Grand Average. Dalle mappe di attivazione dagli *average* dei tre esperimenti risulta evidente come l'attivazione legata al pensiero di produzione si presenti con una latenza di 250ms e nelle aree dei canali c3 c4.. Il voltaggio medio risulta essere costante e tipico dei potenziali evocati. Nel contempo dagli *average* non è stato possibile estrarre delle forme d'onda facilmente riconoscibili e distinguibili per ciascuna vocale, come invece avviene per il segnale audio passato al sonogramma. Ciò è appunto dovuto all'estrema complessità del segnale proveniente dall'EEG. Risulta perciò ancora più necessario affidare la decodifica di questi segnali a strumenti di classificazione come le SVM.

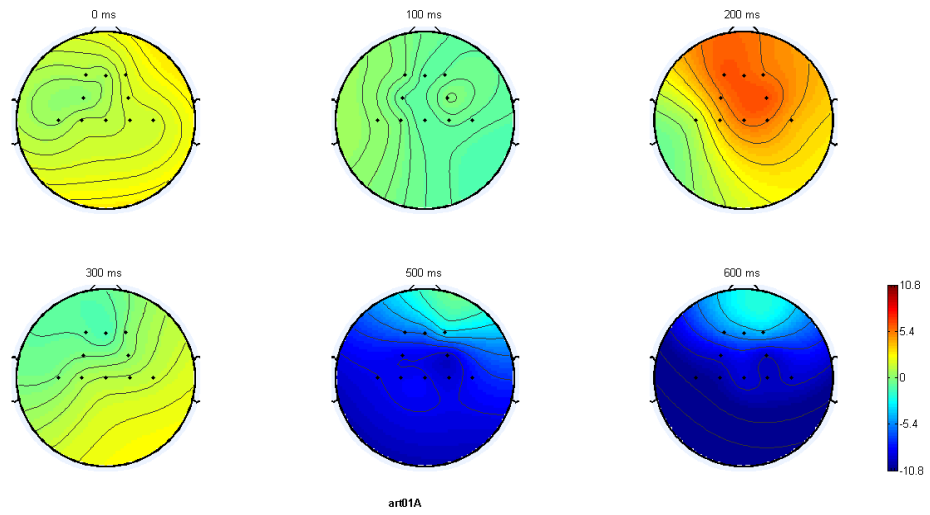


Figura 73: primo esperimento, sessione articolata, soggetto 1, /a/.

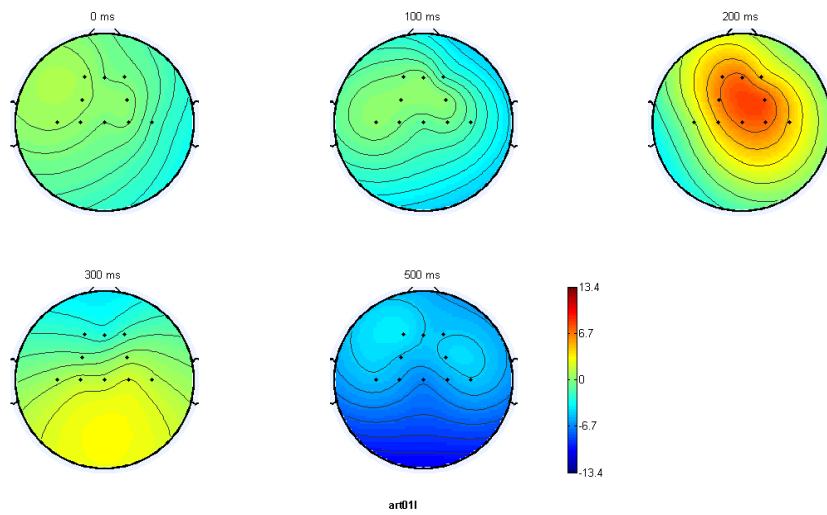


Figura 74: primo esperimento, sessione articolata, soggetto 1, /i/.

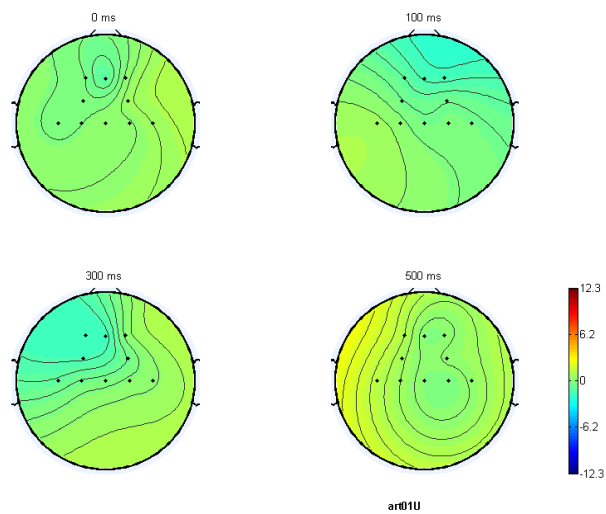


Figura 75: primo esperimento, sessione articolata, soggetto 1, /u/.

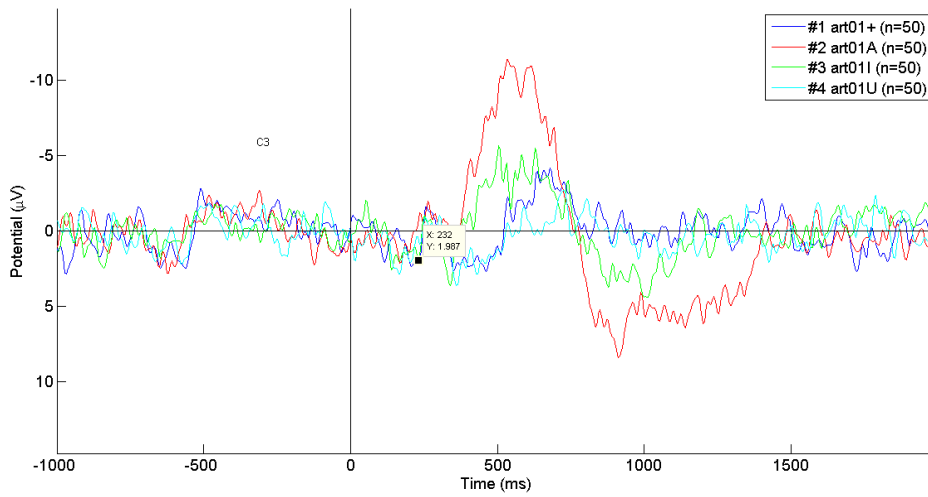


Figura 76: primo esperimento, sessione articolata, soggetto 1, average, canale C3.

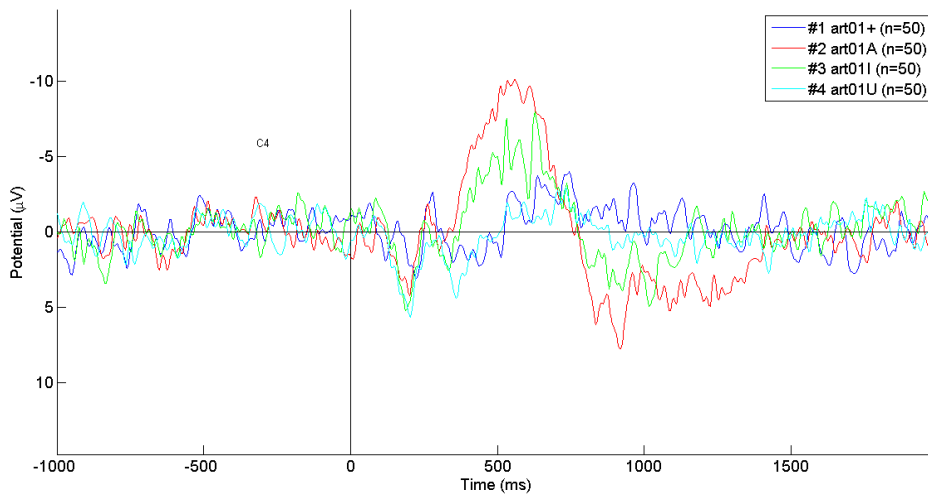


Figura 77: primo esperimento, sessione articolata, soggetto 1, average, canale C4.

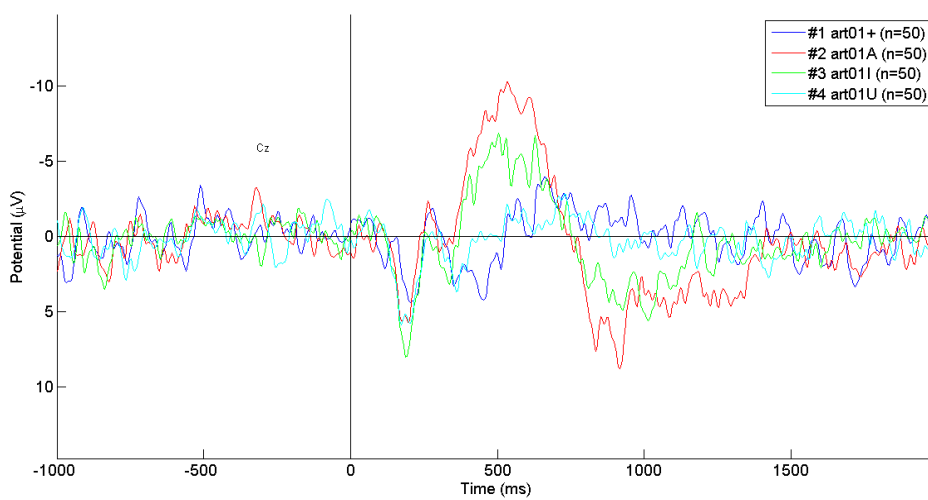


Figura 78: primo esperimento, sessione articolata, soggetto 1, average, canale Cz.

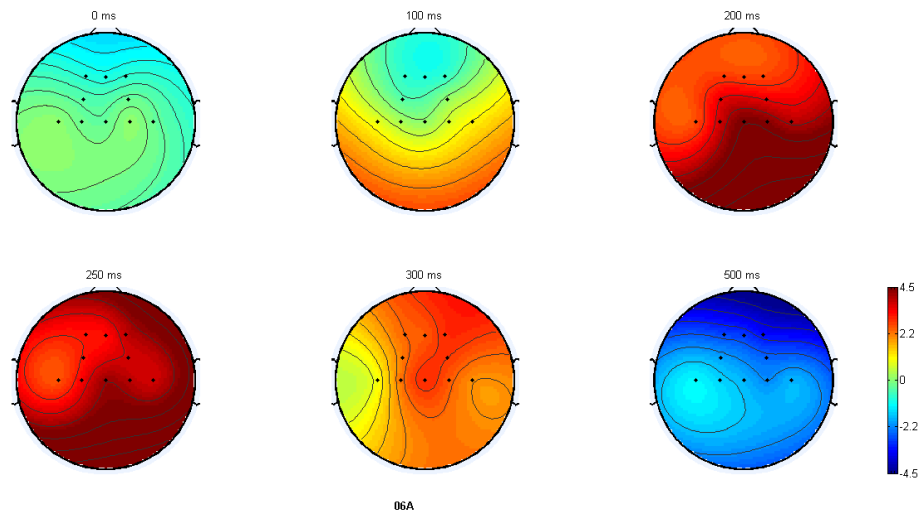


Figura 79: primo esperimento, sessione non articolata, soggetto 6, /a/.

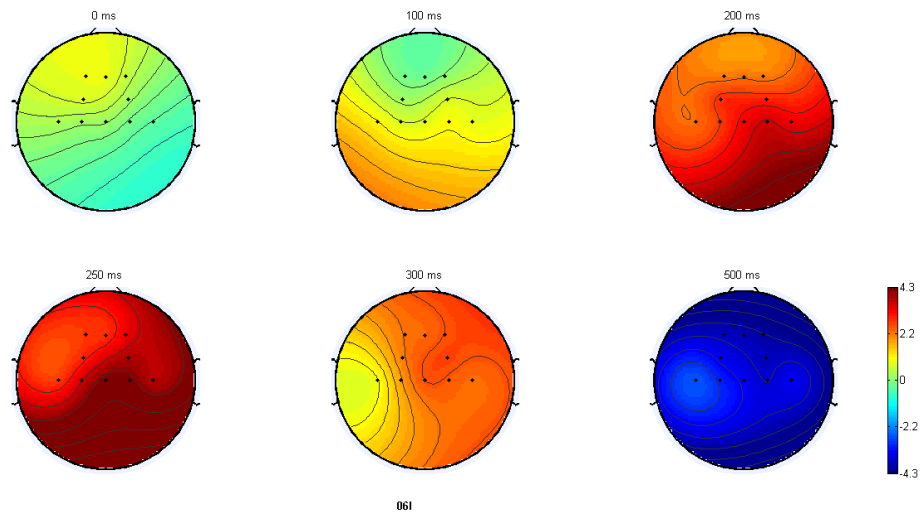


Figura 80: primo esperimento, sessione non articolata, soggetto 6, /i/.

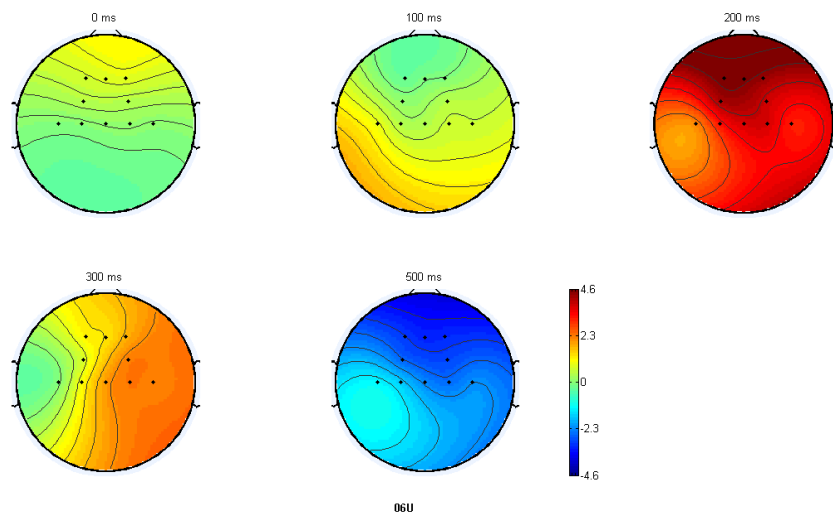


Figura 81: primo esperimento, sessione non articolata, soggetto 6, /u/.

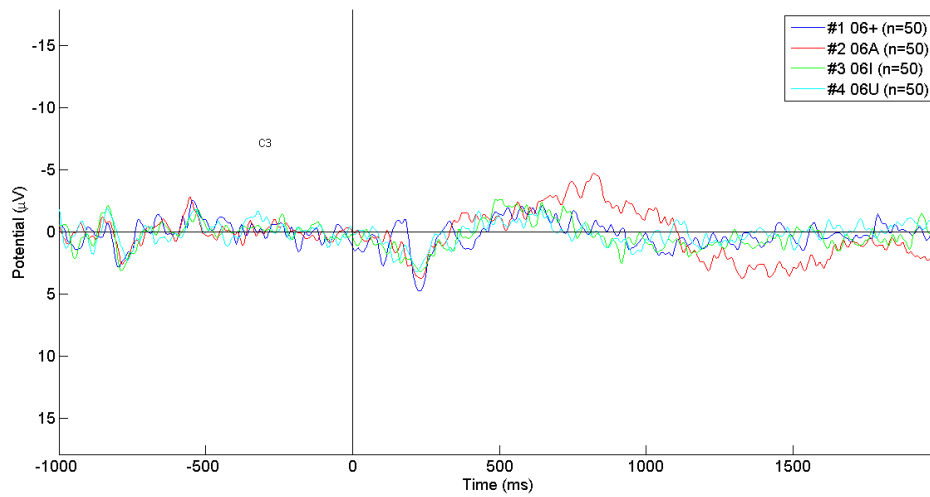


Figura 82: primo esperimento, sessione non articolata, soggetto 6, average, canale C3.

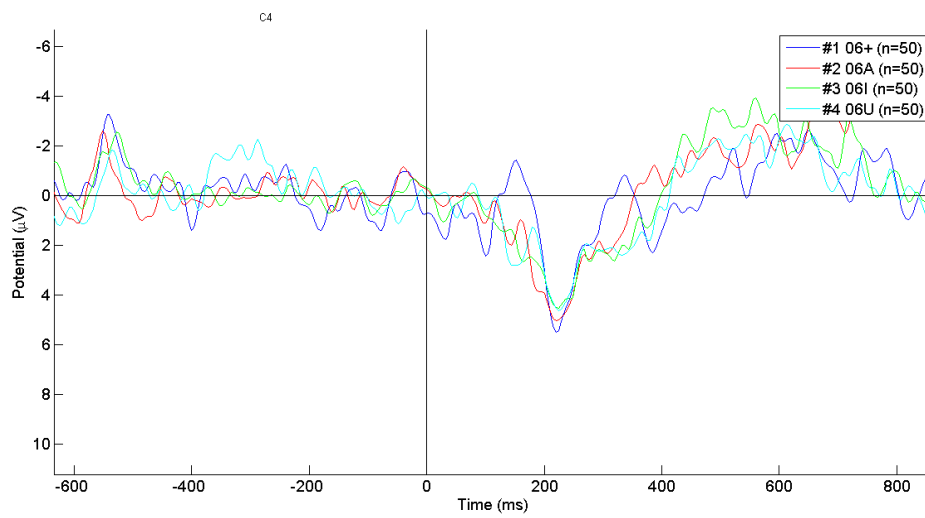


Figura 83: primo esperimento, sessione non articolata, soggetto 6, average, canale C4.

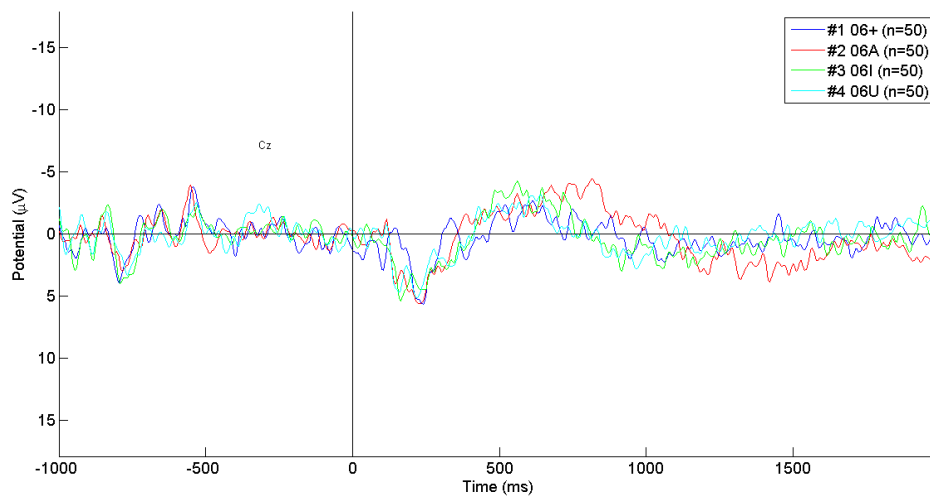


Figura 84: primo esperimento, sessione non articolata, soggetto 6, average, canale Cz.

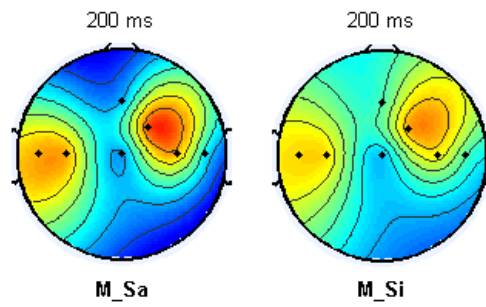


Figura 86: secondo esperimento, sessione non articolata, soggetto M, mappa di attivazione a 200 ms, /a/, /i/.

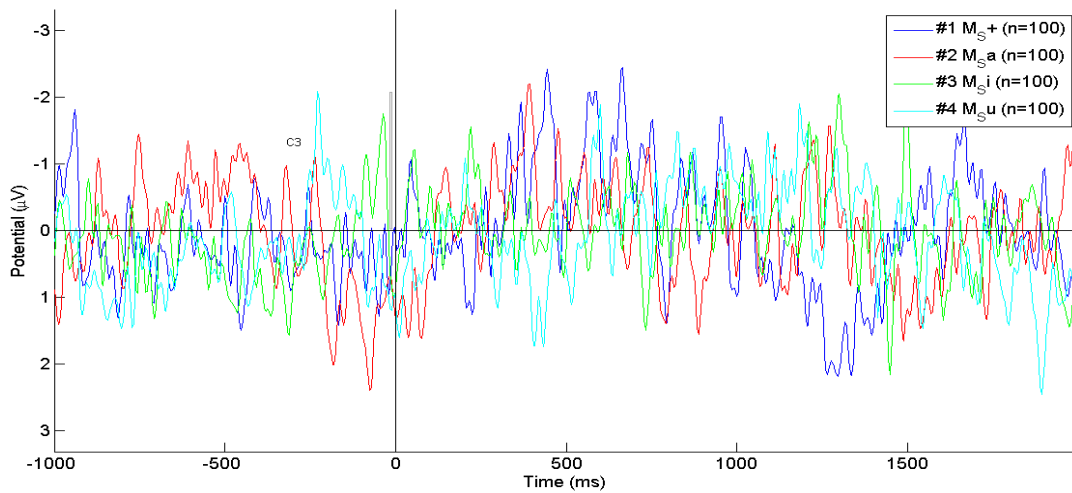


Figura 87: secondo esperimento, sessione non articolata, soggetto M, average, canale C3.

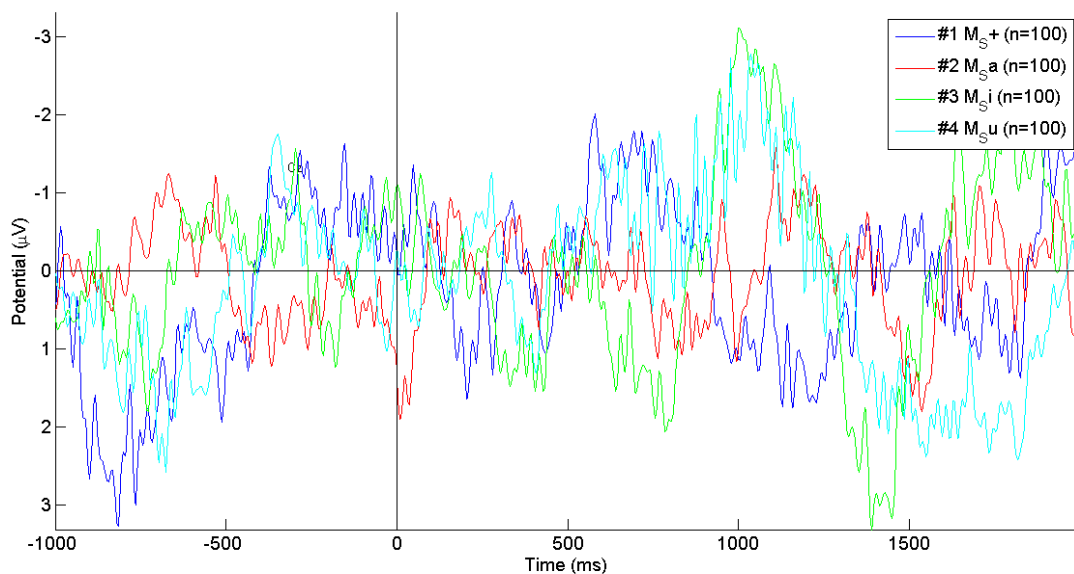


Figura 81: secondo esperimento, sessione non articolata, soggetto M, average, canale Cz.

6. CONCLUSIONI

6.1. Introduzione

Questo nostro lavoro aveva due obiettivi: verificare le prestazioni di un sistema di *Brain Computer Interface* (BCI) basato sulle onde elettroencefalografiche (EEG) nel trasmettere informazioni di tipo linguistico nell'ipotesi di utilizzare questa interfaccia per permettere la comunicazione vocale a persone che non sono in grado di parlare e portare un contributo alla conoscenza del funzionamento della comunicazione vocale.

L'ipotesi che i soggetti comprendano il parlato tramite l'identificazione dei gesti motori del tratto vocale che contribuiscono alla produzione del parlato è stata studiata anche nei suoi aspetti non verbali a partire dai neuroni specchio. Il dato di partenza consiste nell'idea d'identificazione dei gesti motori del tratto vocale che contribuiscono alla produzione del parlato, gesti che l'ascoltatore ripete in sequenza internamente imitando come ad uno specchio quelli prodotti da chi produce quel messaggio articolato.

Nell'ambito degli studi del B.C.I. un sistema di comunicazione diretto fra il cervello dell'utente e la macchina si sono dapprima utilizzati sistemi invasivi di prelievo delle onde cerebrali con elettrodi impiantati all'interno del tessuto cerebrale, ma oggi la strada è quella, non invasiva, di cuffie molto sensibili per tecnologia EEG in grado di captare i segnali con un livello di qualitativo vicino agli impianti. I segnali raccolti da una cuffia hanno, come abbiamo visto, alcune caratteristiche peculiari: la presenza di numerosi artefatti, ovvero segnali causati da movimenti motori come aprire e chiudere gli occhi, ma anche la stessa respirazione, appare quindi necessario riuscire a distinguere tra segnale utile e non. E' possibile individuare questi segnali analizzando e filtrando il segnale attraverso le procedure che di colta in volta si adattano meglio al tipo di indagine condotta e al tipo di segnale da estrarre. Avremo così la localizzazione in particolari aree cerebrali a seconda della tipologia dei segnali che interessa analizzare e la presenza di picchi significativi legati ad alcune caratteristiche rilevanti. Per il BCI con P300 l'elettrodo principale sarà Fz e la componente sarà appunto la P300 (che appare 350ms dopo lo stimolo).

Possiamo perciò definire due modalità il "BCI semplice" e il "BCI complesso" a seconda della strumentazione utilizzata e del livello di complessità di istruzioni da impartire.

Il "BCI semplice" è costituito da un sistema di riconoscimento di parametri neuro fisiologici che vengono modulati dall'utente al fine di far compiere alla macchina una serie di operazioni semplici; per esempio attraverso la modulazione dell'onda cerebrale P300 è possibile spostare il cursore sul monitor. Il "BCI complesso" è costituito da un sistema di riconoscimento delle onde cerebrali con particolare attenzione alle onde provenienti dall'area motoria e pre-motoria. In altri termini noi potremmo chiedere al soggetto di spalancare la bocca

se legge o sente la vocale /a/ e di strizzare un occhio se legge o sente la vocale /i/ ma questo nulla ci direbbe sul funzionamento della comunicazione vocale e certo non ci permetterebbe di realizzare una macchina in grado di comandare un apparato fonatorio sintetico (cioè un elaboratore con un sintetizzatore vocale).

I risultati raggiunti sono frutto di un lavoro che ha avuto diverse fasi, tentativi e rivalutazioni di strade già intraprese, un lavoro fatto di mediazioni tra il lato tecnologico e il lato umano. Appare chiaro come i risultati siano dipendenti da numerose variabili come la strumentazione utilizzata, i settaggi di registrazione e le successive fasi di filtraggio, ma sicuramente in modo preponderante dipendono dai soggetti, dalle loro caratteristiche fisiche e psicofisiche che traspaiono sul segnale EEG.

6.2. Primo esperimento

Per il primo esperimento è stata adottata la procedura di confronto a due perché ritenuta la più comune. Tale procedura è stata adottata nel lavoro di DaSalla del 2009 in cui era previsto che 3 soggetti dovessero produrre i compiti [a] e [u], questo studio è stato un riferimento utile nella fase iniziale del nostro lavoro; infatti si trattava dell'unico esperimento pubblicato che adottasse l'EEG e quindi una tecnica non invasiva per riconoscere il parlato silenzioso. Tuttavia l'utilizzo della stessa metodologia non ha prodotto i risultati sperati.

Una possibile spiegazione è che la metodologia sperimentale DaSalla sia perché troppo legata al tipo di protocolli adottati per il riconoscimento dei compiti nel "BCI semplice" dipendente dalla P300. Infatti l'elaborazione dei primi dati ha evidenziato una importante distinzione tra il presente studio e quello di riferimento: l'individuazione di una diversa componente da analizzare. La componente neurofisiologica legata ai compiti di "pensiero di produzione" è risultata essere la P200 con le sue specifiche caratteristiche di ampiezza e latenza. Ciò è stato ottenuto registrando il segnale proveniente dagli elettrodi Fz, Cz, C3, C4, C5, C6 (e gli EOG poi rimossi in fase di filtraggio) in corrispondenza con l'area premotoria e motoria. Quindi è apparso che si dovevano usare strategie di analisi che rispondessero non a un compito legato alla P300, ma ad una domanda diversa nel tentativo di individuare i singoli compiti differenziandoli gli uni dagli altri. Sono stati adottati apportate una serie di correzioni alla macchina: la finestra di analisi è stata centrata a 250ms dopo lo stimolo; la creazione di un contesto quanto più possibile alla realtà di impiego in cui l'utente pensa di produrre un singolo compito all'interno di una serie di compiti possibili e inseriti nelle conoscenze della macchina. Ad oggi il sistema da noi sperimentato è in grado di riconoscere il pensiero volontario di produzione di 3 foni distinti ed è in grado di addestrare un

algoritmo a distinguere e quindi riconoscere i foni tra di loro con un tasso di riconoscimento pari all'80%. Sia l'algoritmo **misto** (mediato) che il **singolo** danno risultati convincenti ma nelle applicazioni sarà auspicabile utilizzare il secondo perchè si adatta alle necessità di velocità di risposta ed è l'unico che consentirebbe di comandare un sistema di sintesi della voce. Vediamo riassunti nelle tabelle qui di seguito i valori medi ottenuti nei singoli esperimenti.

STIMOLO	Misto		singolo	
	Media	DS	Media	DS
A	74,76	13,21	69,2	5,6
I	72,27	13,7	67,1	5,5
U	78,5	12,7	68,7	5,2
+	80,37	12,3	67,1	5,6

Tabella 10: valori medi di riconoscimento per ogni compito relativi al primo esperimento, sessione 1, solo pensato.

Dalla tabella risulta essere più efficace l'algoritmo **misto** con una percentuale di riconoscimento del 76%, i valori medi di riconoscimento per il compito [a] del 74% per il compito [i] del 77% e per il compito [+] del 75% . L'elevato numero di soggetti (9) ha portato ad un ventaglio percentuali che vanno da un valore minimo del 60% per il soggetto 3 nel caso del riconoscimento della [i], mentre il valore massimo del 91% è ottenuto dal soggetto 6 nel caso del riconoscimento della [u]. Ricordiamo che in generale per questa sessione 4 soggetti ottengono l'80% e più di riconoscimento, anche se la deviazione standard (DS) è elevata al 13 % circa. Per l'algoritmo singolo i risultati percentuali sono più bassi, intorno al 68% ma in questo caso la DS è inferiore di 8 punti percentuali rispetto all'algoritmo precedente. In questo caso il miglior risultato l'ottiene il soggetto 6 per tutti compiti mentre gli altri rimangono nella media.

STIMOLO	MISTO	DS	SINGOLO	DS
a	78,4	12,6	67,6	5,6
i	81	11,9	66	4,6
u	76,8	13,8	68,4	5,6
+	79	12,8	67,2	5,5

Tabella 11: valori medi di riconoscimento per ogni compito relativi al primo esperimento, sessione 2, articolato senza fonazione.

In questa seconda sessione notiamo un miglioramento nel riconoscimento di appena 2 punti medi percentuali per il **misto** mentre per il **singolo** rimane invariato. In questo caso solo un soggetto tende a staccarsi dalla media, si tratta del soggetto 1 per il quale con il **misto** abbiamo il riconoscimento di [a] al 91,1% di [i] al 89,8% di [u] al 91,2%. Lo stesso soggetto ottiene valori medi buoni anche con il **singolo**: 84,9% per [a], 80,8% e 84,7% per [i] e [u]. Per entrambi gli algoritmi la deviazione standard è intorno al 6-5%.

6.3. Secondo esperimento

Nel secondo esperimento i 3 soggetti partecipanti hanno risultati molto vicini tra loro e allineati alla media. Nella prima sessione silente il compito [a] è riconosciuto con il misto al 79,3% mentre con il singolo siamo al di sotto solo di 3 punti percentuali. In generale si osserva una tendenza per cui la deviazione standard del misto è compresa tra 14 e il 17% mentre per il singolo i suoi valori DS sono molto bassi intorno al 3,8%. Nella prima sessione il soggetto M ottiene dei risultati di riconoscimento di poco superiori alla media con l' 83% per [a] e per [u] e spicca il riconoscimento di [i] al 94,2%.

STIMOLO	MISTO	DS	SINGOLO	DS
A	79,3	14,9	73,2	3,8
I	78,1	14,1	73,6	3,9
U	77,3	14,9	74,8	3,7
+	72,8	17,3	76,3	3,7

Tabella 12: valori medi di riconoscimento per ogni compito relativi al secondo esperimento, sessione 1, solo pensato.

STIMOLO	MISTO	DS	SINGOLO	DS
A	76,8	15	70,4	3,9
I	71,5	17	71,1	3,6
U	73,3	15,3	68,2	3,8
+	71	16,3	68,9	3,8

Tabella 13: valori medi di riconoscimento per ogni compito relativi al secondo esperimento, sessione 2, articolazione con fonazione.

6.4. Terzo esperimento

Analizzando i risultati in tabella risulta evidente che il migliore nei test è il soggetto del terzo esperimento con una percentuale media di riconoscimento del 90%. In questo caso l'al-

goritmo singolo ha risultati migliori, intorno al 95%, rispetto al misto che si attesta intorno a 88,8%. Anche la deviazione standard è migliore nell'algorithmo singolo al di sotto della metà del misto.

STIMOLO	MISTO	DS	SINGOLO	DS
A	86	10	94.8	4.2
I	90.5	7	95.5	4
U	90	7	95.8	3.9

Tabella14: valori medi di riconoscimento per ogni compito relativi al terzo esperimento, sessione 1,2,3, solo pensato.

I risultati derivanti dai test sulla “produzione” sono convincenti poiché si aggirano tra il 70% e 90% a seconda dell’algorithmo e dei soggetti che ricordiamo essere il fattore determinante; inoltre i valori minori di DS nel singolo ci fanno preferire questo algorithmo per una fase di utilizzo reale mentre il misto per una fase di studio.

Il risultato ottenuto nel 3° esperimento e le modalità con il quale è stato effettuato ci porta a riconsiderare anche quei soggetti che negli altri esperimenti non hanno ottenuto un sufficiente riconoscimento; infatti si ritiene che con l'opportuno addestramento del soggetto e della taratura della macchina è possibile anche per questi raggiungere i risultati ottimali. Ciò significa che possono esistere due tipologie di macchine, una generica più economica sotto il profilo sia della sua struttura sia in termini di tempo da dedicare al soggetto e alla capacità di adattamento dei comandi dell'uno alle risposte dell'altra. Un macchina generica avrà la possibilità di dare risposte esatte nel 70% dei casi quindi di poter essere utilizzata in contesti comunicativi dove non è necessaria la massima accuratezza o che prevedono un numero limitato di parole su cui la macchina potrebbe essere istruita. La tipologia più complessa di questa macchina per il parlato silenzioso che abbiamo detto essere costosa soprattutto sotto il profilo umano, ma mediamente il 20% più precisa dell'altra, potrebbe essere utilizzata da utenti che hanno necessità di comunicazione molto ampia o che necessitano un'estrema chiarezza.

6.5. Contributi alla conoscenza del funzionamento del linguaggio articolato

Questo studio sull'articolazione del linguaggio è un elemento che si somma a quanto già stato sperimentato di questi tempi dal gruppo del MIT di Boston che fa riferimento al prof . F. Guenther. Un elemento per comprendere le modalità con cui produciamo una verbaliz-

zazione, abbiamo messo sotto analisi una piccola porzione dell'oscuro percorso che va dal pensiero alla parola: i comandi per la produzione. Non stiamo investigando aree della competenza del parlante in una specifica lingua, ma le sue capacità di produrre genericamente dei fonemi per poi comunicare un dato significato.

I nostri dati ci indicano come elettrodi principali per individuare i compiti motori quelli posizionati nell'area motoria/pre-motoria: Fz, Cz, C3, C4, C5 e C6, mentre per quello che riguarda l'individuazione temporale del comando neuromotorio, i nostri esperimenti hanno risultati coincidenti nella finestra comprendente i 250ms dallo stimolo, mentre il gesto motorio vero e proprio appare dopo 500-600ms. Questi risultati non indicano la forma del pensiero o l'idea di lingua ma semplicemente i comandi neurofisiologici che permettono al nostro apparato fonatorio di compiere il gesto comunicativo. Abbiamo cercato e abbiamo trovato l'esistenza di una diversità tra i comandi motori per i fonemi da noi scelti, e una diversità, lieve, anche all'interno dello stesso gruppo di comandi motori relativi allo stesso fonema e ciò si rivela molto importante sotto la prospettiva della teoria motoria che infatti li vedeva come degli invarianti. Si riscontra anche a livello di comandi motori una varietà che prima era assegnata alla sola esecuzione e addebitata a fenomeni di contatto o altro; questo studio sembrerebbe dimostrare che l'invarianza non si riscontra al livello di programmazione del gesto articolatorio di fonazione, ma bensì ad un livello più alto.

Questo approccio sembrerà ai più, di tipo meccanicistico, ma la risposta sulla sostanza delle idee la lasciamo ad altri, offrendogli un punto da cui partire o per chi volesse da cui ripartire.

6.6. Per il Futuro

Partendo da quanto è stato fatto ci si propone di proseguire cercando di creare un inventario di fonemi, ciò potrà ricalcare sia l'inventario della lingua italiana come quella di altre lingue oppure essere ridotto-semplificato per intenderci a quello dei grafemi. Questo renderebbe possibile l'applicazione nella comunicazione al di fuori dell'ambito di ricerca.

Per migliorare e rendere più rapido il lavoro della macchina sarà necessario lavorare sull'hardware; infatti bisognerà avere delle cuffie che possano essere facili da indossare e con elettrodi di materiali che adottino soluzioni nanotecnologiche e di tessuti biologici così da essere ben tollerate dagli utenti e con entrata di segnale che riduca a 0 il rumore.

Così anche la riduzione degli artefatti contribuirà a migliorare le risposte.

Ovviamente queste soluzioni potrebbero non essere del tutto efficaci senza un opportuno lavoro preparatorio che coinvolga gli utenti delle macchine infatti come già si evince dai ri-

sultati del test sul terzo esperimento in cui era coinvolto un solo soggetto (si veda tabella sopra), sarà necessaria la calibrazione delle macchine di volta in volta, un po' come avviene per i sistemi di dettatura che noi tutti oggi conosciamo e utilizziamo e sarà quindi possibile realizzare apparecchi *silent-speaker-dependent* e altri *silent-speaker-independent* a seconda dell'ambito di utilizzo.

Infatti questa tecnologia potrà essere utilizzata in diverse aree della comunicazione in sostituzione di quella verbale di tipo acustico. Gli utenti potranno utilizzare la macchina come supporto nel caso di deficit totale del proprio apparato fonatorio, ma potrà essere utilizzato anche come mezzo di comunicazione in tutti quei contesti in cui l'uso del mezzo acustico è precluso per motivi di sicurezza o di ambiente rumoroso.

Ringraziamenti

Questi 3 anni di dottorato sono stati davvero ricchi di esperienze sia sul piano professionale che su quello emotivo-personale. Sono stati anni di duro impegno per il tipo di indirizzo di ricerca del tutto nuovo non solo per me, ma per la comunità scientifica tutta.

Sono stati faticosi per lo sforzo di conciliare giorno e notte la ricerca e la maternità.

Quindi ringrazio Gigi e i miei genitori per avermi appoggiato, spronato, sopportato e sostenuto nei momenti in cui ho temuto di non farcela sia come madre sia come dottoranda.

Ringrazio l'Ing. A. Paoloni e il Prof M. Grimaldi, per l'aver creduto in questo progetto e nell'avermi messo a disposizione gli strumenti per portare a termine il compito.

Ringrazio lo staff del CRIL per la strumentazione e le competenze messe a disposizione per la registrazione degli esperimenti 1 e 2.

Ringrazio per la parte tecnica di elaborazione dati il dott. Massimiliano Todisco.

Un ringraziamento speciale va a tutti i soggetti salentini e non che si sono prestati a titolo gratuito (Rocco, Vladimir, Laura, Marco, Matteo...) un particolare ringraziamento va all'amico e collega Dott. Marco Maggiore che per ben due volte è stato esposto alla "gogna".

Vorrei quindi ringraziare l'intero reparto di Neurologia dell'ospedale Belcolle, per la disponibilità del suo primario il Dott. Salvatori, un grazie di cuore è tutto per le "ragazze" (Albina, Loredana, Stefania e Sonia) per la professionalità, la gentilezza, l'amicizia e la grande disponibilità.

Per i giorni zurighesi ringrazio il Prof. S. Shmidt e il Prof. M. Loporcaro e un grazie anche a Vincenzo, Alessandra e Martina e al caro Lorenzo.

Grazie a tutti quelli che in maniera diretta hanno contribuito alla riuscita di questo lavoro

7. BIBLIOGRAFIA

Arnal, A., Badin, P., Brock, G., Connan, P.-Y., Florig, E., Perez, N., Perrier, P. Simon, P., Sock, R., Varin, L., Vaxelaire, B., Zerling, J.-P., (2000). *Une base de données ciné radiographiques du français, XXIII^e-mes Journées d'Etude sur la Parole*, Aussois, France, pp. 425–428.

Baken, R.J., Robbins, J., Fisher, H., Blom, E., Singer, M., (1984). *A comparative acoustic study of normal, esophageal and tracheoesophageal speech production*. *J. Speech Hearing Disorders* 49, 202–210.

Bartels, J.L., Andreasen, D., Ehirim, P., Mao, H., Seibert, S., Wright, E.J., Kennedy, P.R., (2008). *Neurotrophic electrode: method of assembly and implantation into human motor speech cortex*. *J. Neurosci. Methods* 174 (2), 168–176.

Betts, B.J., Binsted, K., Jorgensen, C., (2006). *Small-vocabulary speech recognition using surface electromyography*. *Interact. Comput.: Interdisciplinary J. Human–Comput. Interact.* 18, 1242–1259.

Birbaumer, N., (2000). *The thought translation device (TTD) for completely paralyzed patients*. *IEEE Trans. Rehabil. Eng.* 8 (2), 190–193.

Blankertz, B., Dornhege, G., Krauledat, M., Müller, K.-R., Kunzmann, V., Losch, F., Curio, G., (2006). *The Berlin brain–computer interface: EEG-based communication without subject training*. *IEEE Trans. Neural Systems Rehabil. Eng.* 14 (2), 147–152.

Blom, E.D., Singer, M.I., (1979). *Surgical prosthetic approaches for postlaryngectomy voice restoration*. In: Keith, R.L., Darley, F.C. (Eds.), *Laryngectomy Rehabilitation*. Texas College Hill Press, Houston, pp. 251–276.

Blumstein S., Stevens K., (1979), *Acoustic invariance in speech production: evidence from measurements of spectral characteristics of stop consonants*, *JASA*, 66, p. 1001-1017.

Blumstein S., Stevens K., (1980), *Perceptual invariance and onset spectra for stop consonants in different vowel environments*, *JASA*, 67, p. 648-662.

Bos, J.C., Tack, D.W., (2005). *Speech Input Hardware Investigation for Future Dismounted Soldier Computer Systems*, DRCD Toronto CR 2005-064.

Brown, D.R., Ludwig, R., Pelteku, A., Bogdanov, G., Keenaghan, K., (2004). *A novel non-acoustic voiced speech sensor*. Meas. Sci. Technol.15, 1291–1302.

Brown, D.R., Keenaghan, K., Desimini, S., (2005). *Measuring glottal activity during voiced speech using a tuned electromagnetic resonating collar sensor*. Meas. Sci. Technol. 16, 2381–2390.

Brumberg, J.S., Andreasen, D.S., Bartels, J.L., Guenther, F.H., Kennedy, P.R., Siebert, S.A., Schwartz, A.B., Velliste, M., Wright, E.J., (2007). *Human speech cortex long-term recordings: formant frequency analyses*. In: Neuroscience Meeting Planner 2007, Program No. 517.17, San Diego, USA.

Brumberg, J.S., Nieto-Castanon, A., Guenther, F.H., Bartels, J.L., Wright, E.J., Siebert, S.A., Andreasen, D.S., Kennedy, P.R., (2008). *Methods for construction of a long-term human brain machine interface with the Neurotrophic Electrode*. In: Neuroscience Meeting Planner 2007, Program No. 779.5, Washington, DC.

Brumberg, J.S, Nieto-Castanon, A., Kennedy, P.R., Guenther, F.H.,(2010). *Brain–computer interfaces for speech communication*. Speech Comm. 52 (4), 367–379.

Burnett, G.C., Gable, T.G., Holzrichter, J.F., Ng, L.C., (1997). *Voiced excitation functions calculated from micro-power impulse radar information*. J. Acoust. Soc. Amer. 102, 3168(A).

Carstens Medizinelektronik, (2008). <<http://www.articulograph.de/>>.

Chan, A.D.C., Englehart, K., Hudgins, B., Lovely, D.F., (2001). *Myoelectric signals to augment speech recognition*. Med. Biological Eng. Comput. 39, 500–504.

Chan, A.D.C., (2003). *Multi-expert automatic speech recognition system using myoelectric signals*. Ph.D. Dissertation, Department of Electrical and Computer Engineering, University

of New Brunswick, Canada.

Cherkassky V. e F. Mulier,(1998) *Learning from data - Concepts, Theory and Methods*, Wiley .

F. Cincotti, L. Bianchi, G. Birch, C. Guger, J. Mellinger, R. Scherer, R. N. Schmidt, Oscar Y. Suárez, and G. Schalk, IEEE BCI Meeting 2005—Workshop on Technology: Hardware and Software transactions on neural systems and rehabilitation Engineering, vol. 14, no. 2, june 2006

Cole R.A., e Jakimik J.,(1980), *A model of speech perception*, in Cole R.A.,(ed). *Perception and Production of Fluent speech*, Erlbaum, Hillsdale, N.J.

Cole R.A., Scott B., (1974), *The phantom of the phoneme: invariant cues for stop consonants*, *Perception and Psychophysics*, 15, p. 101-107.

Crevier-Buchman, L., (2002). *Laryngectomy patients and the psychological aspects of their tracheostomy*. *Rev. Laryngol. Otolaryngol. Rhinol.* 123, 137–139.

DaSalla, C.S., Kambara, H., Sato, M., Koike, Y., (2009). *Spatial filtering and single-trial classification of EEG during vowel speech imagery*. In: *Proceedings of the 3rd International Convention on Rehabilitation Engineering & Assistive Technology*, Singapore. Article No. 27. ACM, New York, NY, USA. ISBN:978-1-60558-792-9.

D'Ausilio A, Bufalari I, Salmas P, Fadiga L.,(2011). [The role of the motor system in discriminating normal and degraded speech sounds](#), *ELSERVIER* , *Cortex*
doi:10.1016/j.cortex.2011.05.017

D'Ausilio A, Bufalari I, Salmas P, Busan P, Fadiga L.,(2011). [Vocal pitch discrimination in the motor system](#), *Brain Lang.* 2011 Mar 30

D'Ausilio et al.(2009), *The Motor Somatotopy of Speech Perception*, *Current Biology* ,
doi:10.1016/j.cub.2009.01.017

Davidson, L., (2005). *Comparing tongue shapes from ultrasound imaging using smoothing spline analysis of variance*. J. Acoust. Soc. Amer. 120 (1), 407–415.

Delorme A & Makeig S, (2004). *EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics*, Journal of Neuroscience Methods 134:9-21

Delorme, A., Mullen, T., Kothe, C., Bigdely-Shamlo, N., Akalin, Z., Vankov, A., Makeig, S. (2011) *EEGLAB, MPT, NetSIFT, NFT, BCILAB, and ERICA: New tools for advanced EEG/MEG processing*. Computational Intelligence, article ID 130714.

Dekens, T., Patsis, Y., Verhelst, W., Beaugendre, F., Capman, F., (2008). *A multi-sensor speech database with applications towards robust speech processing in hostile environments*. In: Proc. 6th Internat. Language Resources and Evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco, 28–30 May 2008.

B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, J.S. Brumberg, (2010), *Silent speech interfaces*, Speech Communication 52 (2010) 270–287 Elsevier

Denby, B., Stone, M., (2004). *Speech synthesis from real time ultrasound images of the tongue*. In: Proc. IEEE Internat. Conf. on Acoustics, Speech, and Signal Processing, (ICASSP'04), Montre´al, Canada, 17–21 May 2004, Vol. 1, pp. 1685–1688.

Denby, B., Oussar, Y., Dreyfus, G., Stone, M., (2006). *Prospects for a Silent Speech Interface Using Ultrasound Imaging*. IEEE ICASSP, Toulouse, France, pp. 1365–1368.

Dornhege, G., del R. Millan, J., Hinterberger, T., McFarland, D., Mu¨ller, K.-R. (Eds.), (2007). *Towards Brain–Computer Interfacing*. MIT Press.

Drummond, S., Dancer, J., Krueger, K., Spring, G., (1996). *Perceptual and acoustical analysis of alaryngeal speech: determinants of intelligibility*. Percept. Motor Skills 83, 801–802.

R. Duda, P. Hart e G. Stork (2000), *Pattern classification*, Wiley 2000.

Dupont, S., Ris, C., (2004). *Combined use of close-talk and throat microphones for improved speech recognition under non-stationary background noise*. In: Proc. Robust 2004, Workshop (ITRW) on Robustness Issues in Conversational Interaction, Norwich, UK, August 2004.

Elman J.L., McClelland J.L., (1984) “*Speech perception as a cognitive process: the interactive activation model*”, in *Speech and Language: advances in basic research and practice*, vol.10, Academic press, p. 337-374.

Epstein, C.M., (1983). *Introduction to EEG and evoked potentials*. J.B. Lippincot Co.

Epstein, M.A., (2005). *Ultrasound and the IRB*. *Clin. Linguist. Phonet.* 16 (6), 567–572.

Fadiga L, Craighero L, Buccino G, Rizzolatti G (2002) *Speech listening specifically modulates the excitability of tongue muscles: a TMS study*. *European Journal of Neuroscience*;15:399-402.

Fagan, M.J., Eil, S.R., Gilbert, J.M., Sarrazin, E., Chapman, P.M., (2008). *Development of a (silent) speech recognition system for patients following laryngectomy*. *Med. Eng. Phys.* 30 (4), 419–425.

Fant G., (1968), *Analysis and Synthesis of Speech Processes*”, in Malmberg B.,(ed.) *Manual of Phonetics*, North Holland, Amsterdam, p. 173-277.

Fitzpatrick, M., (2002). *Lip-reading cellphone silences loudmouths*, *New Scientist*, edition of 03 April 2002. Furui, S., 2001. *Digital Speech Processing Synthesis and Recognition*, second ed. Marcel Dekker. Fodor J., 1983, *The modularity of mind*, MIT press, Cambridge Masschusetts.

Forester K.J., (1976), *Accessing the mental lexicon*, in Wales R., Walker E., (eds.), *New approaches to language mechanisms*, Amsterdam, p. 257-287.

Forester K.,(1989), *Basic issues in lexical processing*, in Marslen-Wilson, *Lexical representation and process*, Cambridge Massachusetts, MIT Press, p. 75-107.

Fraunfelder U.H., Tyler L.,(1987), *The process of spoken recognition: an introduction*, Cognition 25, p. 1-20.

Fujisaki H., Kawashima T., (1969) *On the modes and mechanisms of speech perception*, Annual report of the Engineering Research Institute, University of Tokyo, 28, p. 63-77.

Fujisaki H., Kawashima T., (1971), *A model of the mechanism for speech perception. Quantitative analysis of categorical effects in discrimination*, Annual report of the Engineering Research Institute, University of Tokyo, 30, p. 59-68.

Galantucci, B., Fowler,C.A., and Turvey, M.T. (2006). The motor theory of speech perception reviewed. *Psychon. Bull. Rev.* 13, 361–377.

Georgopoulos, A.P., Kalaska, J.F., Caminiti, R., Massey, J.T., (1982). *On the relations between the direction of two-dimensional arm movements and cell discharge in primate motor cortex*. *J. Neurosci.* 2 (11), 1527–1537.

Gibbon, F., (2005). Bibliography of Electropalatographic (EPG) Studies in English (1957–2005), Queen Margaret University, Edinburgh, UK, September 2005.
<http://www.qmu.ac.uk/ssrc/cleftnet/EPG_biblio_2005_september.pdf>.

Gracco, V.L., Tremblay, P., Pike, B., (2005). *Imaging speech production using fMRI*. *NeuroImage* 26 (1), 294–301, 15 May.

Guenther, F.H., Brumberg, J.S., Nieto-Castanon, A., Bartels, J.L., Siebert, S.A., Wright, E.J., Tourville, J.A., Andreasen, D.S., Kennedy, P.R., (2008). *A brain–computer interface for real-time speech synthesis by a locked-in individual implanted with a Neurotrophic Electrode*. In:Neuroscience Meeting Planner 2008, Program No. 712.1, Washington,DC.

Guenther, F.H., Satrajit, S.G., & Tourville, J.A. (2006). *Neural modeling and imaging of the cortical interactions underlying syllable production*. *Brain & Language*, 96, 280-301.

Guenther, F.H. (2006). *Cortical interactions underlying the production of speech sounds*. *Journal of Communication Disorders*, 39, 350-365.

Guenther, F.H., Hampson, M., & Johnson, D. (1998). *A theoretical investigation of reference frames for the planning of speech movements*. *Psychological Review*, 105, 611-633.

F.H. Guenther,(1995). *Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production*. *Psychological Review*, 102, 594-621.

Hasegawa, T., Ohtani, K., (1992). *Oral image to voice converter, image input microphone*. In: Proc. IEEE ICCS/ISITA 1992 Singapore, Vol. 20, No. 1, pp. 617–620.

Heracleous, P., Kaino, T., Saruwatari, H., Shikano, K., (2007). *Unvoiced speech recognition using tissue-conductive acoustic sensor*. *EURASIP J. Adv. Signal Process.* 2007 (1), 1–11.

Hirahara, T., Otani, M., Shimizu, S., Toda, M., Nakamura, K., Nakajima, Y., Shikano, K., (2010). *Silent-speech enhancement system utilizing body-conducted vocal-tract resonance signals*. *Speech Comm.*52 (4), 301–313.

Hochberg, L.R., Serruya, M.D., Friehs, G.M., Mukand, J.A., Saleh, M., Caplan, A.H., Branner, A., Chen, D., Penn, R.D., Donoghue, J.P., (2006). *Neuronal ensemble control of prosthetic devices by a human with tetraplegia*. *Nature* 442 (7099), 164–171.

Hochberg, L.R., Simeral, J.D., Kim, S., Stein, J., Friehs, G.M., Black, M.J., Donoghue, J.P., (2008). *More than two years of intracortically based cursor control via a neural interface system*. In: Neuroscience Meeting Planner 2008, Program No. 673.15, Washington, DC.

Holmes, J., Holmes, W., (2001). *Speech Synthesis and Recognition*. Taylor and Francis.

Hoole, P., Nguyen, N., (1999). *Electromagnetic articulography in coarticulation research*. In: Hardcastle, W.H., Hewlitt, N. (Eds.), *Coarticulation: Theory, Data and Techniques*. Cambridge University Press, pp.260–269.

House et alii,(1962), *On the learning of speechlike vocabularies*, *Journal of Verbal Learning and Verbal Behaviour*, vol.1 p. 133-143.

House, D., Granström, B., (2002). *Multimodal speech synthesis: improving information flow in dialogue systems using 3D talking heads*. In: Artificial Intelligence: Methodology, Systems, and Applications. Lecture Notes in Computer Science, Vol. 2443/2002. Springer, Berlin/Heidelberg, pp. 65–84.

Hueber, T., Aversano, G., Chollet, G., Denby, B., Dreyfus, G., Oussar, Y., Roussel, P., Stone, M., (2007a). *Eigen-tongue feature extraction for an ultrasound-based silent speech interface*. In: IEEE Internat. Conf. on Acoustic, Speech, and Signal Processing, ICASSP07, Honolulu, Vol. 1, pp. 1245–1248.

Hueber, T., Chollet, G., Denby, B., Stone, M., Zouari, L., (2007b). *Ouisper: corpus based synthesis driven by articulatory data*. In: Internat. Congress of Phonetic Sciences, Saarbrücken, Germany, pp. 2193–2196.

Hueber, T., Chollet, G., Denby, B., Dreyfus, G., Stone, M., (2007c). *Continuous-speech Phone Recognition from Ultrasound and Optical Images of the Tongue and Lips*. Interspeech, Antwerp, Belgium, pp. 658–661.

Hueber, T., Chollet, G., Denby, B., Dreyfus, G., Stone, M., (2008a). *Phone Recognition from Ultrasound and Optical Video Sequences for a Silent Speech Interface*. Interspeech, Brisbane, Australia, pp. 2032-2035.

Hueber, T., Chollet, G., Denby, B., Stone, M., (2008b). *Acquisition of ultrasound, video and acoustic speech data for a silent-speech interface application*. In: Internat. Seminar on Speech Production, Strasbourg, France, pp. 365–369.

Hueber, T., Benaroya, E.-L., Chollet, G., Denby, B., Dreyfus, G., Stone, M., (2010). *Development of a silent speech interface driven by ultrasound and optical images of the tongue and lips*. Speech Comm. 52 (4), 288–300.

Hummel, J., Figl, M., Birkfellner, W., Bax, M.R., Shahidi, R., Maurer, C.R., Bergmann, H., (2006). *Evaluation of a new electromagnetic tracking system using a standardized assessment protocol*. Phys. Med. Biol. 51, N205–N210.

IEEE, 2008. *Brain Computer Interfaces*, IEEE Computer, Vol. 41, No. 10, October 2008.

Jain A., Duin R. e Mao J. ,(2000). *Statistical Pattern Recognition: a Review*”, IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 22, No. 1, Jan. 2000.

Jorgensen, C., Lee, D.D., Agabon, S., (2003). Sub auditory speech recognition based on EMG signals. In: Proc. Internat. Joint Conf.on Neural Networks (IJCNN), pp. 3128–3133.

Jorgensen, C., Binsted, K., (2005). *Web browser control using EMG based sub vocal speech recognition*. In: Proc. 38th Annual Hawaii International Conference on System Sciences. IEEE, pp. 294c.1–294c.8.

Jorgensen, C., Dusan, S., (2010). *Speech interfaces based upon surface electromyography*. Speech Comm. 52 (4), 354–366.

Jou, S., Schultz, T., Walliczek, M., Kraft, F., (2006). *Towards continuous speech recognition using surface electromyography*. In: INTERSPEECH 2006 and 9th Internat. Conf. on Spoken Language Processing, Vol. 2, pp. 573–576.

Jou, S., Schultz, T., Waibel, A., (2007). *Multi-stream articulatory feature classifiers for surface electromyographic continuous speech recognition*. In: Internat. Conf. on Acoustics, Speech, and Signal Processing. IEEE, Honolulu, Hawaii.

Kennedy, P.R., (1989). *The cone electrode: a long-term electrode that records from neurites grown onto its recording surface*. J. Neurosci.Methods 29, 181–193.

Kennedy, P.R., (2006). *Comparing electrodes for use as cortical control signals: tiny tines, tiny wires or tiny cones on wires: which is best?* In: The Biomedical Engineering Handbook. The Electrical Engineering Handbook Series, third ed., Vol. 1. CRS/Taylor and Francis, Boca Raton.

Kennedy, P.R., Bakay, R.A.E., Moore, M.M., Adams, K., Goldwaithe, J., (2000). *Direct control of a computer from the human central nervous system*. IEEE Trans. Rehabil. Eng. 8 (2), 198–202.

Kennedy, P.R., Bakay, R.A.E., (1998). *Restoration of neural output from a paralyzed patient by direct brain connection*. NeuroReport 9, 1707–1711.

Kent R.D., (1983), *The segmental organization of speech*. in MacNielage (ed.) The production of speech Ch. 4, Springer Verlag, New York.

Kewley Port D., (1983), *Time varying features as correlates of place of articulation in stop consonants*, JASA, 73, p. 322-335.

Klatt D.H., Stevens K.N., (1973), *On the automatic recognition of continuous speech: implications from a spectrogram reading experiment*. IEEE Transactions on Audio and Electroacoustics, AU-21, p. 210-217.

Klatt, D.H., (1979), *Speech perception: a model of acoustic-phonetic analysis and lexical access*, Journal of Phonetics, 3, p. 279-312

Kim S., Simeral J.D., Hochberg L.R., Donoghue J.P., Friehs G.M., Black M.J., (2007). *Multi-state decoding of point-and-click control signals from motor cortical activity in a human with tetraplegia*. In: Neural Engineering, 2007, CNE'07 3rd Internat. IEEE/EMBS Conf., pp. 486–489.

N. Kraus, (2011). *Listening in on the listening brain*, June 2011, p. 40

<http://dx.doi.org/10.1063/1.3603917>

Levinson, S.E., 2005. *Mathematical Models for Speech Technology*. John Wiley. Liberman, A.M., Cooper, F.S., Shankweiler, D.P., and Studdert-Kennedy, M. (1967). *Perception of the speech code*. Psychol. Rev. 74, 431–461.

Liberman A.M., (1957), *Some Results of Research on Speech Perception*, JASA 29, p. 117-123.

Liberman A.M et alii, (1957). *The discrimination of speech sounds within and across phoneme boundaries*, Journal of Experimental Psychology, p. 358-368.

Liberman A.M., et alii, 1967, “*Some observations on a model for speech perception*”. In W.

Wathen-Dunn (Ed.) *Models for the Perception of Speech and Visual Form*, M.I.T Press Cambridge Massachussets, p. 68-87.

Liberman A.M., Mattingly I.G., 1985, "*The Mothor Theory of speech perception revised*", *Cognition*, 21, p. 1-36.

Liberman A.M.,(1982),*On finding that speech is special*, *American Psychologist*, 37, p.148-167.

Lindblom B., (1987), *Adaptive Variability and Absolute Constancy in Speech Signals: two themes in the quest for phonetic invariance* in *Proceedings of the 11th ICPHS, Tallin*, 1-7 Aug. 1987, vol.3, p. 9-18.

Lindblom B.,(1990), *Explaining Phonetic Variation: a Sketch of the H & H theory*, in Hardcastle W.J. e Marchal A. (ed.), *Speech Production and Speech Modelling*, Kluver Academic Press, Dordrecht, p. 403-439.

Lindblom B., (1996) *Role of artiulation in speech perception: clues from production*, *JASA*, 99, 3, p. 1683-1692.

Lively S.E., Pisoni D.B., Goldiger S.D., (1994), *Spoken Word Recognition. Research and Theory*, *Handbook of Psycholinguistics* Morton Ann (ed.) Geinsbacher Ch. 8, p. 265-300.

Lotte, F., Congedo, M., Lecuyer, A., Lamarche, F., Arnaldi, B., (2007). *A review of classification algorithms for EEG-based brain computer interfaces*. *J. Neural Eng.* 4, R1–R13.

Maeda, S., (1990). *Compensatory articulation during speech: Evidence from the analysis and synthesis of vocal-tract shapes using an articulatory model*. In: Hardcastle, W., Marchal, A. (Eds.), *Speech Production and Speech Modelling*. Kluwer Academic Publisher, Amsterdam, pp. 131–149.

Mac Neilage P., (1970), *Motor control of serial ordering of speech*, *Psychological Review* 77, p. 182-196.

Magno Caldognetto E.,(1979),*Introduzione alla percezione dei suoni linguistici*. Cleup Padova.

Magno Caldognetto E.,(1980), *La coarticolazione. Introduzione agli aspetti dinamici della produzione della parola*, CLESP, Padova.

Magno Caldognetto E, (1983) *Aspetti percettivi della coarticolazione* in RIA XIV 1990, p. 53-68.

Maier-Hein, L., Metze, F., Schultz, T., Waibel, A., 2005. *Session independent non-audible speech recognition using surface electromyography*. In: IEEE Workshop on Automatic Speech Recognition and Understanding, San Juan, Puerto Rico, pp. 331–336.

Makeig S., Debener S., Onton J., Delorme A., (2004). *Mining event-related brain dynamics*, *Trends in Cognitive Science*, 8(5):204-210

Manabe, H., Hiraiwa, A., Sugimura, T., (2003). *Unvoiced speech recognition using EMG-mime speech recognition*. In: Proc. CHI, Human Factors in Computing Systems, Ft. Lauderdale, Florida, pp. 794–795.

Manabe, H., Zhang, Z., (2004). *Multi-stream HMM for EMG-based speech recognition*. In: Proc. 26th Annual International Conf. of the IEEE Engineering in Medicine and Biology Society, 1–5 September 2004, San Francisco, California, Vol. 2, pp. 4389–4392.

Maynard, E.M., Nordhausen, C.T., Normann, R.A., (1997). *The Utah intracortical electrode array: a recording structure for potential brain–computer interfaces*. *Electroencephalogr. Clin. Neurophysiol.* 102 (3),228–239.

Marchal, A., Hardcastle, W.J., (1993). *Instrumentation and database for the cross-language study of coarticulation*. In: *Lang. Speech* 36 (1), 3–20.

Marslen-Wilson W.D., (1980). *Speech understanding as a psychological process*, in Simon J., *Spoken language generation and understanding*, Reidel, p. 36-67.

Marslen-Wilson W.D., (1985), *Speed shadowing and speech comprehension*. Speech communication, 4, p. 55-73.

Marslen-Wilson W.D., Tyler L., (1975), *Processing structure of spoken language understanding*, Cognition, 8, p. 1-71.

Massaro D.W., (1972), *Perceptual images, processing time, and perceptual units in auditory perception*, Psychological Review, 79, p. 124-145.

Massaro D.W., Oden G.C., 1980, "Speech perception: a framework for research and theory", in Lass, N.J., (ed.) Speech and language. Advances in basic research and practice. Vol.3, New York Academic press, p. 129-165.

Massaro D.W., (1994), *Psychological aspects of speech perception*, in Handbook of Psycholinguistics Morton Ann Geinsbacher (ed.), Academic Press Ch.7, p. 219-265.

Mclelland J.L., Elman J.L., (1986), *The TRACE model of speech perception*. Cognitive Psychology, 23, p. 1-44.

Miller, L.E., Andreasen, D.S., Bartels, J.L., Kennedy, P.R., Robesco, J., Siebert, S.A., Wright, E.J., (2007). *Human speech cortex long-term recordings: Bayesian analyses*. In: Neuroscience Meeting Planner 2007, Program No. 517.20, San Diego, USA.

Moeller, (2008). <<http://innovationdays.alcatel-lucent.com/2008/documents/Talking.Beyond.Hearing.pdf>>.

Moore B.J., (1982), *An introduction to the Psychology of Hearing*, Academic Press, London.

Moro A. (2006), *I Confini di Babele – il cervello e il mistero delle lingue impossibili –* Longanesi editore

Morse, M.S., O'Brien, E.M., (1986). *Research summary of a scheme to ascertain the availability of speech information in the myoelectric signals of neck and head muscles*

using surface electrodes. Comput.Biol. Med. 16 (6), 399–410.

Morse, M.S., Day, S.H., Trull, B., Morse, H., (1989). *Use of myoelectric signals to recognize speech*. In: Images of the Twenty-First Century –Proc. 11th Annual Internat. Conf. of the IEEE Engineering in Medicine and Biology Society, Part 2, Vol. 11. Alliance for Engineering in Medicine and Biology, pp. 1793–1794.

Morse, M.S., Gopalan, Y.N., Wright, M., (1991). *Speech recognition using myoelectric signals with neural networks*. In: Proc. 13th Annual Internat. Conf. of the IEEE Engineering in Medicine and Biology Society, Vol. 13, No 4, Piscataway, NJ, United States. IEEE, pp. 1877–1878.

Morton J.,(1979),*Word recognition*, in Morton J., Marshall J., Psycholinguistics 2: Structures and processes, MIT press, Cambridge MA p. 107-156.

Munhall, K.G., Vatikiotis-Bateson, E., Tohkura, Y., (1995). *X-ray film database for speech research*. J. Acoust. Soc. Amer. 98, 1222–1224.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). *Early selective-attention effect on evoked potential reinterpreted*. Acta Psychologica, 42, 313-329.

Nakajima, Y., (2005). *Development and evaluation of soft silicone NAM microphone*. Technical Report IEICE, SP2005-7, pp. 7–12

Nakajima, Y., Kashioka, H., Shikano, K., Campbell, N., (2003a). *Nonaudible murmur recognition*. In: Proc. Eurospeech 2003, pp. 2601–2604.

Nakajima, Y., Kashioka, H., Shikano, K., Campbell, N., (2003b). *Nonaudible murmur recognition input interface using stethoscopic microphone attached to the skin*. In: Proc. IEEE ICASSP, pp. 708–711.

Nakajima, Y., Kashioka, H., Campbell, N., Shikano, K., (2006). *Non audible murmur (NAM) recognition*. IEICE Trans. Inform. Systems E89-D (1), 1–8.

Nakamura, H., (1988). *Method of recognizing speech using a lip image*. United States Patent 4769845, September 06.

NessAiver, M.S., Stone, M., Parthasarathy, V., Kahana, Y., Paritsky, A., (2006). *Recording high quality speech during tagged cine-MRI studies using a fiber optic microphone*. J. Magnet. Reson. Imag. 23, 92–97.

Neuper, C., Muñller, G.R., Kuñbler, A., Birbaumer, N., Pfurtscheller, G., (2003). *Clinical application of an EEG-based brain computer interface: a case study in a patient with severe motor impairment*. Clin. Neurophysiol. 114, 399–409.

Nguyen, L., Burnett, G., Holzrichter, J., Gable, T., (2000). *Denoising of human speech using combined acoustic and EM sensor signal processing*. In: Internat. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Istanbul, Turkey, 5–9 June 2000, Vol. 1, pp. 229–232.

Nguyen, N., Marchal, A., Content, A., (1996). *Modeling tongue-palate contact patterns in the production of speech*. J. Phonetics 1996, 77–98.

Nijholt, A., Tan, D., Pfurtscheller, G., Brunner, C., Millañn, J.d.R., Allison, B., Graitmann, B., Popescu, F., Blankertz, B., Muñller, K.R., (2008). *Brain–Computer Interfacing for Intelligent Systems*. Intell. Systems 23 (3), 72–79.

Otani, M., Shimizu, S., Hirahara, T., (2008). *Vocal tract shapes of non-audible murmur production*. Acoust. Sci. Technol. 29, 195–198.

O’Shaughnessy D., (1987), Speech communication, Human and Machine, cap. 9 “*Speech Synthesis*”, Addison-Wesley publ., p. 380-412.

O’Shaughnessy D., (1987), Speech communication, Human and Machine, cap.4 “*Hearing*”, Addison-Wesley publ., p. 128-164.

Ouisper, (2006). *Oral Ultrasound synthetic SpEech SouRce*, “*Projet Blanc*”, National

Research Agency (ANR), France, Contract No. ANR-06-BLAN-0166, 2006–2009.

Papeo L., Vallessi A., Isaja A., Rumiati R. I., (2010). *Effects of TMS on Different Stages of Motor and Non-Motor Verb Processing in the Primary Motor Cortex*,

Patil, S.A., Hansen, J.H.L., (2010). *A competitive alternative for speaker assessment: physiological Microphone (PMIC)*. *Speech Comm.* 52 (4), 327–340.

Paoloni A., Cerrato L., 1996, *Sulla trascrizione di intercettazioni ambientali*, RIA vol.20, n.4, p. 159-163.

Perkell, J., Cohen, M., Svirsky, M., Matthies, M., Garabieta, I., Jackson, M., (1992.) *Electro-magnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements*. *J. Acoust. Soc. Amer.* 92, 3078–3096.

Petajan, E.D., (1984). *Automatic lipreading to enhance speech recognition*. In: IEEE Communications Society Global Telecommunications Conf., Atlanta, USA.

Pisoni D.B., (1985), *Speech Perception: some new directions in research and theory*, in *JASA*, 78 (1) p. 381-388.

Pisoni D.B., Luce P.A, (1986), *Speech Perception: Research, Theory and the Principal Issues*”, in *Pattern Recognition By Humans and Machine: Speech Perception Vol. 1*, p. 11-50.

Pisoni D.B., Carrell, T.D., Gans S.J., (1983), *Perception of the duration of rapid spectrum changes in speech and non-speech signals*. *Perception and Psychophysics*, v.34, p. 314-322.

Pisoni D.B., Sawush J.R., (1975), *Some stages of processing in speech perception*, in Cohen A., e Noteboom S. (eds.) *Structure and process in speech perception* Springer-Verlag Heidelberg , p. 16-34.

Porbadnigk, A., Wester, M., Calliess, J., Schultz, T., (2009). *EEG-based speech recognition – impact of temporal effects*. In: Biosignals 2009, Porto, Portugal, January 2009, pp. 376–381.

Preuss, R.D., Fabbri, D.R., Cruthirds, D.R., (2006). *Noise robust vocoding at 2400 bps*. In: 8th Internat. Conf. on Signal Processing, ICSP 2006, Guilin, China, November 16–20, 2006, Vol. 1, pp. 16–20.

Quatieri, T.F., Messing, D., Brady, K., Campbell, W.B., Campbell, J.P., Brandstein, M., Weinstein, C.J., Tardelli, J.D., Gatewood, P.D., (2006). *Exploiting nonacoustic sensors for speech enhancement*. IEEE Trans. Audio Speech Lang. Process. 14 (2), 533–544.

Repp B.H., (1987), *The role of psychophysics in understanding speech perception*, in Schouten M.E.H., (ed) *The Psychophysics of speech perception*, NATO-SED Series, Nijhoff Publisher, Dordrecht p. 3-27.

Repp B.H.,(1984), *Categorical perception. Issues, methods, findings*, in Lass N.J., (ed.) *Speech and language. Advances in basic research and practice*. Vol.10 New York Academic press, p. 243-271.

Repp B.H. (1982), *Phonetic trading relations and context effects: new experimental evidence for a speech mode of perception*. Psychological Bulletin, 92, p. 81-110.

Rothenberg, M., (1992). *A multichannel electroglottograph*. Journ. Voice 6 (1), 36–43.

Rubin, P., Vatikiotis-Bateson, E., (1998). *Talking heads*. Proc. Audio Visual Speech Process. 1998, 233–238.

Sajda, P., Mueller, K.-R., Shenoy, K.V. (Eds.), (2008). *Brain Computer Interfaces*. IEEE Signal Process. Mag. (special issue).

Salmelin R. (2007) *Clinical Neurophysiology of Language: The MEG Approach in Clinical Neurophysiology*, 118, pp. 237-54

G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J.R. Wolpaw, (2005), *BCI2000: A General-Purpose Brain-Computer Interface (BCI) System*, IEEE Transactions on biomedical engineering, vol. 51, no. 6, June 2004

Schoenle, P.W., Graebe, K., Wenig, P., Höhne, J., Schrader, J., Conrad, B., (1987). *Electromagnetic articulography: Use of alternating magnetic fields for tracking movements of multiple points inside and outside the vocal tract*. Brain Lang. 31, 26–35.

Shokey L., Reddy R., (1974), *Quantitative analysis of speech perception: results from transcription of connected speech from unfamiliar languages*. Speech Communication Seminar August 1974, Stockholm.

Schroeter, J., Ostermann, J., Graf, H.P., Beutnagel, M., Cosatto, E., Syrdal, A., Conkie, A., Stylianou, Y., (2000). *Multimodal speech synthesis*. In: IEEE Internat. Conf. on Multimedia and Expo 2000, pp.571–574.

Schultz, T., Wand, M., (2010). *Modeling coarticulation in large vocabulary EMG-based speech recognition*. Speech Comm. 52 (4), 341–353.

Sophie Scott, K., Rosen, Stuart; Beaman, C. Philip; Davis, Josh P.; Wise, Richard J. S., (2009), *The neural processing of masked speech: Evidence for different mechanisms in the left and right temporal lobes* The Journal of the Acoustical Society of America, vol. 125, issue 3, p. 1737

Siebert, S.A., Velliste, M., (2007). *Human speech cortex long-term recordings: neural net analyses*. In: Neuroscience Meeting Planner 2007, Program No. 517.18, San Diego, USA.

Soros P., Kornelissen K., Laine M., Salmelin R., (2003) *Naming Actions and Objects: Cortical Dynamics in Healthy Adults and in Anomic Patient with a Dissociation in Action/ Object Naming*, in NeuroImage, 2003, 19, pp. 1787-8018

Stevens K.N., House A.S., (1967), *Remarks on Analysis by Synthesis and Distinctive Features*, in Wathen-Dunn W (ed.) Models for Perception of Speech and Visual Form, Cambridge Massachusetts, p. 88-102.

Stevens K.N., House A.S.,(1972), "*Speech Perception*", in Tobias J.V (ed.) *Foundations of Modern Auditory Theory* New York, vol.2, p. 3-62.

Stevens S.S., Davis H., (1983), *Hearing, its psychology and physiology*, American Institute of Physics New York.

Streeter L., et alii, (1983), *Acoustic and perceptual indicators of emotional stress*, JASA 73, p. 1354-1360.

Studdert-Kennedy M., (1970), *The perception of speech*, Haskins labs Status Report on speech research, 23, p. 15-48.

Summerfield Q., (1987), *Some preliminaries to a comprehensive Account of Audio-Visual Speech perception*, in *Hearing by eye: the psychology of lip-reading* ch. 1, Dodd B., Campbell R., (eds) Lawrence Erlbaum Ass. Publ. London, p. 3-51.

Syrdal A.K.,() *Perception of consonantal place of articulation*, in *Speech and Language: advances in Basic research and practice*, vol.9, p. 313-349.

Stone, M., Sonies, B., Shawker, T., Weiss, G., Nadel, L., (1983). *Analysis of real-time ultrasound images of tongue configuration using a griddigitizing system*. J. Phonetics 11, 207–218.

Stone, M., Shawker, T., (1986). *An ultrasound examination of tongue movement during swallowing*. Dysphagia 1, 78–83.

Stone, M., Davis, E., (1995). *A head and transducer support (HATS) system for use in ultrasound imaging of the tongue during speech*. J. Acoust. Soc. Amer. 98, 3107–3112.

Stone, M., (2005). *A guide to analyzing tongue motion from ultrasound images*. Clin. Linguist. Phonet. 19 (6–7), 455–502.

Sugie, N., Tsunoda, K., (1985). *A speech prosthesis employing a speech synthesizer-vowel discrimination from perioral muscle activities and vowel production*. IEEE Trans.

Biomed. Eng. BME-32 (7), 485–490. Suppes, P., Lu, Z.-L., Han, B., 1997. Brain wave recognition of words.

Proc. Nat. Acad. Sci. USA 94, 14965–14969.

Tardelli, J.D. (Ed.), (2003). Pilot Corpus for Multisensor Speech Processing. MIT Lincoln Labs Report ESC-TR-2004-084.

Tatham, M., (1971). *The place of electromyography in speech research*. Behav. Technol. 6.

TERC, 2009. <<http://spinlab.wpi.edu/projects/terc/terc.html>>.

Titze, I.R., Story, B.H., Burnett, G.C., Holzrichter, J.F.N.G.L.C., Lea, W.A., (2000). *Comparison between electroglottography and electromagnetic glottography*. J. Acoust. Soc. Amer. 107 (1), 581–588, January.

Tran, V.-A., Bailly, G., Loevenbruck, H., Jutten C., (2008a). *Improvement to a NAM captured whisper-to-speech system*. In: Interspeech 2008, Brisbane, Australia, pp. 1465–1498.

Tran, V.-A., Bailly, G., Loevenbruck, H., Toda, T., (2008b). *Predicting F0 and voicing from NAM-captured whispered speech*. In: Proc. Speech Prosody, Campinas, Brazil.

Tran, V.-A., Bailly, G., Loevenbruck, H., Toda, T., (2010). *Improvement to a NAM-captured whisper-to-speech system*. Speech Comm. 52 (4), 314–326.

Truccolo, W., Friehs, G.M., Donoghue, J.P., Hochberg, L.R., (2008). *Primary motor cortex tuning to intended movement kinematics in humans with tetraplegia*. J. Neurosci. 28 (5), 1163–1178.

Walliczek, M., Kraft, F., Jou, S.-C., Schultz, T., Waibel, A., (2006). *Sub-word unit based non-audible speech recognition using surface electromyography*. In: Proc. Interspeech, Pittsburgh, USA, pp. 1487–1490.

Wand, M., Schultz, T., (2009). *Speaker-adaptive speech recognition based on surface electromyography*. In: BIOSTEC – BIOSIGNALS 2009 best papers. Communications in

Computer and Information Science (CCIS) series. Springer, Heidelberg, 2009.

S. Watanabe, *Pattern Recognition: Human and Mechanical*, Wiley, New York 1985.

Wester, M., Schultz, T., (2006). *Unspoken speech – speech recognition based on electroencephalography*. Master's Thesis, Universität Karlsruhe (TH), Karlsruhe, Germany.

Wolpaw, J.R., Birbaumer, N., McFarland, D., Pfurtscheller, G., Vaughan, T., (2002). *Brain-computer interfaces for communication and control*. *Clin. Neurophysiol.* 113 (6), 767–791.

Wrench, A.A., Scobbie, J.M., (2003). *Categorising vocalisation of English /l/ using EPG, EMA and ultrasound*. In: 6th Internat. Seminar on Speech Production, Manly, Sydney, Australia, 7–10 December 2003, pp. 314–319.

Wrench, A., Scobbie, J., Linden, M., (2007). *Evaluation of a helmet to hold an ultrasound probe*. In: *Ultrafest IV*, New York, USA.

Wright, E.J., Andreasen, D.S., Bartels, J.L., Brumberg, J.S., Guenther, F.H., Kennedy, P.R., Miller, L.E., Robesco, J., Schwartz, A.B., Van Son R.J., Pols L.C., (1996), “*A comparison between the acoustics of vowel and consonant reduction*”, in *Proceeding of the Institute of Phonetic Sciences University of Amsterdam*, 20, p. 13-25.

Yost W., Nielsen W., (1986), *Le basi della funzione uditiva*, trad.italiana a c. di Bernabei L., Cipolla G., Piccin Nuova Libreria Padova.

Zien A., Rätsch G., Mika S., Schölkopf B., Lemmen C., Smola A., Lengauer T., Müller K.-R., (2000) *Engineering Support Vector Machine Kernels That Recognize Translation Initiation Sites*, in *silico Biology GCB '99 German Conference on Bioinformatics*, October 4 - 6, 1999 Hannover, Germany <http://www.bioinfo.de/isb/gcb99/talks/zien/>

8. APPENDICE

Qui di seguito sono riportati i punti dell'elaborazione dei segnali

Punto 1

```
load 1_1.mat;
Ch1train=Ch1(1:750,1:25);
Ch2train=Ch2(1:750,1:25);
Ch3train=Ch3(1:750,1:25);
Ch4train=Ch4(1:750,1:25);
Ch1test=Ch1(1:750,26:50);
Ch2test=Ch2(1:750,26:50);
Ch3test=Ch3(1:750,26:50);
Ch4test=Ch4(1:750,26:50);
```

Punto 2

```
i=1;
k=1;
for i=1:750;
for k=1:23;
riga=Ch1train;
A=riga(i,k);
x=k+1;
B=riga(i,x);
y=k+2;
C=riga(i,y);
PARZIALE=(A+B+C)/3;
w1(i,k)=PARZIALE
end
```

Punto 3

```
riga=Ch2train;
riga=Ch3train;
riga=Ch4train;
riga=Ch1test;
riga=Ch2test;
riga=Ch3test;
riga=Ch4test;
```

Punto 4

w_1, w_2, w_3, w_4 = sono le matrici 750×23 relative al TRAIN del caso "nulla".

a_1, a_2, a_3, a_4 = sono le matrici 750×23 relative al TRAIN del caso "a".

i_1, i_2, i_3, i_4 = sono le matrici 750×23 relative al TRAIN del caso "i".

u_1, u_2, u_3, u_4 = sono le matrici 750×23 relative al TRAIN del caso "u".

$w_{1t}, w_{2t}, w_{3t}, w_{4t}$ = sono le matrici 750×23 relative al TEST del caso "nulla".

$a_{1t}, a_{2t}, a_{3t}, a_{4t}$ = sono le matrici 750×23 relative al TEST del caso "a".

$i_{1t}, i_{2t}, i_{3t}, i_{4t}$ = sono le matrici 750×23 relative al TEST del caso "i".

$u_{1t}, u_{2t}, u_{3t}, u_{4t}$ = sono le matrici 750×23 relative al TEST del caso "u".

Punto 5

TRAIN:

$w_{1freq}, w_{2freq}, w_{3freq}, w_{4freq}$ = sono le matrici 12×23 relative al passaggio in frequenza di w_1, w_2, w_3, w_4 ;

$a_{1freq}, a_{2freq}, a_{3freq}, a_{4freq}$ = sono le matrici 12×23 relative al passaggio in frequenza di a_1, a_2, a_3, a_4 ;

$i_{1freq}, i_{2freq}, i_{3freq}, i_{4freq}$ = sono le matrici 12×23 relative al passaggio in frequenza di i_1, i_2, i_3, i_4 ;

$u_{1freq}, u_{2freq}, u_{3freq}, u_{4freq}$ = sono le matrici 12×23 relative al passaggio in frequenza di u_1, u_2, u_3, u_4 ;

TEST:

$w_{1tfreq}, w_{2tfreq}, w_{3tfreq}, w_{4tfreq}$ = sono le matrici 12×23 relative al passaggio in frequenza di $w_{1t}, w_{2t}, w_{3t}, w_{4t}$;

$a_{1tfreq}, a_{2tfreq}, a_{3tfreq}, a_{4tfreq}$ = sono le matrici 12×23 relative al passaggio in frequenza di $a_{1t}, a_{2t}, a_{3t}, a_{4t}$;

$i_{1tfreq}, i_{2tfreq}, i_{3tfreq}, i_{4tfreq}$ = sono le matrici 12×23 relative al passaggio in frequenza di $i_{1t}, i_{2t}, i_{3t}, i_{4t}$;

$u_{1tfreq}, u_{2tfreq}, u_{3tfreq}, u_{4tfreq}$ = sono le matrici 12×23 relative al passaggio in frequenza di $u_{1t}, u_{2t}, u_{3t}, u_{4t}$;

Tabella A

Procedimento seguito per la costruzione delle matrici di *train* e *test*

file	channels	traintest	sottomatrice
1_1m/nulla/	ch1	ch1train	W1
		ch1test	w1t
	ch2	ch2train	w2
		ch2test	w2t
	ch3	ch3train	w3
		ch3test	w3t
	ch4	ch4train	w4
		ch4test	w4t
1-2m/a/	ch1	ch1train	a1
		ch1test	a1t
	ch2	ch2train	a2
		ch2test	a2t
	ch3	ch3train	a3
		ch3test	a3t
	ch4	ch4train	a4
		ch4test	a4t
1_3m/i/	ch1	ch1train	i1
		ch1test	i1t
	ch2	ch2train	i2
		ch2test	i2t
	ch3	ch3train	i3
		ch3test	i3t
	ch4	ch4train	i4
		ch4test	i4t
1-4m/u/	ch1	ch1train	u1
		ch1test	u1t
	ch2	ch2train	u2
		ch2test	u2t
	ch3	ch3train	u3
		ch3test	u3t
	ch4	ch4train	u4
		ch4test	u4t

Tabella B

Matrice nel tempo 750x23	LF FUNCTION	Matrice in frequenza
train		
w1	⇒	w1freq
w2	⇒	w2freq
w3	⇒	w3freq
w4	⇒	w4freq
test		
w1t	⇒	w1tfreq
w2t	⇒	w2tfreq
w3t	⇒	w3tfreq
w4t	⇒	w4tfreq

Tabella C

Sottomatrici ausiliarie TRAIN-TEST

	Nulla	a	i	u
12	w1freq	a1freq	i1freq	u1freq
12	w2freq	a2freq	i2freq	u2freq
12	w3freq	a3freq	i3freq	u3freq
12	w4freq	a4freq	i4freq	u1freq
	23	23	23	23
	nullat	at	it	ut
12	w1tfreq	a1tfreq	i1tfreq	u1tfreq
12	w2tfreq	a2tfreq	i2tfreq	u2tfreq
12	w3tfreq	a3tfreq	i3tfreq	u3tfreq
12	w4tfreq	a4tfreq	i4tfreq	u1tfreq
	23	23	23	23

Punto 6

```
>>nulla=[w1freq;w2freq;w3freq;w4freq];
>>a=[a1freq;a2freq;a3freq;a4freq];
>>i=[i1freq;i2freq;i3freq;i4freq];
>>u=[u1freq;u2freq;u3freq;u4freq];
>>>nullat=[w1tfreq;w2tfreq;w3tfreq;w4tfreq];
>>at=[a1tfreq;a2tfreq;a3tfreq;a4tfreq];
>>it=[i1tfreq;i2tfreq;i3tfreq;i4tfreq];
>>ut=[u1tfreq;u2tfreq;u3tfreq;u4tfreq];
```

Punto 7

```
>>in_train=[nulla,a,i,u];
>>in_train=in_train';
>>in_val=[nullat,at,it,ut];
>>in_val=in_val';
```

Punto 8

```
out_train_nulla=zeros(4,23);
out_train_nulla=[1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1];
out_train_a=zeros(4,23);
out_train_a=[-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1];
out_train_i=zeros(4,23);
out_train_i=[-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1];
out_train_u=zeros(4,23);
out_train_u=[-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1];
```

```

-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1,-1;
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1];
out_train_lin=[out_train_nulla,out_train_a,out_train_i,out_train_u];
out_train=out_train_lin'

```

Tabella D

Matrici out_train e out_val

nulla	a	i	u
1111	-1-1-1-1...	-1-1-1-1...	-1-1-1-1...
-1-1-1-1...	1111	-1-1-1-1...	-1-1-1-1...
-1-1-1-1...	-1-1-1-1...	1111	-1-1-1-1...
-1-1-1-1...	-1-1-1-1...	-1-1-1-1...	1111

Punto 9

```

for k=1:17;
riga=Ch1test;
A=riga(i,k);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k
x=k+1;
B=riga(i,x);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+1
y=k+2;

C=riga(i,y);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+2
b=k+3;
D=riga(i,b);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+3
c=k+4;
E=riga(i,c);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+4
d=k+5;
F=riga(i,d);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+5
e=k+6;
G=riga(i,e);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+6
f=k+7;
H=riga(i,f);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+7
g=k+8;
I=riga(i,g);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+8

```

```

PARZIALE=(A+B+C+D+E+F+G+H+I)/9;
w1(i,k)=PARZIALE;
for k=1:17; x
riga=Ch1test;
A=riga(i,k);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k
x=k+1;
B=riga(i,x);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+1
y=k+2;

C=riga(i,y);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+2
b=k+3;
D=riga(i,b);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+3
c=k+4;
E=riga(i,c);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+4
d=k+5;
F=riga(i,d);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+5
e=k+6;
G=riga(i,e);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+6
f=k+7;
H=riga(i,f);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+7
g=k+8;
I=riga(i,g);%SELEZIONO L'ELEMENTO DI RIGA i E COLONNA k+8
PARZIALE=(A+B+C+D+E+F+G+H+I)/9;
w1(i,k)=PARZIALE;

```

Punto 10

```

>>[Ea, Ei]=salla('5_4', '5_1')
>>Eatrain=Ea(1:504,1:25)
>>Eitrain=Ei(1:504,1:25)
>>Eatest=Ea(1:504,26:50)
>>Eitest=Ei(1:504,26:50)
>>in_train=[ Eatrain, Eitrain]
>>in_train=in_train'
>>in_val=[ Eatest, Eitest]
>>in_val=in_val'

```

Punto 11

```
[L,ind] = sort(diag(L),'descend');  
Vc = Vc(:,ind);  
W=sqrt(inv(diag(L))) * Vc';
```

Punto 12

```
Ch1=LF(Ch1);  
Ch2=LF(Ch2);  
Ch3=LF(Ch3);  
Ch4=LF(Ch4);  
Ea=[Ch1;Ch2;Ch3;Ch4];  
Ch1=LF(Ch1);  
Ch2=LF(Ch2);  
Ch3=LF(Ch3);  
Ch4=LF(Ch4);  
Ei=[Ch1;Ch2;Ch3;Ch4];
```

Algoritmo 1 mixture

```
s1=ALLEEG(1,1).data;  
s1(5:7,,:)=[]; % C3 C4 Fz Cz Fcz C5 C6  
  
s2=ALLEEG(1,2).data;  
s2(5:7,,:)=[]; % C3 C4 Fz Cz Fcz C5 C6  
  
% SCELTA TEMPO  
  
ti=1; % ms  
tf=1000; % ms  
  
sample1=fix(ti*250/1000)+1;  
sample2=fix(tf*250/1000);  
  
s1=s1(:,sample1:sample2,:);  
s2=s2(:,sample1:sample2,:);  
  
% SCELTA CANALI  
  
ch=[1 2 4 5]; % C3 C4 Fz Cz Fcz C5 C6  
% ch=1:64;  
  
sf1=[];
```

```

sf2=[];

for i=1:length(ch)

    sf1=[sf1; s1(ch(i),:,:)];
    sf2=[sf2; s2(ch(i),:,:)];

end

[sf1_csp, sf2_csp] = cspp(sf1,sf2);

for i=1:100

    i

    DATA1=reshape(sf1_csp,size(sf1_csp,1)*size(sf1_csp,2),size(sf1_csp,3));
    DATA2=reshape(sf2_csp,size(sf2_csp,1)*size(sf2_csp,2),size(sf2_csp,3));

    [DATA_M1_train, DATA_M1_test, DATA_M2_train,
    DATA_M2_test]=mean_mixture(DATA1,DATA2);

    DATA_train=double([DATA_M1_train DATA_M2_train]');
    actual_train=vertcat(ones(size(DATA_M1_train,2),1),-1*ones(size(DATA_M2_train,2),1));

    DATA_test=double([DATA_M1_test DATA_M2_test]');
    actual_test=vertcat(ones(size(DATA_M1_test,2),1),-1*ones(size(DATA_M2_test,2),1));

    M_a(i) = s_v_m(DATA_train, DATA_test, actual_train, actual_test);

end

M_acc=mean(M_a)
SD_acc=std(M_a)

CSP
function [Za, Zu] = cspp(a,u)

Ca=zeros(size(a,1),size(a,1));
Cu=zeros(size(u,1),size(u,1));

for i=1:size(a,3)
    Ca=Ca+(a(:, :, i)*a(:, :, i)')/trace(a(:, :, i)*a(:, :, i)');
end
Ca=Ca/size(a,3);

for i=1:size(u,3)

```

```

    Cu=Cu+(u(:, :, i)*u(:, :, i)')/trace(u(:, :, i)*u(:, :, i)');
end
Cu=Cu/size(u,3);

Cc=Ca+Cu;

[Vc,LAMBDAc] = eig(Cc);
[LAMBDAc,ind] = sort(diag(LAMBDAc),'descend');
Vc = Vc(:,ind);

W = sqrt(inv(diag(LAMBDAc))) * Vc';

Sa=W*Ca*W';
% Su=W*Cu*W';

[Ua,LAMBDAA] = eig(Sa);
% [Uu,LAMBDAAu] = eig(Su);

[~,ind] = sort(diag(LAMBDAA)); Ua = Ua(:,ind);

P = (Ua'*W)';

for i=1:size(a,3)
    Za(:, :, i)=P*a(:, :, i);
end

for i=1:size(u,3)
    Zu(:, :, i)=P*u(:, :, i);
end

```

La funzione per la media mobile mean-mixture è:

```

function [DATA_M1_train, DATA_M1_test, DATA_M2_train,
DATA_M2_test]=(DATA1,DATA2)

```

```

DATA_M1_train=[];
DATA_M2_train=[];
DATA_M1_test=[];
DATA_M2_test=[];

```

```

DATA1=DATA1(:,randperm(end));
DATA2=DATA2(:,randperm(end));

```

```

DATA1_1=DATA1(:,1:ceil(75/100*(size(DATA1,2))));
DATA1_2=DATA1(:,ceil(75/100*(size(DATA1,2)))+1:end);

```

```

for i=1:size(DATA1,2)

```

```

    DATA1_1_random=DATA1_1(:,randperm(end));
    DATA_M1_train=[DATA_M1_train

```

```
mean(DATA1_1_random(:,1:ceil(50/100*(size(DATA1_1_random,2))))),2);
```

```
end
```

```
for i=1:size(DATA1,2)
```

```
DATA1_2_random=DATA1_2(:,randperm(end));
```

```
DATA_M1_test=[DATA_M1_test  
mean(DATA1_2_random(:,1:ceil(50/100*(size(DATA1_2_random,2))))),2);
```

```
end
```

```
DATA2_1=DATA2(:,1:ceil(75/100*(size(DATA2,2))));
```

```
DATA2_2=DATA2(:,ceil(75/100*(size(DATA2,2))+1:end));
```

```
for i=1:size(DATA2,2)
```

```
DATA2_1_random=DATA2_1(:,randperm(end));
```

```
DATA_M2_train=[DATA_M2_train  
mean(DATA2_1_random(:,1:ceil(50/100*(size(DATA2_1_random,2))))),2);
```

```
end
```

```
for i=1:size(DATA2,2)
```

```
DATA2_2_random=DATA2_2(:,randperm(end));
```

```
DATA_M2_test=[DATA_M2_test  
mean(DATA2_2_random(:,1:ceil(50/100*(size(DATA2_2_random,2))))),2);
```

```
end
```

L'SVM

```
function [M_acc, SD_acc] = s_v_m(DATA_train, DATA_test, actual_train, actual_test)
```

```
for i=1:1
```

```
% iter=i
```

```
% DATA_random=DATA(randperm(end),:);
```

```
%
```

```
%
```

```
% train=DATA_random(1:ceil(75/100*(size(DATA,1))),2:end);
```

```
% actual_train = DATA_random(1:ceil(75/100*(size(DATA,1))),1:1);
```

```
%
```

```
% test=DATA_random(ceil(75/100*(size(DATA,1))+1:size(DATA,1),2:end);
```

```
% actual_test = DATA_random(ceil(75/100*(size(DATA,1))+1:size(DATA,1),1:1);
```

```

SVMStruct = svmtrain(DATA_train,actual_train,...
    'method', 'SMO','kernel_function','quadratic','showplot',0,'boxconstraint',1); % C
predict = svmclassify(SVMStruct,DATA_test,'showplot',0);
[C,order]=confusionmat(actual_test,predict);
Acc(i)=sum(diag(C))/(sum(diag(C))+C(1,2)+C(2,1));
end

M_acc=100*mean(Acc);
SD_acc=100*std(Acc);

```

Il secondo algoritmo

```

s1=ALLEEG(1,1).data;
s1(6:8,:)=[]; % C3 C4 Fz Cz Fcz C5 C6

s2=ALLEEG(1,2).data;
s2(6:8,:)=[]; % C3 C4 Fz Cz Fcz C5 C6

% SCELTA TEMPO

ti=1000;% ms
tf=1070;% ms

sample1=fix(ti*250/1000)+1;
sample2=fix(tf*250/1000);

s1=s1(:,sample1:sample2,:);
s2=s2(:,sample1:sample2,:);

% SCELTA CANALI

ch=[1 2 3 4 5 ];% C3 C4 Fz Cz Fcz C5 C6

sf1=[];
sf2=[];

for i=1:length(ch)

    sf1=[sf1; s1(ch(i),:,:)];
    sf2=[sf2; s2(ch(i),:,:)];

end

[M_acc, SD_acc] = dasalla(sf1,sf2)

```



```
s1=ALLEEG(1,1).data;
s1(6:8,:,:)=[]; % C3 C4 Fz Cz Fcz C5 C6
```

```
s2=ALLEEG(1,2).data;
s2(6:8,:,:)=[]; % C3 C4 Fz Cz Fcz C5 C6
```

```
% SCELTA TEMPO
```

```
ti=1000;% ms
tf=1200;% ms
```

```
sample1=fix(ti*250/1000)+1;
sample2=fix(tf*250/1000);
```

```
s1=s1(:,sample1:sample2,:);
s2=s2(:,sample1:sample2,:);
```

```
% SCELTA CANALI
```

```
ch=[1 2 3 4 5];% C3 C4 Fz Cz Fcz C5 C6
```

```
sf1=[];
sf2=[];
```

```
for i=1:length(ch)
```

```
    sf1=[sf1; s1(ch(i),:,:)];
    sf2=[sf2; s2(ch(i),:,:)];
end
```

```
[M_acc, SD_acc] = dasalla(sf1,sf2)
```

La funzione *dasalla.m*

```
function [M_acc, SD_acc] = dasalla(a,u)
```

```
Ca=zeros(size(a,1),size(a,1));
Cu=zeros(size(u,1),size(u,1));
```

```
for i=1:size(a,3)
    Ca=Ca+(a(:,:,i)*a(:,:,i)')/trace(a(:,:,i)*a(:,:,i)');
end
Ca=Ca/size(a,3);
```

```
for i=1:size(u,3)
    Cu=Cu+(u(:,:,i)*u(:,:,i)')/trace(u(:,:,i)*u(:,:,i)');
end
Cu=Cu/size(u,3);
Cc=Ca+Cu;
```

```
[Vc,LAMBDAC] = eig(Cc);
```

```

[LAMBDaC,ind] = sort(diag(LAMBDaC),'descend');
Vc = Vc(:,ind);

W = sqrt(inv(diag(LAMBDaC))) * Vc';

Sa=W*Ca*W';
% Su=W*Cu*W';

[Ua,LAMBDaA] = eig(Sa);
% [Uu,LAMBDaU] = eig(Su);

[~,ind] = sort(diag(LAMBDaA)); Ua = Ua(:,ind);

P = (Ua'*W)';

for i=1:size(a,3)
    Za(:,i)=P*a(:,i);
end

for i=1:size(u,3)
    Zu(:,i)=P*u(:,i);
end

DATAa=reshape(Za,size(a,1)*size(a,2),size(a,3));
DATAu=reshape(Zu,size(u,1)*size(u,2),size(u,3));
DATA=[DATAa DATAu]';
actual_data=vertcat(ones(size(a,3),1),-1*ones(size(u,3),1));
DATA=double([actual_data DATA]);
for i=1:100
display(['Iter' '' num2str(i)]);
DATA_random=DATA(randperm(end),:);
train=DATA_random(1:fix(75/100*(size(a,3)+size(u,3))),2:end);
actual_train = DATA_random(1:fix(75/100*(size(a,3)+size(u,3))),1,1);
test=DATA_random(fix(75/100*(size(a,3)+size(u,3)))+1:size(a,3)+size(u,3),2:end);
actual_test = DATA_random(fix(75/100*(size(a,3)+size(u,3)))+1:size(a,3)+size(u,3),1,1);
SVMStruct = svmtrain(train,actual_train,...
    'method', 'SMO','kernel_function','quadratic','showplot',0,'boxconstraint',1); % C

predict = svmclassify(SVMStruct,test,'showplot',0);

[C,order]=confusionmat(actual_test,predict);
Acc(i)=sum(diag(C))/(sum(diag(C))+C(1,2)+C(2,1));
end

M_acc=100*mean(Acc);
SD_acc=100*std(Acc);

```

Tabella dei risultati

Qui di seguito riportiamo i risultati in percentuale e la deviazione standard (DS) inseriti nelle tabelle per ciascuno dei test effettuati per l'algoritmo misto e singolo.

Stimolo	Non articolato 1°esperimento			
	Alg. misto	DS	Alg. semplice	DS
01A vs 01IU+	82,1	4,47	80,6	4,8
01I vs 01AU+	73,7	16,0	64,6	6
01U vs 01AI+	78,5	12,7	64,4	5,7
01+ vs01AIU	78,1	14,4	64,3	5,9
02A vs 02IU+	70,5	13,	61,8	8
02I vs 02AU+	70,6	17,3	64,3	6,1
02U vs 02AI+	70,5	15,8	64,4	4
02+ vs02AIU	73,3	16,1	65,7	6
03A vs 03IU+	68,4	17,9	64	6
03I vs 03AU+	60,3	18,5	57,7	5,9
03U vs 03AI+	71,4	15,8	67	6,3
03+ vs03AIU	83,8	12,5	63,2	6
04A vs 04IU+	79	12,5	67,7	5,4
04I vs 04AU+	81,2	13,5	68,8	5,2
04U vs 04AI+	83,9	11,5	69,9	5,4
04+ vs04AIU	84,1	10	68,3	5,2
05A vs 05IU+	70,5	15,3	66,2	5
05I vs 05AU+	80,9	13,8	70,5	4,7
05U vs 05AI+	75,7	13,2	65,4	5,8
05+ vs05AIU	71,3	17	64,1	5
06A vs 06IU+	88,7	7,5	88,9	4,3
06I vs 06AU+	89,8	6,8	88,4	4,6
06U vs 06AI+	91,5	6,6	92,2	4
06+ vs06AIU	88,2	8,5	87,2	5,2
07A vs 07IU+	72,6	18	64,1	6,3
07I vs 07AU+	73	15	61,8	6,5
07U vs 07AI+	71,3	16,4	66	5,6
07+ vs07AIU	82	9,8	66,9	5,5
08A vs 07IU+	73,8	12	64,2	5,1
08I vs 07AU+	84,9	11,6	61,3	5,4
08U vs 07AI+	82,3	11	64,7	5
08+ vs07AIU	80,5	12,9	61,4	5,8
09A vs 07IU+	67,3	18,3	65,5	5,9
09I vs 07AU+	81,1	11	66,6	5,7
09U vs 07AI+	81,5	11,5	64,5	5,7
09+ vs07AIU	82,1	10,2	63,5	6

stimolo	Articolato 1°esperimento			
	Alg. misto	DS	Alg. singolo	DS
01A vs 01IU+	91,1	5,9	84,9	4,8
01I vs 01AU+	89,8	6,4	80,8	5
01U vs 01AI+	91,2	7,9	84,7	5
01+ vs01AIU	89,3	6,1	82,7	5,2
02A vs 02IU+	79	13	68,6	5,6
02I vs 02AU+	75,1	15,7	63,5	1
02U vs 02AI+	77,5	14,6	64,2	6,3
02+ vs02AIU	80,5	14,9	69,7	5
03A vs 03IU+	81,1	11,8	62,9	6
03I vs 03AU+	80,2	15,2	66,8	1
03U vs 03AI+	70,9	16,8	68,8	5,8
03+ vs03AIU	79,7	13	62,0	6,3
04A vs 04IU+	80,3	11,7	65,7	5,4
04I vs 04AU+	84,4	7,9	66,6	5,6
04U vs 04AI+	76,7	14	65,7	6,2
04+ vs04AIU	83,9	8,9	67,3	5,9
05A vs 05IU+	67,5	16	64,4	6
05I vs 05AU+	69,5	17	60,3	6
05U vs 05AI+	69	19	61	5
05+ vs05AIU	73	17	62,9	6
06A vs 06IU+	78,4	12,8	67,8	5,5
06I vs 06AU+	78,8	14	63,5	6
06U vs 06AI+	72,6	15	70,9	5,5
06+ vs06AIU	72,5	16	66,2	5,1
07A vs 07IU+	70,9	17,8	67	5,6
07I vs 07AU+	83,2	8,5	70,9	5,5
07U vs 07AI+	77,9	12	66,2	5,1
07+ vs07AIU	77	16	65,1	5,6
08A vs 07IU+	76,9	13	64,6	5,7
08I vs 07AU+	85,4	11,7	59,4	6
08U vs 07AI+	79,9	11,5	66,2	5,8
08+ vs07AIU	75,9	13	65,7	5,2
09A vs 07IU+	80,9	12	62,6	5,9
09I vs 07AU+	82,8	10,9	62,7	5,6
09U vs 07AI+	75,7	14	67,9	6,2
09+ vs07AIU	80	11	64	6

Non articolato 2°esperimento				
stimolo	Alg. misto	DS	Alg. singolo	DS
L_SA	82,3	12,7	65,5	3,8
L_SI	73,8	16,9	66,6	4
L_SU	75,2	16,9	68,3	4,3
L_S+	73,3	17	66,4	4,1
M_SA	84,5	15,2	81,8	3,7
M_SI	94,2	6,4	89,6	3,5
M_SU	82,4	14	85,7	3,0
M_S+	81,6	16	83,1	3,1
V_SA	71,1	16,9	72,4	3,9
V_SI	66,4	19	73,4	4,2
V_SU	74,3	14	70,4	4
V_S+	63,5	19	71,5	3,9

Articolato 2°esperimento				
stimolo	Alg. misto	DS	Alg. singolo	DS
L_A	74,1	12	69,8	3,9
L_I	62,8	17	67,3	4
L_U	65,3	14	61	4
L_+	65,4	18	63,	3 4
M_A	76	16	71,6	4,4
M_I	71,8	18	71,5	3,9
M_U	75,7	16	70,4	4
M_+	71,3	15	72,8	3,8
V_A	80,5	17	69,9	3,6
V_I	80,1	16	74,6	3
V_U	79,1	17	73,2	3,5
V_+	76,5	16	71	3,8

Dati relativi alla media di riconoscimento delle singole vocali nell'esperimento N°1
SESSIONE 1 che prevedeva il solo pensato,

STIMOLO	Misto	DS	Semplice	DS
A	74,76	13,21	69,2	5,6
I	72,27	13,7	67,1	5,5
U	78,5	12,7	68,7	5,2
+	80,37	12,3	67,1	5,6

Medie dei dati

Dati relativi alla media di riconoscimento delle singole vocali nell'esperimento N°1
SESSIONE 2 che prevedeva l'articolazione senza fonazione.

STIMOLO	Misto	DS	Semplice	DS
A	78,45	12,6	67,6	5,6
I	81	11,9	66	4,6
U	76,8	13,8	68,4	5,6
+	79	12,8	67,2	5,5

Dati relativi alla media di riconoscimento delle singole vocali nell'esperimento N°2

SESSIONE 1 che prevedeva il solo pensato.

STIMOLO	Misto	DS	Semplice	DS
A	79,3	14,9	73,2	3,8
I	78,1	14,1	73,6	3,9
U	77,3	14,9	74,8	3,7
+	72,8	17,3	76,3	3,7

Dati relativi alla media di riconoscimento delle singole vocali nell'esperimento N°2
SESSIONE 2 che prevedeva l'articolazione. fonazione,

STIMOLO	Misto	DS	Semplice	DS
A	76,8	15	70,4	3,9
I	71,5	17	71,1	3,6
U	73,3	15,3	68,2	3,8
+	71	16,3	68,9	3,8

Tabella riassuntiva dei valori medi di riconoscimento per ogni compito relativi al terzo esperimento, sessioni 1,2,3, solo pensato.

STIMOLO	MISTO	DS	SINGOLO	DS
A	86	10	94,8	4.2
I	90.5	7	95.5	4
U	90	7	95.8	3.9