



UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Psicologia Generale (DPG)

Corso di laurea magistrale in Psicologia Cognitiva Applicata

Tesi di laurea magistrale

**Analisi multivariata di dati EEG per lo studio dell'inibizione
di ritorno**

Multivariate analysis of EEG data for studying inhibition of return

Relatore

Prof. Marco Zorzi

Laureando: Francesco Resta

Matricola: 2017382

Anno Accademico 2022/2023

Introduzione

Capitolo I: MVPA

1 Cos'è la Multivariate pattern analysis

1.2 Storia della MVPA

1.2.1 Statistica classica e Apprendimento statistico

1.2.2 Haxby e il riconoscimento dei volti

1.3 Analisi univariata vs analisi multivariata

1.3.1 Univariato vs Multivariato

1.3.2 Segni di risposta uniformi vs non uniformi

1.3.3 Analisi direzionali vs non direzionali

1.3.4 Encoding vs Decoding

1.3.5 Stima statistica all'interno del campione vs predizione fuori dal campione

1.3.6 Attivazione vs Informazione

1.4 Predizione vs interpretazione

1.5 Time-resolved Decoding

1.6 Fra Neuroscienze e BCI: l'utilizzo di classificatori lineari su dati M/EEG

1.6.1 Support Vector Machine

1.6.2 Linear Discriminant Analysis

1.7 Analisi aggiuntive

1.7.1 Temporal generalization

1.7.2 Weight projection

Capitolo II: Studio sull'inibizione di ritorno

2.1 L'inibizione di ritorno

2.1 Origini del fenomeno

2.2 Teoria della facilitazione nella ricerca

2.3 Disancoraggio dell'attenzione e inibizione di ritorno

2.4 Inibizione di input e di output come due fenomeni distinti

2.5 Correlati neurofisiologici dell'inibizione di ritorno

Capitolo III: MVPALAB Toolbox

3.1 Mvpalab

3.2 Configurazione Generale

3.3 Specifiche per la Temporal Generalization

3.4 Specifiche del modello di classificazione

3.5 Calcolo delle statistiche di significatività

Capitolo IV: Applicazione della multivariate pattern analysis

4.1 Descrizione dell'esperimento

4.1.1 Metodologia

4.1.2 Descrizione del paradigma sperimentale

4.1.3 Registrazione dei dati EEG

4.1.4 Analisi dei dati comportamentali

4.1.5 Analisi dei potenziali evento relati (ERP)

4.1.6 Risultati comportamentali

4.1.7 Risultati segnali ERP

4.2 Descrizione della pipeline

4.3 Risultati e discussione

4.3.1 Analisi delle accuratèzze per istante di tempo

4.3.2 Temporal Generalization

4.3.3 Weights projection

Capitolo IV: Conclusioni generali

5.1 Limiti della MVPA

5.2 Questioni etiche

5.3 Conclusioni

Introduzione

La presente tesi si propone di esaminare l'applicazione della Multivariate pattern analysis (MVPA) su dati EEG per lo studio dell'inibizione di ritorno. L'inibizione di ritorno è un fenomeno prodotto dall'orientamento automatico dell'attenzione che si manifesta solo quando l'intervallo fra il segnale esogeno e lo stimolo target è maggiore di 200 millisecondi. In questo caso, i tempi di reazione sono rallentati quando lo stimolo-target compare nella stessa posizione del segnale esogeno. Questo processo è importante per la regolazione del comportamento e può essere influenzato da diversi fattori, come l'età e le condizioni neurologiche. Nonostante i potenziali evento-relati (ERP) siano ampiamente utilizzati negli studi sull'orientamento e controllo dell'attenzione visuo-spaziale le tecniche di analisi di questi dati risultano ancora non del tutto adeguate. L'applicazione dell'MVPA permette di ovviare ad alcuni dei limiti dell'analisi classica univariata, consentendo, tramite l'utilizzo di classificatori, di analizzare i dati con una maggiore sensibilità e accuratezza, tenendo conto dei cambiamenti che avvengono contemporaneamente fra le diverse variabili. In questo studio il disegno sperimentale prevedeva due condizioni, in una in cui lo stimolo-target appariva nella stessa posizione del segnale esogeno, nell'altra, invece, lo stimolo target appariva in una posizione differente. L'applicazione dell'MVPA consiste nell'addestrare classificatori ad ogni intervallo temporale in grado di classificare sulla base dei dati EEG a disposizione se il soggetto si trova in una o nell'altra condizione. L'accuratezza calcolata di ogni modello funge, poi, da indicatore al fine di comprendere se in quel determinato istante c'è una differenza nei correlati neuronali in grado di guidare il classificatore nel discernere le due condizioni. Questo disegno viene poi ripetuto per tre sotto condizioni che dipendono dal lasso di tempo che intercorre fra la presentazione del segnale esogeno e la presentazione del target (250 ms, 400 ms, 500 ms). I risultati hanno mostrato che l'accuratezza di classificazione è significativamente superiore al caso in diversi intervalli temporali, permettendo di identificare la dinamica temporale con cui l'attività elettrica cerebrale discrimina le due condizioni.

Capitolo 1: MVPA

1.1 Cos'è la multivariate pattern analysis

La “Multivariate Pattern Analysis” o MVPA è una tecnica innovativa per l'analisi multivariata dei dati di neuroimmagine. L'applicazione di questa tecnica trova riscontro nella letteratura più recente in quanto, grazie ai suoi numerosi risvolti applicativi, permette di superare molti dei limiti dell'analisi classica e univariata (Carlson, 2015).

L'MVPA nel settore neuroscientifico trova le sue prime applicazioni sulla risonanza magnetica funzionale, essa permetteva di considerare l'intera attivazione dei voxel all'interno di una registrazione fMRI piuttosto che analizzare le attivazioni attraverso la considerazione dei voxel come misure indipendenti e distinte (Haxby, 2012). *La multivariate pattern analysis* è, quindi, uno strumento sempre più ampiamente utilizzato in letteratura in quanto permette di ottenere una più sensibile e accurata analisi dei dati, tenendo conto dei cambiamenti che avvengono in contemporanea fra le diverse variabili (Grootswagers, 2017; Norman, 2006). Nell'analisi multivariata il materiale grezzo è l'insieme dei dati provenienti da diverse variabili che vengono analizzate simultaneamente. La sua applicazione offre, quindi, una maggiore comprensione e interpretazione dei dati, a differenza dell'analisi univariata che presenta diverse limitazioni nell'analizzare e comprendere la complessità dei fenomeni neuropsichici (Hebart, Baker 2018). La principale differenza tra analisi univariata e analisi multivariata (meglio approfondita nel paragrafo 3) risiede principalmente nella considerazione dell'oggetto di analisi. Per l'EEG, ad esempio, l'analisi univariata considera il picco di attivazione del potenziale post-sinaptico mentre un'analisi multivariata ha come oggetto di analisi l'informazione, ovvero, qualsiasi variazione nelle misurazioni di tutte le variabili costituisce un'informazione utile a discriminare diverse condizioni. L'MVPA nasce, quindi, come tecnica multivariata applicata all'analisi dei dati di risonanza magnetica funzionale. Con questo metodo l'attivazione emodinamica di ogni voxel non viene più considerata come una misura indipendente ma vengono analizzate le variazioni concomitanti dei numerosi voxel all'interno della misurazione complessiva (Haxby, 2012). Successiva all'utilizzo della MVPA su dati di fMRI (Haynes, 2015; Haynes, Rees,

2006; Tong, Pratte 2015), l'applicazione di questa tecnica si sposta su strumenti di neuroimaging ad alta risoluzione temporale come la MEG e EEG.

Gli strumenti tipici dell'analisi multivariata, sviluppati, inizialmente, come strumenti in grado di generare previsioni accurate nel mondo reale, nascono fuori dalla ricerca psicologica. L'utilizzo di queste metodologie ha, poi, suscitato un notevole interesse nella letteratura neuroscientifica in quanto molto efficaci per lo studio delle attivazioni cerebrali. L'indagine sui correlati di attivazione neuronale, prima dell'avvento delle tecniche multivariate, era dominato da una ideologia inadeguata a comprendere a pieno la complessità del fenomeno.

1.2 Storia della MVPA

Dal punto di vista storico la multivariate pattern analysis si inserisce in un progresso statistico applicato all'analisi dei dati nelle neuroscienze che segue due filoni distinti. La statistica classica e l'apprendimento statistico. Nel corso del processo evolutivo della ricerca neuroscientifica, i ricercatori sono passati da studi qualitativi su alcuni pazienti che riportavano lesioni cerebrali ad un tipo di analisi quantitativa tramite tecniche innovative di neuroimaging su centinaia di pazienti. Si è passati dalla colorazione manuale e l'ispezione microscopica di parti anatomiche del cervello a mappature tridimensionali in scala micrometrica. I dettagli all'interno dei dataset di neuroimaging continuano ad aumentare aumentando la risoluzione degli strumenti e quindi anche la complessità dei dati e delle informazioni contenute all'interno.

1.2.1 Statistica classica e Apprendimento statistico

Con il termine statistica classica e apprendimento statistico si intende identificare due filoni culturali distinti che hanno origini diverse (Breiman, 2001). La prima, la statistica classica, nasce con le ricerche della prima metà del ventunesimo secolo, la seconda, la statistica basata sull'apprendimento, prende piede principalmente nella seconda parte del Novecento (Efron, Tibshirani 1991). Nel corso di diversi decenni, le conoscenze in psicologia, come anche in medicina e nelle neuroscienze, è stata monopolizzata dall'approccio della statistica classica con la stima dei modelli e i conseguenti test di

significatività per verificare l'esistenza dell'effetto nel campione. L'apprendimento statistico (*statistical learning*), invece, si è concentrato per lo più sulla previsione su dati di notevole estensione (Bzdok, 2017).

È interessante notare come l'incongruenza fra i due diversi approcci si traduce anche attraverso l'uso della terminologia di base. La variabile di input e la variabile di output assumono nomi diversi a seconda dell'approccio. Nel paradigma classico, di solito, la variabile di input viene identificata come variabile indipendente, esplicativa o predittore, nel paradigma d'apprendimento, invece, viene riconosciuta come feature (caratteristica). Per quanto riguarda la variabile di output, viene chiamata dipendente nel primo approccio e target (obiettivo) nel secondo.

La statistica classica e lo "*statistical learning*" si intrecciano innumerevoli volte all'interno del processo evolutivo scientifico e metodologico. Nei primi anni del 1900

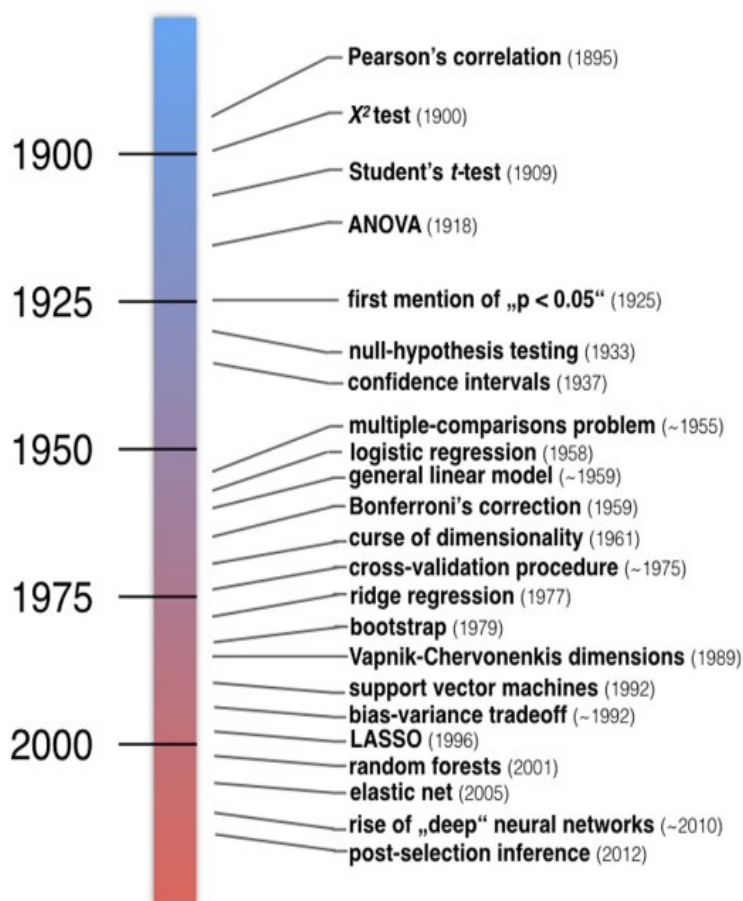


Figura 1. Rappresentazione grafica dello sviluppo storico della statistica classica e dell'apprendimento statistico. La parte in blu racchiude alcune delle metodologie della statistica applicata alla medicina e psicologia. La parte in rosso riguarda alcune tra le metodologie e le tecniche algoritmiche nate nella seconda metà del '900 (immagine estratta da: Bzdok, 2017),

emergono le prime nozioni centrali della statistica classica come la deviazione standard e il p-value (Cowles, Davis 1982). In questo stesso periodo appaiono anche importantissimi strumenti: William S. Gosset pubblica l'articolo t-test con lo pseudonimo di "Student", Fisher inventa l'analisi della varianza, il test dell'ipotesi nulla (Bzdok, 2017). Successivamente Pearson propone l'utilizzo dell'ipotesi alternativa introducendo le nozioni statistiche di falso positivo e falso negativo (Neyman, Pearson, 1933).

Durante gli anni 40 e 50 le inferenze di Fisher e il test t-Student vengono inseriti nei libri di testo americani e successivamente anche i lavori di Neyman e Pearson seguono lo stesso corso. Gli strumenti e le metodologie tipiche della statistica basata sull'apprendimento vengono invece utilizzati per lo più da computer scientist, ingegneri o fisici che ricoprono posizioni diverse, distanti dal contesto accademico, applicate al campo dell'industria. Uno dei dibattiti ancora presenti nella letteratura contemporanea riguarda proprio la considerazione dei due approcci. Il tema del dibattito consiste nel capire se considerare i due paradigmi come discipline separate (Chambers, 1993; Breiman, 2001; Friedman, 2001; Efron and Hastie, 2016) o ritenere la statistica come una classe metodologica molto più ampia che comprende le diverse metodologie della statistica classica e dell'apprendimento statistico (Tukey, 1962; Cleveland, 2001; Jordan, Mitchell, 2015; Blei, Smyth, 2017).

Il punto di inizio nell'approccio dello statistical learning può essere riconosciuto nella progettazione del *perceptron*, il primo algoritmo ispirato al funzionamento del cervello (Rosenblatt, 1958). Questa scoperta, assieme ad altre, hanno alimentato l'entusiasmo della ricerca verso l'intelligenza artificiale portando a periodi di forte e deciso sviluppo nel campo della statistica d'apprendimento intervallato da periodi poco fertili denominati in seguito come "AI winters" (termine apparso per la prima volta nel 1984 in un dibattito pubblico della "American Association of Artificial Intelligence"; Russell, Norvig, 2002; Crevier, 1993). Successivamente l'applicazione del machine learning inizia ad essere più presente specialmente in domini scientifici di tipo quantitativo, supportando l'enorme richiesta dovuta alla complessificazione dei dataset nella ricerca (Donoho, 2017). Per riassumere, la vera differenza fra il pre-guerra e il dopo-guerra risiede nel grado di automazione delle pratiche statistiche (Efron, Tibshirani, 1994). Con questo concetto i ricercatori, Efron e Tibshirani, puntano ad evidenziare gli sviluppi indipendenti che hanno caratterizzato le due metà del '900. L'emergere di un campo coerente dello statistical learning, infatti, sopraggiunge nella seconda metà del ventesimo secolo attraverso un certo numero di sviluppi indipendenti all'interno delle aziende e in alcuni dipartimenti universitari non statistici.

Nello stesso periodo, nel settore neuroscientifico, la comparsa delle innovative tecniche di neuroimmagine portano ad una trasformazione radicale nella tipologia e nella complessità dei dati oggetto di analisi. Strumenti come la Tomografia a Emissione di

Positroni (PET) e la risonanza magnetica funzionale (fMRI) permettono una descrizione “*in vivo*” dei correlati neurali di diversi task cognitivi e stimolazioni sensoriali. Il, così chiamato, “Brain Scanning” guida la ricerca neuroscientifica trasformando i dati in misure quantitative con diverse variabili per ogni osservazione. L’approccio da subito dominante nell’analisi dei dati provenienti dalle nuove tecniche di neuroimaging come la PET e la fMRI è quello tipico delle metodologie appartenenti alla statistica classica, ovvero, analisi univariate sul singolo voxel attraverso l’applicazione di strumenti come test statistici e utilizzo delle ipotesi e l’applicazione successiva dei General Linear Model (GLM; Scheffé, 1959; Poline, Brett, 2012). Questi modelli, per quanto migliorano la qualità di analisi rimangono saldi alla filosofia univariata. È importante ricordare che l’informazione distribuita fra i voxel viene rilevata con scarsa efficacia dagli strumenti dell’analisi statistica classica e dalle diverse applicazioni dei general linear model rispetto a tecniche dell’apprendimento statistico.

Nel 2001 Haxby e colleghi hanno ideato una interessante analisi di correlazione tra voxel, nel tentativo di avere prove contro l’allora presunta iper-specificità della corteccia ventrale nella risposta a stimoli visivi quali i volti. La concretizzazione della classificazione tramite quello che, poi, a posteriori verrà definito come algoritmo “*one-neares-neighbor*” (Mitchell, 2008) attraverso strumenti ancora tipici della statistica classica riporta numerose considerazioni e notevoli sviluppi metodologici e interpretativi:

1. L’analisi congiunta di set di zone cerebrali nel tentativo di includere rappresentazioni distribuite e sovrapposte;
2. Analisi ripetute con diverse parti dei dati nel tentativo di comparare con il livello del caso;
3. Analisi tra diverse categorie di stimoli nel tentativo di comprendere la specificità della risposta neurale.

Dopo questo fondamentale punto di svolta, l’applicazione dei metodi dello statistical learning su dati di neuroimaging comincia a divenire sempre più popolare tanto da presentarsi con diversi nomi, tra cui: mind-reading, brain-decoding o MVPA.

Il termine MVPA, inizialmente riferito alla “MultiVoxel Pattern Analysis”, viene sostituito con “MultiVariate Pattern Analysis” (Haxby, 2012). Questo, dato dal progresso

storico di questa particolare metodologia, inizialmente applicata esclusivamente su dati di risonanza magnetica funzionale (quindi principalmente attraverso i voxel) e successivamente applicata ad altre tecniche di neuroimaging, tra cui EEG e MEG.

L'interesse verso le diverse applicazioni del decoding nella ricerca sui correlati neuronali inizia a crescere sempre di più con i primi articoli e tutorial applicativi del *brain-image* (Haynes and Rees, 2006; Mur et al., 2009; Pereira et al., 2009). Nella maggior parte di queste guide viene messo in risalto come, il passaggio da una metodologia inquadrata in un paradigma filosofico basato sull'attivazione dei singoli elementi, ad una metodologia che invece si preoccupa maggiormente dell'informazione estratta dai dati, porta a numerosi vantaggi, tra cui uno spiccato aumento della sensibilità dell'analisi. Mur e colleghi nella loro "introductory guide" al pattern-information fMRI" descrivono attentamente la differenza dei due approcci nell'applicazione del multivariate decoding su dati di risonanza magnetica, specificano i limiti dell'approccio classico.

Il paradigma della statistica classica, nel tentativo di identificare le regioni cerebrali che si attivano nell'insieme, sottopongono i dati ad un processo di smoothing spaziale e l'attività dei singoli voxel viene, poi, mediata all'interno di una regione di interesse (ROI). Questo tipo di analisi, nonostante comporti una maggiore sensibilità verso attivazioni spazialmente estese, portano ad una drastica riduzione di sensibilità verso attivazioni nel dettaglio o a "grana fine" del modello spaziale. La riduzione della sensibilità da parte delle metodologie dell'analisi classica univariata è stata sottolineata più volte in letteratura, anche Norman nel 2006 nel suo articolo "Beyond mind-reading" sottolinea come l'averaging fra i voxel che rispondono significativamente ad una determinata condizione comporta numerosi svantaggi (Norman 2006). Oltre l'offuscamento dei pattern presenti nelle informazioni spaziali a grana fine, Norman, sottolinea una delle problematiche principali all'interno dell'analisi classica ovvero l'enorme rischio di trascurare informazioni rilevanti che risiedono in voxel che non presentano un'attivazione significativa, che mostrano, quindi, un'attivazione più debole rispetto ad altri voxel ma che possono comunque avere un ruolo fondamentale nella modulazione della risposta di attivazione a determinati stimoli o condizioni esterne.

La coscienza riguardo i limiti dell'analisi classica su dati di neuroimmagine diventa sempre più consapevole nella letteratura con il conseguente spostamento dell'attenzione

verso un nuovo tipo di analisi capace di comprendere le diverse sfumature di attivazione che risiedono nei dati. Con l'aumento costante di dimensione e complessità dei dati di neuroimaging, nei primi anni del ventunesimo secolo, l'applicazione di metodologie dell'apprendimento statistico diviene sempre più usuale. Al fine di una maggiore comprensione delle diverse metodologie statistiche si può pensare ad esse come situate su di un continuum i cui due poli sono la statistica classica e quella di apprendimento (Efron, Hastie 2016). Il primo polo, la statistica classica si interessa maggiormente di problemi con una minore dimensione del campione che possono essere affrontati da modelli con pochi parametri scelti dal ricercatore in modo analitico. Il secondo polo, invece, è stato concepito per problemi con più variabili e campioni con un'ampia dimensione. I processi della statistica classica assumono, quindi, che il comportamento dei dati sia guidato da meccanismi noti, mentre lo *statistical learning* sfrutta particolari algoritmi per evitare specifiche a priori riguardo i meccanismi alla base dei dati.

Dopo questa digressione sugli sviluppi storici dei due filoni, è importante aggiungere che fra le varie metodologie poste in questo continuum nessuna di esse può essere considerata migliore delle altre, infatti, sia la statistica classica che l'apprendimento statistico possono solo essere considerate relativamente al problema oggetto di considerazione. Il "No free lunch theorem" riassume perfettamente questa impossibilità di poter dichiarare la superiorità di un modello o analisi (Wolpert, 1996). Questo teorema, molto popolare nel campo della matematica e del machine learning afferma che tutti gli algoritmi di ottimizzazione performano egualmente se le loro prestazioni vengono mediate fra tutti i possibili problemi (Gigerenzer, 2004). Queste assunzioni permettono di comprendere intuitivamente che ogni algoritmo e/o tipologia di analisi presenta differenti caratteristiche che risultano performanti a seconda del problema presentato.

1.2.2 Haxby e il riconoscimento dei volti

Come già affermato in precedenza le basi su cui poggia il multivariate decoding nascono molto prima della loro piena considerazione da parte della letteratura scientifica. Il primo studio neuroscientifico in cui compare l'applicazione della MVPA su dati di neuroimaging arriva nel 2001. Dal titolo "*Rappresentazioni distribuite e sovrapposte di volti e oggetti nella corteccia temporale ventrale*" l'articolo pioniere nell'utilizzo

dell'MVPA, indaga le rappresentazioni neurali all'interno della corteccia temporale ventrale della presentazione dei volti (Haxby, 2001). Nello specifico è stata esaminata l'architettura funzionale del percorso della visione di oggetti e volti tramite risonanza magnetica. Anno dopo anno sono emerse in letteratura ricerche con titoli simili che utilizzano la multivariate pattern analysis (per una rassegna vedi Gessell, Geib, 2021). Nello specifico un utilizzo molto comune nei primi anni successivi alla prima applicazione dell'MVPA è quello di comprendere se un particolare modello di attività cerebrale rappresenta le categorie da cui vengono selezionati gli stimoli. Ad esempio, nel 2003 Carlson utilizza l'MVPA per comprendere le attività cerebrali che corrispondono a "rappresentazioni categoriali di oggetti", Serences e Boynton nel 2007 indagano la rappresentazione della scelta comportamentale del movimento, Wang nel 2012 studia invece le rappresentazioni concettuali astratte o concrete (Carlson, 2003; Serences, Boynton 2007; Wang, 2012). L'MVPA viene usata anche per altri numerosi scopi, come ad esempio: "sondare direttamente come le informazioni vengono rappresentate in aree cerebrali visivamente reattive" (Johnson, Johnson, 2014). Riassumendo, la multivariate pattern analysis rappresenta più di una semplice evoluzione dei classici General Linear Model che erano identificate come le tecniche più avanzate per avere una comprensione quanto più ottimale possibile delle attivazioni neuronali su dati di neuroimmagini.

Nello studio del 2001 Haxby e colleghi hanno ideato un prototipo di analisi molto ingegnoso, in quanto nel corso dello studio, sono riusciti ad applicare una multivariate pattern analysis attraverso strumenti tipici dell'analisi univariata. I risultati di questa ricerca hanno da subito dimostrato la grande potenza e sensibilità di questo particolare tipo di analisi rispetto ai metodi classici. Questo articolo, tuttavia, è diventato molto famoso in letteratura per il suo contenuto teorico-neuroscientifico (la rappresentazione neurale distribuita delle percezioni visive di alto livello), piuttosto che metodologico. Il riconoscimento dell'importanza dell'innovazione metodologica è avvenuto in seguito.

Haxby stesso in un articolo successivo, nel 2012, dove illustra le domande e i bisogni metodologici che hanno guidato la ricerca nel 2001 pioniera della MVPA, propone tre principali fattori in grado di comprendere la difficoltà da parte della letteratura di quel tempo nell'accettare del tutto l'innovazione metodologica della multivariate pattern analysis.

1. Non era ancora ben chiaro cosa rappresentassero i risultati di una MVPA in termini di rappresentazione neurale. Questo particolare tipo di analisi, infatti, non portava risposte alle domande tipiche dell'approccio statistico classico, localizzazionista (Dov'è l'area dei volti, Dov'è l'area del movimento, ecc..), bensì essa era particolarmente efficace a domande che investigavano i vari stati cerebrali all'interno di determinate aree e come queste codificano le diverse informazioni. Domande del genere erano non congruenti con la concezione della modularità cerebrale presente in quegli anni portando, quindi, anche ad una conseguente difficoltà nel comprendere la metodologia dell'analisi MVPA connessa ad esse;
2. Il decoding multivariato, appariva molto complicato, ed inoltre, al tempo non esisteva alcun software per l'implementazione che potesse semplificarne il processo;
3. L'analisi MVPA veniva fatta separatamente per ogni soggetto, perché avendo un altissima sensibilità a pattern di attività presenti nelle topografie a grana fine, questo portava ad una difficoltà nell'allineamento ottimale attraverso l'anatomia del cervello dei singoli soggetti.

Le ragioni che hanno portato Haxby e collaboratori nello svolgere una ricerca nel tentativo di indagare le rappresentazioni dei volti e degli oggetti nell'area temporale ventrale erano diverse. Nello specifico, questa ricerca era mossa dal dubbio sull'esistenza di aree specifiche nella percezioni di volti o oggetti, come suggerito dai risultati ottenuti in quello stesso periodo (Puce et al. 1996; McCarthy et al. 1997; Kanwisher et al. 1997). L'idea di fondo era proprio quella di sradicare il pensiero della modularità forte connesso alla logica della statistica classica univariata, applicata alle tecniche di neuroimmagine, che suggerivano come risposte cerebrali più deboli, anche se significative, non avessero alcun ruolo nelle rappresentazioni e che quindi dovessero venire scartate. La tesi dei ricercatori, infatti, verteva su di una rappresentazione neuronale distribuita. Nelle loro ipotesi era molto più probabile che le risposte deboli svolgessero un ruolo importante identificando la codifica di rappresentazioni di volti o oggetti come attuata da schemi di attività neurale piuttosto che da picchi di attività in moduli specifici per ogni categoria. Riassumendo, l'ipotesi modulare prevede che l'informazione per la classificazione degli oggetti per categoria fosse contenuta principalmente nei voxel che rispondono in maniera

massimale alla categoria in questione. Haxby et al. (2001) hanno scoperto che i voxel attivati al massimo in risposta a particolari categorie possono essere eliminati dall'input del modello con un costo minimo per l'accuratezza della classificazione. L'idea metodologica alla guida del progetto di ricerca si basava sul concetto di convalida incrociata per campioni multipli: Se una categoria di stimoli comporta un pattern di attivazione unico e distinto da altre categorie, allora, i dati di neuroimmagine appartenente ad una categoria dovrebbero essere più simili tra loro rispetto al confronto con categorie diverse. In questo tipo di analisi, i dati di attivazione, suddivisi nelle categorie di stimoli presentati erano, in seguito, divise in due parti distinte. Queste sottocategorie fungevano successivamente da elementi di confronto. Veniva calcolata la correlazione fra gruppi: "Within category correlation" e "Between category correlation". La correlazione fra i pattern è stata la misura scelta per la somiglianza. Sulla base di queste assunzioni, quindi, la correlazione fra due gruppi appartenenti alla stessa categoria doveva risultare maggiore rispetto alla correlazione fra gruppi di categorie diverse.

In seguito, è stata compiuta un'analisi sui pattern di attivazione sub-massimali. Ovvero, sono stati esclusi i voxel che presentavano un'attivazione massima ad una determinata categoria e sono stati ripetuti i confronti between e within. Una conclusione significativa riguarda come l'assenza dei pattern di attivazione massimale (ovvero i voxel considerati tipicamente nelle analisi classiche) non ledesse l'accuratezza nella classificazione dei correlati neurali delle rispettive categorie che rimaneva alta (94 %). Questo, oltre all'importanza teorica nel presentare prove contro la concezione della modularità forte, rendevano evidente il ruolo delle attivazioni sub-massimali molto spesso escluse dalla considerazione statistica classica. La rilevanza di queste ricerche è stata dimostrata, poi, dalla collaborazione di altri colleghi che hanno rianalizzato i dati attraverso l'utilizzo di classificatori a reti neurali (Hanson et al. 2004), LDA (O'Toole et al. 2005) o SVM (Cox & Savoy 2003). Successivamente questa modalità metodologica presente nell'articolo di Haxby del 2001 verrà considerata da Mitchell come "*regression-based one nearest neighbour (INN)*". Anche se in ricerche successive sono stati identificati classificatori che presentano una migliore efficienza la *INN regression-based* è stata riutilizzata in ricerche successive (Peelen et al. 2010). Per quanto l'interesse verso questa nuova e potente tecnologia sia cresciuto molto lentamente, l'entusiasmo verso l'MVPA è aumentato in modo significativo dopo due articoli pubblicati su Nature Neuroscienze nel

2005 ((Kamitani, Tong, 2005; Haynes, Rees, 2005). Entrambi hanno dimostrato l'applicazione dell'MVPA su caratteristiche ben definite quale l'orientamento dei bordi. Kamitani, Tong e Haynes, inoltre, tramite anche lavori successivi hanno mostrato come l'utilizzo dell'MVPA poteva permettere la decodifica di stati cognitivi come l'intenzione di svolgere un compito (Haynes et al. 2007) e l'obiettivo dell'attenzione visuo-spaziale (Kamitani, Tong, 2005). La decodifica multivariata nasce, quindi, da due correnti distinte: le tecniche multivariate utilizzate per fare previsioni nel mondo reale e l'utilizzo univariato nello studio delle attivazioni cerebrali. Questa parentela con due strumenti diversi porta a due principali vie di confusione nella concezione alla base del decoding. La prima riguarda l'analisi univariata classica e lo studio delle funzioni cerebrali tramite tecniche di neuroimaging. Il secondo, riguarda invece, l'applicazione del multivariate decoding per la predizione di eventi nel mondo reale. Questo comporta la necessità di distinguere due poli distinti per ogni corrente di partenza. La prima riguarda la differenza fra l'approccio classico univariato e l'approccio multivariato, la seconda invece tratta la distinzione fra modelli per la predizione e modelli per l'interpretazione dei risultati. Di seguito verranno illustrate i due principali poli di contrapposizione in modo da eliminare le confusioni originate dai due paradigmi di origine concettuale del multivariate decoding

1.3 Analisi multivariata vs analisi univariata

Una delle sorgenti di confusione emersi nella comprensione generale della decodifica multivariata risiede nel rapporto fra due correnti filosoficamente e pragmaticamente distinte: l'approccio della statistica classica univariata e quello della statistica multivariata. Questa dicotomia si aggiunge alla precedente differenziazione tra la statistica classica e l'apprendimento statistico, permettendo di inquadrare perfettamente il contesto pratico-metodologico dell'MVPA. Anche qui, la differenza fra i due gruppi metodologici parte da una ben più ampia differenziazione concettuale fra le due filosofie di origine: l'analisi classica univariata con la filosofia basata sull'attivazione e l'analisi multivariata con la filosofia basata sull'informazione.

Dall'avvento dei principali strumenti di neuroimmagine la mappatura funzionale del cervello ha subito notevoli sviluppi nel corso dell'evoluzione tecnico scientifica degli ultimi anni. Discostandosi anche dall'idea di base che il cervello consista nell'insieme di

regioni macroscopiche altamente specializzate (Haxby, 2012). Nei primissimi esperimenti di neuroimmagine dove veniva utilizzata la PET, l'attività cerebrale veniva misurata attraverso una risoluzione spaziale dell'ordine dei centimetri. I singoli elementi (voxel) erano di dimensioni molto più ampie di quelle che sono attualmente, con quel particolare ordine di grandezza il singolo voxel presentava misure accomunabili alle presunte regioni funzionali. Quindi l'unico modo di studiare l'attivazione dei correlati neurofisiologici era quello di considerare l'attivazione media del singolo voxel. La regione veniva, quindi, localizzata calcolando una statistica di attivazione per ciascuna posizione del voxel fissando una soglia della mappa statistica risultante. Queste particolari condizioni tecnologiche hanno portato la sedimentazione dell'approccio basato sull'attivazione (Hebart, Baker 2018). Con l'avvento della risonanza magnetica funzionale, la risoluzione spaziale è notevolmente aumentata, la larghezza dei voxel arriva a 4 millimetri. Nonostante le presunte aree funzionali fossero, ora, rappresentate da più voxel l'analisi standard per la fMRI è rimasta per molto tempo ancorata al paradigma basato sull'attivazione, in cui si presume che la regione coinvolta in un determinato processo cognitivo si attivi uniformemente nel suo insieme (Kriegeskorte, Goebel 2006). È proprio la concezione dell'attivazione di insieme delle determinate aree che porta poi all'utilizzo di metodologie per migliorare la potenza statistica delle analisi di neuroimmagini, ovvero lo smoothing spaziale, che, come è stato già precedentemente affermato, porta alla perdita delle informazioni sui pattern di attività presenti nella mappa spaziale a grana fine. La quantità delle informazioni rimosse attraverso lo spatial smoothing aumenta proporzionalmente con l'aumentare della risoluzione spaziale della misurazione. Riassumendo, il paradigma activation-based (ovvero la filosofia alla base dell'approccio predominante all'interno della letteratura neuroscientifica per numerosi anni) si basa, per lo più, su di un'analisi di differenti livelli di attivazione neuronale. In questo senso una più o meno ampia attivazione di un neurone (e/o gruppi di neuroni) indica un più o meno coinvolgimento del neurone stesso nel processo cognitivo oggetto di studio. Questa sistema di pensiero viene adottato, poi, sulle varie tecniche di neuroimaging: maggiore segnale BOLD nell'fMRI, un maggiore voltaggio negli elettrodi in un EEG o una maggiore banda di frequenza nella MEG. Una pratica comune all'interno di questo contesto riguarda l'utilizzo di modelli lineari che indagano relazioni, anch'esse spesso lineari, fra le variabili dello studio e i dati raccolti, successivamente,

vengo prodotte inferenze sulla base delle stime precedentemente ottenute dai modelli lineari.

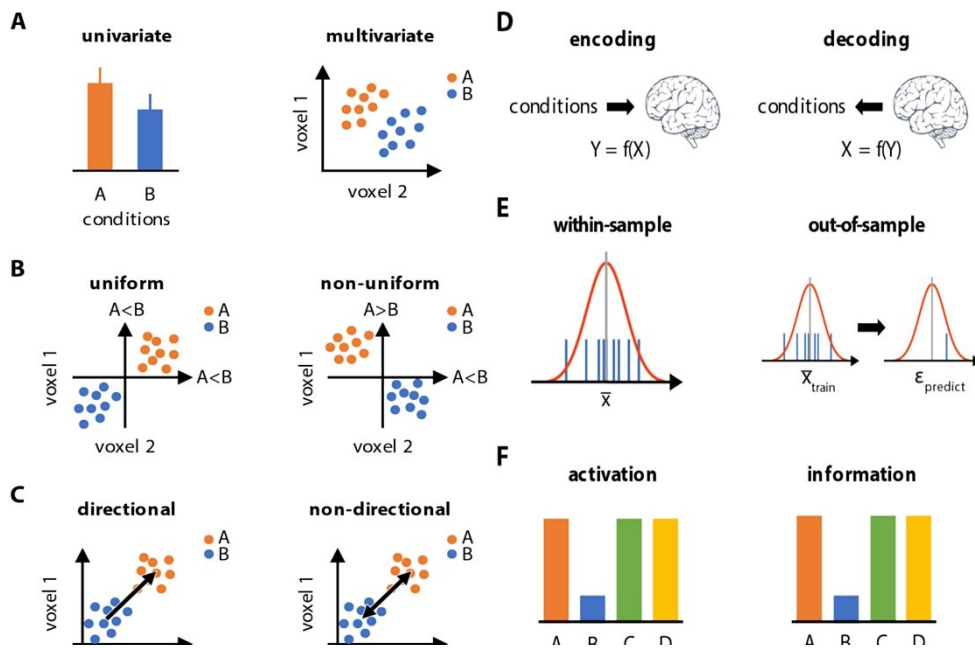


Figura 2. Le sei differenze tra analisi classica univariata e analisi multivariata (immagine estratta da: Hebart, Baker 2018)

L'approccio multivariato segue, invece, una filosofia basata sull'informazione. Qui ogni differenza misurabile tra le condizioni oggetto di studio risulta fondamentale in quanto viene utilizzata come discriminante nel processo di classificazione. Infatti, a differenza dell'approccio standard, il paradigma information-based, considera ogni singola differenza nella distribuzione dei dati accogliendo anche relazioni fra le variabili e i dati di tipo non lineare. In questo senso l'approccio basato sull'informazione risulta essere un orientamento più generico del primo. Nello specifico, considerando la decodifica multivariata come una particolare tecnica applicativa del paradigma information-based, è l'accuratezza del modello di decodifica che rappresenta la stima dell'informazione ottenuta dai dati. Oltre alle differenze principali che intercorrono fra i diversi approcci: l'univariato e il multivariato, riassunti dalle discrepanze presenti fra i due rispettivi paradigmi di appartenenza vi sono ulteriori discrepanze, nello specifico sei:

Univariato vs multivariato

La prima riguarda l'aspetto più generico dei due diversi approcci, che può essere intuitivamente appreso dal nome degli stessi. Univariato, infatti, si riferisce ad un tipo di analisi che avviene separatamente per ogni elemento dei dati (voxel), Multivariato invece indica un'analisi che tratta gli elementi di ricerca come un insieme congiunto di oggetti (Figura 2A), l'analisi di questi, infatti, viene conseguita attraverso la comprensione delle differenze di informazioni fra insiemi di elementi (gruppi di voxel per fMRI o canali per M/EEG). L'approccio univariato, infatti, tratta la moltitudine di segnali solo attraverso misure di raggruppamento successive ignorando in maniera considerevole la rilevanza di ogni misura nella comprensione della distinzione delle condizioni sperimentali e anche la loro covarianza. Il decoding multivariato invece risulta più sensibile al contributo che ogni elemento (voxel e canale) apporta alla comprensione di tale differenza (Hebart, Baker 2018).

Segni di risposta uniformi vs non uniformi

Una seconda differenza riguarda l'uniformità della risposta d'attivazione. Nell'analisi univariata, le regioni di interesse (ROI) vengono identificate come quelle regioni che mostrano un'attivazione congiunta e uniforme in risposta ad un determinato stimolo. Questa considerazione, nonostante non sia eccessivamente rigida in quanto considera che il livello di attivazione nella risposta all'interno di ogni regione può variare, assume comunque che fra le regioni di interesse vi debba essere lo stesso segno. Prendendo in esempio l'area fusiforme che risponde ai volti (FFA) sebbene sia noto che essa risponda in maniera differente a diverse tipologie di volti si assume comunque che questa presenti un'attivazione relativamente uniforme in risposta a qualsiasi volto (Haxby, 2001).

Nella codifica multivariata, invece, viene presa in considerazione a priori la possibilità che i voxel in una determinata regione possano manifestare segni di risposta differenti: sia l'attivazione che la de-attivazione dei voxel presenti viene considerata come informativa sulla variabile di interesse ed entrambi i segni contribuiscono alla stima

complessiva del contenuto informativo. Risposte significativamente positive o negative, quindi, rappresentano un contenuto informativo di eguale importanza.

Analisi direzionali vs non direzionali

L'utilizzo sempre più comune dell'approccio di decodifica multivariata rispetto alle metodologie della statistica classica univariata ha implicato un graduale passaggio da una definizione direzionale ad una non direzionale del tipo di analisi.

Nel paradigma classico univariato si dà per assunto che, una determinata regione cerebrale risulta correlata ad un determinato processo cognitivo, quando presenta un'attivazione maggiore ad una condizione sperimentale rispetto a quella di controllo, o anche quando presenta una qualche tipo di correlazione positiva o negativa con i vari livelli della variabile di interesse. Questo tipo di analisi viene eseguita per ogni singolo voxel. In seguito, sulla base della mappa dei risultati per singolo elemento, si determina se una regione cerebrale risulta coinvolta o meno. In questo caso si parla di analisi direzionale perché il segno, ancora una volta, ricopre un ruolo molto importante.

Nel multivariate decoding, invece, l'analisi è quasi sempre condotta in maniera non direzionale soprattutto perché all'interno di un ambiente multivariato la direzionalità perde di significato. Un voxel può presentare un livello di attivazione più alto in una condizione rispetto ad un'altra. Da queste premesse si evince l'enorme difficoltà nel definire la direzione complessiva di una specifica regione perché le attivazioni di una regione di interesse non devono necessariamente avere una uniformità di risposta, potendo presentare, così, una miscela di direzioni (Hebart, Baker 2018). Nella maggior parte delle analisi MVPA il focus principale è sulla differenziazione e discriminabilità dei pattern di attivazione piuttosto che sulla differenziazione fra i singoli voxel, l'idea di direzionalità non presenta alcuna rilevanza. È interessante comunque notare che esistono delle modalità di analisi direzionali nel contesto del multivariate decoding. L'aspetto di non direzionalità non sembra quindi essere un aspetto necessario dell'MVPA, ma il contesto applicativo nella discriminazione dei pattern rende l'analisi non direzionale una caratteristica considerabile dell'approccio di decodifica.

Encoding vs Decoding

La codifica descrive la previsione dei dati (variabili dipendenti) dalle condizioni sperimentali (variabili indipendenti), mentre la decodifica descrive la previsione delle condizioni sperimentali dai dati. Per fare un esempio, un modello generale lineare all'interno di un'analisi univariata classica è un modello di codifica, perché permette una descrizione di come il processo di studio viene per l'appunto, codificato all'interno dell'attività cerebrale (Hebart, Baker 2018). Il processo di decodifica è spesso applicato all'interno delle analisi multivariate, in particolar modo per la sua apparente semplicità. Questa tipologia di analisi, infatti, è divenuta sempre più comoda da eseguire grazie alla presenza di algoritmi di classificazione già preesistenti e preparati che portano ad una semplificazione del processo. Come è già largamente sostenuto in letteratura il decoding e l'encoding sono processi complementari quando l'obiettivo è quantificare la relazione fra variabili dipendenti e variabili indipendenti (Friston et al., 2008; Kriegeskorte, 2011; Naselaris et al., 2011). I modelli di decodifica permettono di comprendere la possibilità che una determinata informazione sia presente in una regione cerebrale (o in un determinato istante di tempo, nel caso della decodifica temporale). I modelli di codifica, invece, fanno previsioni sugli spazi di rappresentazione. In contesti sperimentali dove vengono indagati fenomeni sensoriali i modelli di codifica permettono di verificare teorie neuro-computazionali.

Stima statistica all'interno del campione vs predizione fuori dal campione

L'utilizzo della stima statistica all'interno del campione è uno specifico approccio tipico dell'analisi classica univariata. In questa metodologia viene inizialmente compiuta una stima statistica utilizzando l'intero set di dati disponibili in modo da poter comprendere la relazione che intercorre fra le variabili del disegno sperimentale e i dati a disposizione. Una volta compiuto questo passaggio le stime ottenute vengono esaminate attraverso l'utilizzo di test statistici per poter comprendere il grado di generalizzabilità dei risultati. Nel paradigma multivariato l'obiettivo del decoding non comprende l'acquisizione delle stime di attivazione sul campione di riferimento, bensì di ottenere stime dell'informazione sulle variabili sperimentali presenti nei dati. Questa stima dell'informazione viene

utilizzata all'interno di uno strumento predittivo utilizzato per predire le variabili sperimentali da dati diversi da quelli del campione di riferimento.

Nel “out-of sample prediction” la stima dell'informazione ottenuta da un set di dati, definito come dati di addestramento, viene impiegata per costruire un modello in grado di prevedere le variabili sperimentali da un set diverso definito dati di test. Nel decoding multivariato, questo particolare procedimento viene eseguito molteplici volte con diversi sottoinsiemi di dati in modo da rendere i risultati quanto più generalizzabili possibili. La capacità di prevedere le variabili da parte del modello sull'insieme dei dati viene spesso tradotta con il grado di accuratezza (Bzdok, Yeo 2017).

Attivazione vs Informazione

L'ultima differenza riguarda i già citati approcci di riferimento. Attivazione e informazione. Nell'approccio di tipo informativo sia l'attivazione che la disattivazione di una determinata area in risposta ad uno stimolo sono informativi della corrispondenza di una specifica regione ed il processo cognitivo. Nell'approccio basato esclusivamente sull'attivazione vi è una mancanza di sensibilità nei confronti di una disattivazione di un'area.

1.4 Predizione vs Interpretazione

Nel 2006 un gruppo di neuroscienziati organizza una competizione sulla predizione degli stati mentali tramite modelli di apprendimento automatico “Pittsburgh Brain Activity Interpretation Competition”. Ad ogni partecipante di questa competizione erano state assegnate le registrazioni fMRI che presentavano le attivazioni di alcuni soggetti impegnati nella visione di spezzoni di alcuni film. In seguito, gli stessi partecipanti hanno poi generato delle valutazioni su livelli differenti. La competizione consisteva nel prevedere queste valutazioni sulla base delle registrazioni fMRI. Il primo elemento che porta a suscitare interesse riguarda il gruppo di vincitori di questa competizione (Nature Neuroscience, 2006). Il gruppo vincente, infatti, era composto per lo più da data scientist privi di competenza specifica nel settore delle neuroscienze cognitive (Sona et al., 2007).

Un altro fattore interessante è che nell'addestramento del modello, fra le variabili più predittive, vi erano elementi che usualmente (nella ricerca neuropsicologica) vengono considerate come attivazioni rumore. I voxel maggiormente predittivi, quindi, non riflettevano un'attivazione cerebrale in risposta al segmento del film ma elementi collaterali con un forte valore predittivo. Ad esempio, nel tentativo di decodificare eventi a carattere umoristico è emerso che la zona cerebrale più informativa era la regione dei ventricoli. Questa tendeva ad agitarsi ogni qual volta il partecipante presentava la necessità di ridere. Nonostante un'elevata accuratezza del modello nell'identificare i momenti comici degli spezzoni, risulta evidente l'impossibilità di trarre conclusioni sul ruolo funzionale di questa zona nel rispondere a scene a carattere umoristico (Tong, Pratte 2011). Questo spinse gli ideatori della competizione di Pittsburgh a inserire nell'anno successivo un premio per ricompensare non solo il valore predittivo del modello ma anche la possibilità di estrarre contenuti interpretativi sul funzionamento cognitivo. Questa vicenda introduce uno degli aspetti fondamentali per l'utilizzo del multivariate decoding nel campo neuroscientifico: l'interpretazione.

La seconda sorgente di confusione nasce origini applicative del decoding multivariato. Esso, come già descritto in precedenza, è nato per fornire previsioni di eventi nel mondo reale partendo dai dati a disposizione (Hebart, Baker 2018). Il primo uso, infatti, riguarda la mera predizione e quindi un'attenzione maggiore a strategie di ottimizzazione che mirano al miglioramento dell'accuratezza di decodifica a discapito di un approccio che cerca una linearità e una maggiore interpretazione dei risultati. Il decoding multivariato orientato alla predizione è ampiamente utilizzato nel campo della BCI (brain computer interface). Esso punta all'identificazione di marcatori biologici che vengono utilizzati successivamente per la predizione puntuale di alcuni stati cerebrali. L'obiettivo primario risulta proprio quello della massimizzazione della performance, i modelli utilizzati prevedono strategie complesse di apprendimento che vengono spesso utilizzati come "black-box", dove, per l'appunto, il processo viene messo in secondo piano. Il decoding multivariato per l'interpretazione si preoccupa, invece, dell'aspetto maggiormente interpretativo del cervello umano e non richiede necessariamente alti standard di accuratezza. Nel momento in cui l'accuratezza supera il livello del caso (change level), vi sono i presupposti per poter affermare l'esistenza di una struttura di dati che differisce fra le condizioni prese in esame. Per questo tipo di analisi, le variabili altamente predittive

utilizzate dal gruppo di data scientist, vincitori della competizione di Pittsburgh nel 2006, sarebbero considerate variabili di disturbo in quanto utili per la performance ma altamente fuorviante nel caso dell'interpretazione. È opportuno tenere a mente che gli strumenti alla base del multivariate decoding, tra cui gli algoritmi di apprendimento utilizzati sono nati per l'approccio predittivo e che quindi bisogna utilizzare numerosi accorgimenti per gestire il delicato processo interpretativo.

1.5 Time-resolved Decoding

L'avvento del brain decoding per l'analisi dei dati di neuroimaging ha permesso un notevole sviluppo nella comprensione di diversi pattern di attivazione neuronale. Nonostante ci siano stati molteplici utilizzi del decoding multivariato per le analisi su serie temporali di dati di neuroimaging provenienti da strumenti quali magnetoencefalografia ed elettroencefalografia nel settore della Brain Computer Interface (BCI; Müller et al., 2008; Curran & Stokes, 2003; Wolpaw, Birbaumer, McFarland, Pfurtscheller, Vaughan, 2002; Kübler, Kotchoubey, Kaiser, Wolpaw, Birbaumer, 2001; Vidal, 1973), sono ancora molto recenti e ridotte le applicazioni nel campo delle neuroscienze cognitive. A questo punto, preme rimarcare la sostanziale differenza che si pone fra gli obiettivi della ricerca neuroscientifica e quelli della brain computer interface. Quest'ultima, infatti, utilizza il decoding per estrapolare informazioni nel modo più accurato possibile. Ancora una volta le differenze poggiano sul binomio: Performance del modello e interpretazione dei risultati. Questa discrepanza di obiettivi si traduce in utilizzi differenti del processo di decodifica (vedi paragrafo 1.6). Gli obiettivi da parte delle neuroscienze cognitive nelle analisi di decoding multivariato su dati di neuroimaging sono molto diversi da quelli delle BCI (approfondito nel paragrafo successivo).

Una particolare variante dell'analisi di decodifica, quale la time resolved MVPA, risulta possibile grazie all'utilizzo di strumenti di neuroimmagine diversi dalla classica risonanza magnetica. La particolarità della tecnologia M\EEG rispetto ad altri strumenti di neuroimaging risiede nell'altissima risoluzione temporale a discapito di una più bassa risoluzione spaziale. L'ottima precisione temporale, nell'ordine dei millisecondi permette l'applicazione della decodifica multivariate a risoluzione temporale. Nella risonanza

magnetica funzionale il delay dovuto alla ricezione del segnale BOLD non permette di dividere le attivazioni cerebrali per istanti di tempo, in quanto il ritardo della risposta emodinamica potrebbe portare a sfasamenti di alcuni istanti nella risposta fra i soggetti. Una precisione di registrazione al millisecondo tipica della MEEG e dell'EEG, invece, permette di inquadrare le attivazioni in determinate finestre di tempo, permettendo la sovrapposizione delle finestre di attivazione fra i soggetti.

Nel corso dello sviluppo storico e metodologico avvenuto in letteratura le analisi compiute attraverso l'utilizzo di strumenti come la MEG o l'EEG sono passate dall'essere confinate ad un tipo di statistica univariata ad una metodologia più ampia e multivariata, permettendo di comprendere maggiormente la complessità dei processi cognitivi intuendone la natura ampiamente distribuita delle rappresentazioni neurali (Huth et al., 2012; Cichy et al., 2014). Molte analisi dominate dal paradigma univariato si limitavano ad analizzare le differenze di ampiezze tra singoli canali o istanti di tempo tra le varie condizioni sperimentali, tralasciando la possibilità che alcuni stati cognitivi possano essere riconosciuti in base ad un modello di attività più distribuito e che quindi coinvolge l'attivazione di più canali in diversi istanti di tempo (Ng, Reh 2022). L'obiettivo dell'analisi di codifica nella maggior parte dei casi è quello di identificare dai dati di neuroimmagine a quale condizione o stimolo sperimentale è sottoposto il partecipante. Il pensiero di fondo, quindi, riguarda la capacità dei classificatori nel discriminare fra le varie condizioni partendo dai dati. Se questo è possibile, allora si può concludere che nei dati di partenza vi sono informazioni (in termini di correlati neurofisiologici) che guidano il classificatore nella discriminazione delle condizioni (Grootswager, 2016).

La prima parte di questa analisi consiste nella registrazione dei dati attraverso strumenti di neuroimmagine. Per i motivi già elencati in precedenza, al fine di una più corretta e possibile interpretazione, questa fase risulta molto delicata perché è necessario escludere, quanto più possibile, le variabili (o features) che risulterebbero predittive ma irrilevanti per la fase interpretativa, come ad esempio i correlati neurofisiologici del movimento oculare in un compito che studia l'orientamento dell'attenzione visuo-spaziale (Hong, Bo 2020). In seguito alla raccolta dei dati, la fase successiva si preoccupa del processamento dei dati grezzi e la successiva selezione del range di tempo su cui verrà eseguita l'analisi e la selezione delle feature.

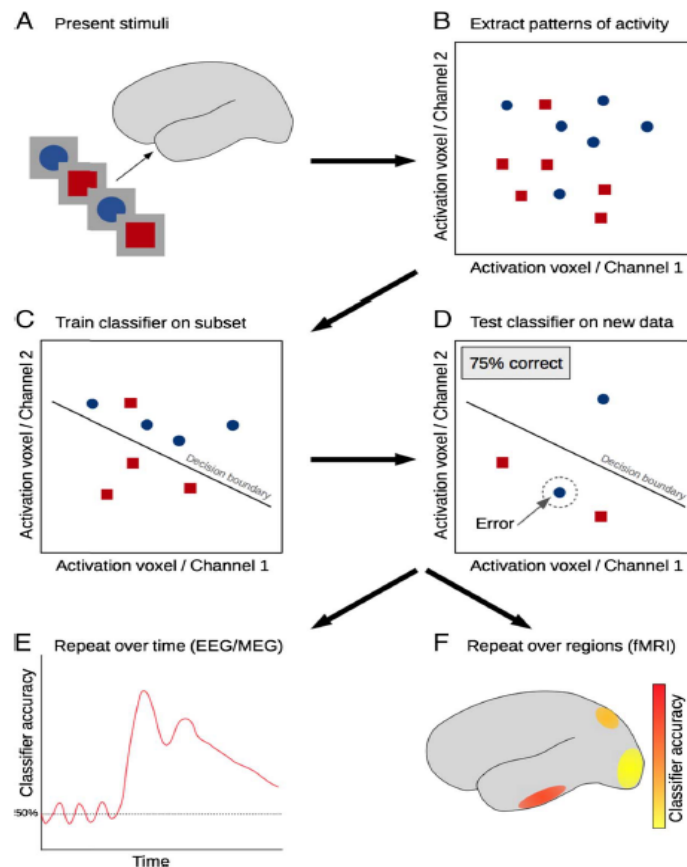


Figura 3. In questa immagine è rappresentato l'approccio generale alla decodifica multivariata. **(A)** L'attività cerebrale correlata alla presentazione di stimoli (quadrati rossi o cerchi blu) viene registrata attraverso le tecniche di neuroimaging. **(B)** Ogni misurazione viene inserita in uno spazio n -dimensionale (in questo caso vengono raffigurate solo due dimensioni per semplicità). **(C)** Viene addestrato un classificatore al fine di discriminare le due condizioni **(D)** Il classificatore viene testato su un set di dati diverso da quello di addestramento. **(E)** **(F)** Le ultime immagini raffigurano i possibili risultati. In un'analisi Time-resolved, i modelli vengono implementati per ogni modello creando una mappa temporale delle accuratze (E), per l'analisi MVPA spaziale invece i modelli vengono implementati in diverse aree (F). (immagine tratta da: Grootswagers, 2016)

Per la fase di selezione delle features bisogna considerare la grandezza del campione è il numero di variabili presenti nei dati. Quando il numero delle features (canali o voxel) supera il numero del campione bisogna considerare il rischio di generare una superficie eccessivamente complessa che performa efficacemente nella fase di addestramento ma presenta difficoltà nella classificazione nella fase di test. Una possibilità per arginare questo problema è rimuovere le features con una scarsa discriminabilità, apportando, così

una riduzione delle dimensioni dello spazio e una classificazione più generalizzabile (PCA). Il passo successivo è rappresentato dalla fase di addestramento del classificatore. Il modello di apprendimento automatico viene addestrato su una parte dei dati a disposizione, permettendo di comprendere gli eventuali pattern di attivazione presenti nelle diverse condizioni o stimoli sperimentali. Nello specifico, il classificatore, in questa fase, traccia il così detto “*Decision Boundary*” ovvero il confine decisionale che permette di distinguere una condizione rispetto ad un'altra. Per comprendere meglio il ruolo del classificatore nella decodifica multivariata si possono considerare i diversi livelli di una variabile come punti in un iperpiano, le quali coordinate corrispondono alle misurazioni di ogni feature. Nella Figura 3 è possibile vedere uno dei casi più semplici. In questo caso gli stimoli sperimentali sono rappresentabili in un piano bidimensionale in quanto si considerano solo due feature (voxel o canali). In questo esempio si parla di una classificazione binaria in quanto vi sono solamente due categorie. Come è possibile notare, una volta disposti gli stimoli nel piano, il compito del classificatore è quello di tracciare un confine che permette di scegliere, sulla base delle coordinate date dalle features considerate, a quale condizione appartiene. Questo confine decisionale viene tracciato nella fase di addestramento del modello (Grootswagers, 2016).

Nella fase successiva, quella di test, il modello viene, per l'appunto, testato su dati diversi da quelli usati per l'addestramento. Calcolando, in fine, la misura della performance del classificatore che, in questo specifico esempio, verrà definita tramite l'accuratezza. Essa rispecchia il rapporto fra gli stimoli classificati correttamente e gli stimoli classificati erroneamente. In seguito, a seconda dello strumento di neuroimmagine utilizzato e del tipo di analisi scelto si ottiene una distribuzione delle accuratèzze che può essere di tipo spaziale o temporale (Figura 4).

Molte ricerche negli ultimi anni hanno prodotto numerosi risultati applicando l'MVPA partendo da dati M\EEG, sono stati studiati numerosi processi cognitivi tra cui: l'emergere delle rappresentazioni degli oggetti a livello di categoria e di esemplare utilizzando la MEG (Carlson, Tovar, Alink, Kriegeskorte, 2013), l'EEG (Cauchoix, Barragan-Jason, Serre, & Barbeau, 2014); le rappresentazioni invarianti degli oggetti emergono nel tempo (Kaiser et al. 2016; Isik et al. 2014; Carlson et al, 2011); e come gli oggetti sono rappresentati in altre modalità (ad esempio, scritta o uditiva; Simanova et al. 2010,2015; Murphy et al., 2011; Chan, et al. 2010).

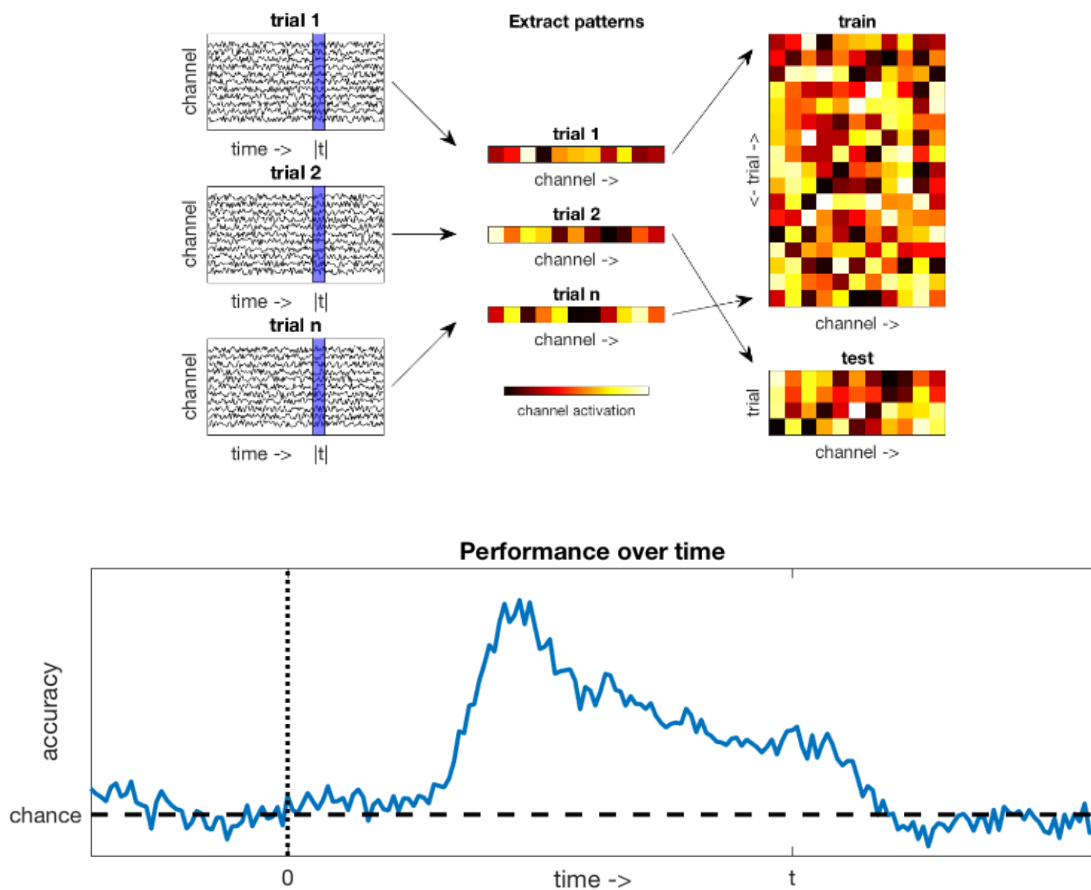


Figura 4. Nella prima parte è raffigurata la modalità con cui vengono creati i dataset per ogni istante di tempo. Nella seconda parte è raffigurata la distribuzione delle accuratze ottenute per ogni istante di tempo. (Immagine estratta da Grootswagers 2016)

Altri studi hanno utilizzato questo approccio per decodificare l'orientamento e la frequenza spaziale dei reticoli MEG (Wardle, Kriegeskorte, Grootswagers, Khaligh-

Razavi, & Carlson, 2016; Cichy, Ramirez, & Pantazis, 2015) e per studiare il processo decisionale (Stokes et al, 2013; Bode et al., 2012), le illusioni (Hogendoorn, Verstraten, & Cavanagh, 2015) o la memoria di lavoro (Wolff, Ding, Myers, & Stokes, 2015; van Gerven et al., 2013).

Per quanto concerne la time resolved MVPA, un sotto-obiettivo riguarda l'ottenere una mappa temporale che presenta le accuratèzze o in maniera piú generica, la performance del modello per ogni istante di tempo. Una volta ottenuto questo, si calcola la significatività oltre il livello del caso, per ogni istante di tempo (Ashton et al. 2022). Il calcolo della significatività per ogni istante di tempo avviene attraverso 'utilizzo di test statistici per confrontare la distribuzione delle accuratèzze con le accuratèzze nulle, per ogni istante di tempo. Il calcolo delle accuratèzze nulle avviene di solito attraverso l'utilizzo di numerose permutazioni. Il processo di addestramento e test viene ripetuto numerosissime volte mischiando ogni volta in maniera casuale i dati, in modo da ottenere una distribuzione di accuratèzze dettate dal caso. Vi sono diverse modalità con cui può essere eseguita un'analisi dei pattern multivariati per istanti di tempo. Una prima differenza deriva dalla considerazione dei soggetti. L'analisi può essere svolta, infatti, "single subjects" o "across subjects", nella prima i trial di ogni condizione per ogni soggetto vengono utilizzati separatamente per ottenere una stima della performance del modello indipendente per ogni soggetto. Nella seconda tipologia di analisi, invece, i trial di ogni soggetto vengono utilizzati per l'addestramento e valutazione di un singolo modello creando così una singola misura della performance per ogni istante di tempo. È intuibile che differenti modalità di analisi comportano vantaggi e svantaggi. Nel primo caso creare misure indipendenti per soggetto permette di controllare le differenze individuali che possono ridurre, la performance del classificatore comportandosi da elementi di disturbo nell'analisi *across-subject*. Nel seconda tipologia di analisi invece la considerazione dell'intero set di dati utilizzato per un unico modello permette di avere a disposizione piú dati nel processo di addestramento e quindi ridurre il rischio di overfitting del modello e migliorare la generalizzabilità dei risultati. In ogni caso, il set di dati del modello viene formato dall'insieme delle misure di ogni istante dei diversi trial. Le misure continue di ogni prova vengono segmentate in numerose finestre di tempo, seguendo un passo determinato a priori in modo da ottenere un dataset di dati contenente la stessa finestra di tempo di ogni trial.

Da questo set di dati, successivamente, si eseguono le suddivisioni nei gruppi di addestramento e di test. Le modalità di suddivisione rispettano rigorose procedure fondamentali nella classificazione del modello. Questa fase prende il nome di convalida incrociata (traduzione di *cross-validation*). In questo fondamentale passaggio, il processo di suddivisione, addestramento e test viene ripetuto numerose volte a seconda della strategia di *cross-validation* utilizzata. La validazione incrociata consiste, quindi, nella suddivisione dell'intero dataset in diversi sotto-insieme di dati che verranno, poi, rimescolati per formare i due gruppi (di addestramento e di test), ogni volta diversi. L'elemento fondamentale della convalida incrociata riguarda l'uso di ogni sottoinsieme di dati per la fase di addestramento e per la fase di test. Questo continuo riutilizzo dei sottoinsiemi nei due gruppi rende il processo statisticamente efficiente e robusto (Kriegeskorte, 2015).

Una delle modalità di suddivisione dei gruppi (Fold) viene definita come "*leave-one-out*". Questo tipo di *cross-validation* consiste nel suddividere i dati a disposizione ottenendo n gruppi per n trial, il processo di addestramento, così, avviene sull'intero set di dati meno un trial per condizione su cui poi viene testato il modello. La tecnica *leave-one-out* può essere considerata come una particolare tipologia di K-Fold cross-validation, che indica, per l'appunto, la suddivisione del set di dati in k fold (contenendo un singolo trial) con k-1 fold per la fase di addestramento e 1 fold per quella di test. Come per la maggior parte delle forme di analisi, la scelta della migliore strategia di validazione incrociata è strettamente guidata dalla domanda sperimentale, è quindi fondamentale considerare le varie tipologie di splitting dei dati in linea con l'obiettivo dello studio. Ci sono, quindi, numerose scelte che vengono compiute all'interno dell'analisi MVPA che rendono l'intera procedura altamente diversificata all'interno della moltitudine dei disegni di ricerca. Le scelte che vengono compiute riguardano, non solo la differenza fra analisi fra i soggetti o con i soggetti, ma numerosi aspetti che appartengono alle diverse fasi dell'analisi. Partendo dai dati una considerazione di notevole rilevanza riguarda l'applicazione dell'*averaging*. È ormai noto che i dati come quelli EEG abbiano una concreta quantità di rumore, rendendo molto complesso il riconoscimento di pattern da parte del classificatore. Una procedura standard nel tentativo di ridurre il rumore è proprio quella di utilizzare degli pseudo-trial che rappresentano il risultato della media dei canali di n trial. L'utilizzo di pseudo-trial permette di eliminare gran parte del rumore presente

ma ad un costo molto importante, ovvero, la riduzione del numero di prove utilizzate per l'addestramento del modello (Blankertz, 2011). Altre scelte che bisogna compiere all'interno del processo di decoding riguardano: l'ampiezza della finestra temporale oggetto di analisi, l'applicazione di filtri nel pre-processing, la scelta del classificatore, la scelta della tipologia di misura della performance del modello e come già menzionata anche la scelta della tipologia di convalida incrociata. Tutte queste scelte comportano un'ampia variabilità con una conseguente flessibilità di ricerca da parte del multivariate decoding alle domande e obiettivi dello studio.

1.6 Fra Neuroscienze e BCI: l'utilizzo di classificatori lineari su dati M/EEG

L'applicazione del decoding su dati M/EEG è stata inizialmente dominata dall'approccio tipico della Brain Computer Interface. Un'interfaccia cervello-computer (BCI) segue obiettivi diversi dalle neuroscienze cognitive. Mentre le neuroscienze, infatti, si preoccupano della comprensione e interpretazione dei pattern di attività neuronale, l'obiettivo principale della decodifica nella BCI è quello di tradurre l'informazione data dall'attività neuronale in comandi o azioni utilizzabili per controllare un dispositivo esterno, come ad esempio un braccio robotico o il cursore di uno schermo. Un'interfaccia cervello-computer si preoccupa, quindi, di fornire al sistema nervoso centrale nuovi output, che non sono di tipo muscolare o ormonale (Wolpaw, Wolpaw 2012). L'attività del sistema nervoso centrale comprende una moltitudine di fenomeni elettrici, chimici e metabolici che possono essere registrati attraverso le diverse tecnologie delle neuroimmagini. Il ruolo della BCI consiste, quindi, nel catturare questi fenomeni e convertirle in output artificiali che agiscono direttamente nel mondo esterno. L'utilizzo di queste interfacce riveste un ruolo molto importante, in quanto la loro applicazione permette di sostituire l'output naturale, danneggiato o perso a causa di una malattia, con uno artificiale. Nel processo di decodifica nelle interfacce cervello-computer, l'informazione deve essere estratta in tempo reale e con il più alto grado di sensibilità. Sulla base di queste informazioni risulta spontaneo intuire quanto le metodologie dei due diversi approcci divergono sostanzialmente. Nelle BCI l'efficienza è una caratteristica fondamentale, portando all'applicazione di algoritmi di decodifica sempre più complessi e avanzati. Per quanto riguarda le neuroscienze, invece, il processo di decodifica deve

seguire diversi accorgimenti per mantenere il protocollo di ricerca quanto più semplice e lineare in modo da garantire una corretta generalizzabilità e interpretabilità dei risultati. L'utilizzo di procedure lineari permette, quindi, di arginare la concezione del modello come scatola nera, consentendo anche di interpretare le caratteristiche del classificatore. Oltre ad una minore rielaborazione delle features la minore complessità all'interno delle procedure di decodifica delle neuroscienze cognitive comporta anche ad una scelta di algoritmi di classificazione semplificati, quali i classificatori lineari (Poline, Brett 2012).

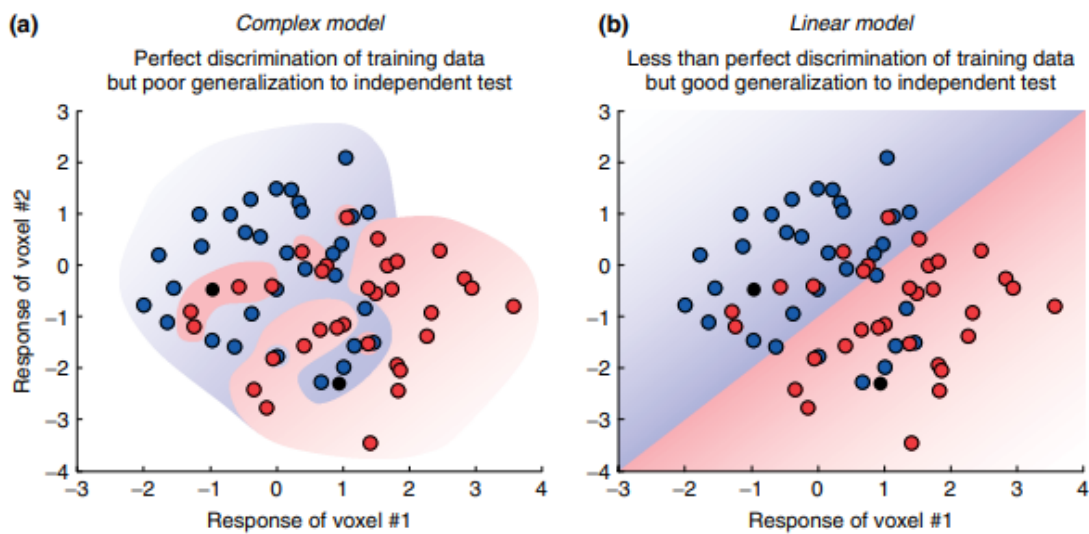


Figura 5. (a) Rappresentazione geometrica del funzionamento di classificatori con algoritmi complessi non-lineari. (b) rappresentazione di algoritmi di classificazione più semplici con kernel lineari.

Esistono numerosi algoritmi di classificazione per l'analisi multivariata dei pattern ma quasi la totalità di essi condivide la considerazione multidimensionale dei dati, dove ogni feature rappresenta una dimensione. Come già descritto in precedenza, il classificatore viene prima addestrato in fase di training (calcolando così il limite decisionale), e successivamente testato nella fase seguente. La fase di test ricopre un ruolo fondamentale nell'individuare i modelli che vanno in "overfitting" ovvero quando il modello performa molto bene con i dati di training e male nei dati di test (Ying, 2019). Le cause che portano all'overfitting sono molteplici, una di queste riguarda l'elevata quantità di rumore presente nei dati di neuroimmagine. Le informazioni di disturbo presenti nei dati di addestramento possono portare il classificatore a modellarsi su di esse anziché su delle caratteristiche discriminanti delle diverse categorie. Uno dei modi per tamponare la

probabilità che un modello vada in overfitting è proprio quello di utilizzare classificatori più semplici in quanto modelli non lineari con algoritmi più complessi hanno più probabilità di modellarsi eccessivamente sui dati di addestramento (Schwarzkopf, Ress 2011).

Nella figura 5 si può notare come l'utilizzo di algoritmi di classificazione più complessi porti ad una maggiore performance nella fase di training dovuta ad una modellazione iperspecifica sui dati di addestramento. Un confine più flessibile, infatti, può presentare un'alta probabilità di adattamento alle idiosincrasie del rumore presenti nei dati di addestramento. Questo si traduce in una maggiore generalizzabilità del modello con un kernel lineare. Una domanda importante per comprendere quali classificatori è meglio utilizzare nell'analisi dei pattern multivariati riguarda, per l'appunto, quali sono le caratteristiche che differenziano i vari classificatori (Lotte, Congedo 2007).

In primo luogo, essi differiscono dalle forme ammesse per il confine decisionale. Per i classificatori lineari esso sarà rappresentato da un iperpiano di $n-1$ dimensioni, dove n indica il numero di features (nell'esempio semplificato della figura 5 dove erano rappresentate le misure delle attivazioni di due features il confine decisionale è rappresentato da una linea). Per i classificatori non lineari invece il confine decisionale sarà non planare con una forma più complessa (Misaki, Kim 2010).

Tra i classificatori lineari più comuni ci sono: Linear Discriminant Classifier, Naive Bayes, Logistic Regression, Perceptron, SVM (kernel lineare). Di seguito verranno approfonditi i due classificatori utilizzati nella presente analisi MVPA: la Support Vector Machine e la Linear Discriminant Analysis.

Support Vector Machine (SVM)

La support vector machine (SVM) è un modello di apprendimento supervisionato basato su algoritmi di classificazione tra i più semplici ed efficaci con importanti proprietà di regolarizzazione e solide basi teoriche (Bhavsar, Panchal 2012). Come è stato descritto in precedenza l'interpretazione geometrica del ruolo del classificatore consiste nel calcolare il confine decisionale in grado di separare le classi disposte come punti all'interno di uno spazio n -dimensionale. La support vector machine è un insieme di

algoritmi molto duttile in quanto non è necessariamente lineare ma può utilizzare kernel diversi (tra cui quello lineare). Per comprendere a pieno il concetto di kernel, bisogna prima chiarire la nozione di iperpiano di separazione (ovvero l'interpretazione geometrica del decision boundary menzionato precedentemente) e la nozione di massimizzazione dei margini.

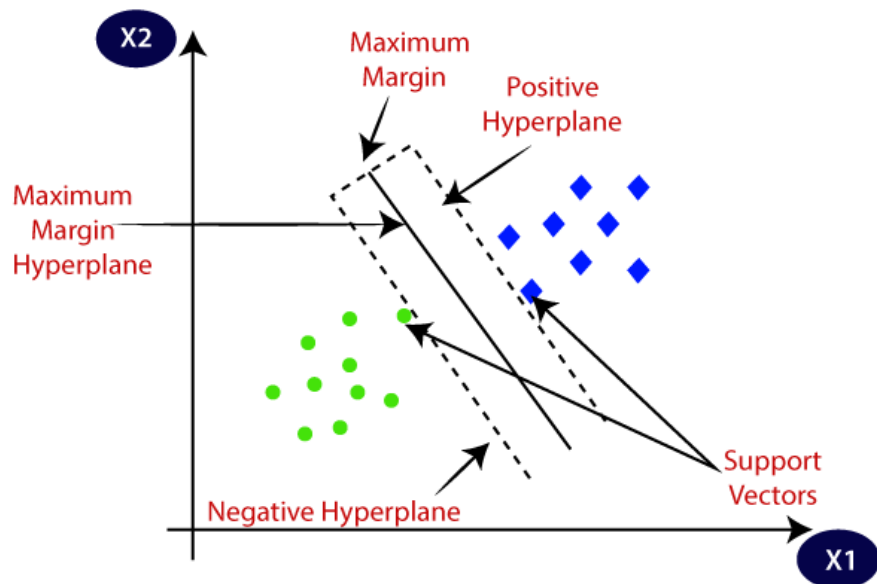


Figura 6. Rappresentazione geometrica della Support Vector Machine

Nell'esempio semplificato di misure di attivazioni con solo due features (figura 6), il modello SVM (lineare) calcola una retta di separazione fra le due categorie (in questo caso la retta è un iperpiano di $n-1$ dimensioni dove n è il numero di features). Come è intuibile le rette che permettono una distinzione fra le due classi nell'addestramento del modello sono molteplici, ciò che differenzia, infatti, i modelli lineari fra loro è il criterio con cui viene scelta la migliore retta fra le diverse alternative. Il problema principale di un classificatore, infatti, riguarda come scegliere il migliore iperpiano di separazione che permette la migliore generalizzazione. Questo problema è stato in parte risolto nel 1965 (Vapnik, 1982). Per la support vector machine un iperpiano ottimale è definito quindi come il limite decisionale con il massimo margine tra gli elementi delle due classi. Per trovare, in questo caso, la retta migliore per distinguere le due classi bisogna prima individuare i punti più vicini alla linea delle due classi. Questi punti vengono definiti "Support vectors". La distanza fra i support vectors e il confine decisionale viene definito margine. L'obiettivo dell'algoritmo della support vector machine è quello di trovare la

retta, la quale presenta i margini più grandi. La selezione di questo particolare limite decisionale massimizza la capacità di previsione e generalizzabilità della SVM (Noble 2006). Ci sono situazioni in cui i dati a disposizione non sono discriminabili da un iperpiano lineare, in questo caso si ricorre alla funzione del Kernel che permette di aumentare la dimensione geometrica dove sono disposte le misurazioni aumentando le possibilità di discriminabilità degli stessi. L'iperpiano di separazione così diventa non-lineare. Concludendo l'SVM (lineare) è uno degli strumenti più robusti utilizzati nel decoding multivariato grazie all'ottimo bilanciamento fra performance del classificatore e rischio di una bassa generalizzazione dei risultati (overfitting).

Linear discriminant analysis (LDA)

L'analisi discriminante lineare, o LDA, è un algoritmo di classificazione di tipo lineare che presenta molte differenze con la support vector machine. Proposta da Fischer nel 1936, l'LDA, consiste nell'individuare l'iperpiano che permetta di minimizzare la varianza fra le classi e massimizzare la distanza delle medie per ogni categoria. L'analisi discriminante lineare presenta molti punti in comune con la PCA (Principal Component Analysis), una tecnica utilizzata per ridurre la dimensione del campione eliminando dai dati le componenti che vengono reiterate e quindi superflue. Anche la LDA può essere utilizzata per ridurre la dimensionalità del campione, in quanto il principio di funzionamento di questo classificatore è riportare le misure n-dimensionale delle osservazioni su una linea (nel caso di una classificazione binaria). Le osservazioni vengono, infatti, proiettate su una linea come punti ed in questo spazio viene definito il confine decisionale fra le classi (Xanthopoulos, 2012). Riassumendo, l'interpretazione geometrica di questo particolare classificatore riguarda la costruzione di un asse su cui vengono proiettate le misure, in modo da ridurre la dimensionalità del campione. La posizione dell'asse viene decisa seguendo due criteri: massimizzazione della distanza fra le medie e minimizzazione della varianza fra gruppi. Una volta proiettate le misure il classificatore pone il limite decisionale su uno spazio monodimensionale.

1.7 Analisi aggiuntive

Una volta completata l'intera procedura del *Time-resolved decoding* si ottiene la distribuzione della performance del classificatore. Grazie al calcolo della significatività risulta poi possibile identificare i momenti in cui l'accuratezza supera il livello del caso e quindi permette di individuare gli istanti di tempo in cui sono presenti attività neuronali in grado di guidare il classificatore nella discriminazione. Ottenuti questi risultati, attraverso il processo di ricerca tramite il multivariate decoding, si possono ottenere altre informazioni per affinare il contenuto interpretativo dell'analisi. Tra queste vi è la Temporal generalization e la Weight projection.

Temporal Generalization

La temporal generalization consiste principalmente nell'addestrare il modello nella fase di training su un determinato istante di tempo e testarlo nella fase successiva su dati appartenenti ad un altro istante. Ripetendo sistematicamente questo approccio si ottiene una matrice di performance del classificatore generalizzato nel tempo. Sulle righe vi è la performance del modello nei dati di addestramento, sulle colonne, invece, vi è la performance nella fase di test, la diagonale di questa matrice corrisponde invece al vettore delle accuratèzze ottenute dall'analisi di decoding.

Questa procedura permette di comprendere il livello di generalizzazione nel tempo dei pattern di attivazioni neuronali. L'idea di fondo è che quando un modello addestrato su dati di un determinato istante di tempo, presenta performance relativamente alte quando viene testato su dati di istanti di tempo diversi, allora è possibile pensare che la struttura dei pattern risulta simile nel tempo (Grootswager 2016). Questo tipo di informazione aumenta notevolmente il valore interpretativo dell'analisi, in quanto permette non solo di individuare quando e per quanto tempo un'informazione rilevante è presente nel cervello ma con analisi aggiuntive quale la temporal generalization è possibile comprendere se questa informazione ricorre nel tempo e soprattutto per quanto tempo ricorre (King, Dehaene, 2014).

Weight Projection

Un'ulteriore analisi possibile è la “*Weight Projection*”. Questa tipologia di analisi risulta molto interessante in quanto permette di ottenere informazioni spaziali in un'analisi che presenta un'altissima risoluzione temporale. La *Weight Projection*, o proiezione dei pesi, consiste nell'indagare per ogni modello creato quali sono state le features più significative. In ogni classificatore, infatti, viene attribuito un peso per ogni feature che indica il valore predittivo e quindi l'importanza che esso ha nel guidare la discriminazione fra le condizioni sperimentali. Bisogna fare molta attenzione quando si trattano i pesi di un classificatore, in quanto si potrebbe cadere nell'erronea convinzione che il valore per ogni feature abbia una corrispondenza diretta con il suo ruolo all'interno delle attivazioni relative ad un determinato processo cognitivo. Purtroppo, i pesi grezzi non sono così facilmente interpretabili, valori alti di un determinato peso, infatti, non implicano necessariamente una maggiore quantità di informazioni classe-specifiche da parte della variabile corrispondente, così come un peso diverso da zero non indica la presenza di informazioni da parte del canale (o voxel per fMRI). Questo è dovuto principalmente al fatto che molto spesso i classificatori attribuiscono rilevanza ad alcune features semplicemente perché permettono di controllare il rumore nei dati e non perché possiedono informazioni discriminanti. Una soluzione a questo problema è stata proposta da Haufe e colleghi nell'articolo pubblicato nel 2014 “*On the Interpretation of weight vectors of linear models in multivariate neuroimaging*”. In questo lavoro Haufe e colleghi propongono una metodologia in grado di trasformare i pesi grezzi in indicatori di attivazione, creando così una mappa di attivazioni a distribuzione spaziale. Un'importante distinzione per comprendere quando e come interpretare i pesi dei modelli è quella tra i modelli così detti “forward” e i modelli “backward”. La differenza principale fra i due modelli è che i primi esprimono i dati come funzioni di alcuni variabili, la seconda tipologia, invece, parte dalla risposta comportamentale e cerca di risalire da essa ai dati di riferimento. L'analisi dei pesi di un classificatore è un tipico modello di analisi backward che parte dall'output e cerca di risalire alle attivazioni neuronali presenti nei dati. La differenza sostanziale fra le due tipologie di modelli è che i modelli forward sono di solito interpretabili mentre i modelli backward non lo sono (Haufe et al 2014). Quello che il processo ideato da Haufe compie riguarda proprio la trasformazione dei modelli backward nei corrispondenti modelli forward rendendoli, così, interpretabili. Questo

procedimento ha senso esclusivamente sui modelli lineari apportando un sostegno all'utilizzo dei classificatori lineari nel processo di analisi di decodifica multivariata. Una volta ottenuti i pesi corretti è possibile indagare la mappa di attivazione ottenuta considerando, quindi, gli elettrodi (nel caso di un'analisi EEG) più significativi. Grazie a questa analisi è possibile individuare anche l'evoluzione spaziale delle attivazioni nel tempo. Infatti, reiterando l'intero procedimento di acquisizione dei pesi e correzione per ogni istante di tempo si ottiene una sequenza di mappe che indica l'attivazione di ogni canale nel tempo.

Capitolo II: Studio sull'inibizione di ritorno

Il seguente capitolo tratterà il fenomeno cognitivo oggetto dell'analisi MVPA, ovvero, l'inibizione di ritorno. Questo fenomeno è stato oggetto di numerosi dibattiti nel corso della storia. Successivamente verranno discusse alcune delle teorizzazioni più rilevanti nella comunità neuroscientifica (Berlucchi, 2006).

L'inibizione di ritorno

L'orientamento spaziale rappresenta una delle funzioni cognitive fondamentali nell'essere umano, che ha avuto un ruolo cruciale nel corso dell'evoluzione storica. Esistono due modalità principali di orientamento dell'attenzione nel campo visuo-spaziale: volontaria e involontaria. Nell'orientamento volontario, il soggetto dirige consapevolmente la propria attenzione verso una specifica area del campo visivo. Nell'orientamento involontario o riflessivo, uno stimolo esterno nel campo visivo "cattura" l'attenzione del soggetto senza il suo controllo diretto. Sia il controllo intenzionale dell'attenzione che quello riflessivo hanno svolto un ruolo essenziale per la sopravvivenza umana. Una maggiore capacità di selezionare attentamente le risorse visive consente di generare risposte comportamentali più efficaci e di migliorare la codifica delle informazioni rilevanti nella memoria. Allo stesso tempo, una rapida reattività all'orientamento involontario dell'attenzione permette agli individui di sfuggire a predatori o pericoli presenti nell'ambiente circostante (Klein 2000).

Con il termine "inibizione di ritorno" (IOR) ci si riferisce a un fenomeno cognitivo noto come effetto inibitorio dell'orientamento visuo-spaziale. In pratica, si osserva che le risposte ai bersagli visivi presenti in un'area già oggetto di attenzione sono più lente (Posner, 1984). Questo effetto ha suscitato un grande interesse nella comunità scientifica, stimolando numerosi studi volti a comprenderne le cause, gli effetti e le implicazioni neurali. Nel contesto dell'evoluzione funzionale, molti studiosi considerano l'IOR come un meccanismo che promuove l'esplorazione di nuovi spazi una volta che le aree precedentemente esplorate sono state considerate non rilevanti dal soggetto (Klein, 1999). L'IOR comporta infatti un'inibizione delle zone del campo visivo che sono state già esplorate, poiché l'attenzione è già stata spostata su un'altra regione.

Questa capacità di inibizione consente all'individuo di ottimizzare l'orientamento visuo-spaziale, favorendo l'esplorazione di nuovi stimoli e ambienti. È importante sottolineare che l'IOR non è un semplice effetto passivo, ma rappresenta un meccanismo attivo che contribuisce all'adattamento comportamentale dell'individuo. È possibile, infatti, comprendere l'enorme valore evolutivo dell'IOR confrontandola con il fenomeno complementare, ovvero l'effetto di facilitazione dell'attenzione nel processamento degli stimoli presenti nel campo visivo. Se l'attenzione è dispiegata su una determinata zona, il target stimolo che appare nello stesso punto godrà di un effetto facilitatorio, riducendo così il tempo di processamento. Se l'intervallo tra lo spostamento dell'attenzione e la comparsa dello stimolo aumenta eccessivamente, la stessa zona sarà soggetta ad un effetto inibitorio, in quanto l'attenzione è stata già disingaggiata e quindi quella stessa regione è stata indicata come non rilevante (Posner, 1985).

In sintesi, l'inibizione di ritorno è un fenomeno complesso che ha implicazioni sia a livello cognitivo che comportamentale. La sua comprensione richiede una visione integrata che consideri il ruolo dell'orientamento volontario e involontario dell'attenzione nell'ottimizzazione delle risorse cognitive. Continua a essere oggetto di ricerca approfondita per svelare i meccanismi neurali e i fattori che influenzano l'IOR, aprendo la strada a ulteriori studi e scoperte nel campo della percezione e dell'attenzione visiva.

Origini del fenomeno

Il fenomeno dell'inibizione di ritorno (IOR) è stato per la prima volta descritto negli studi di Posner e Cohen nel loro lavoro intitolato "*Components of visual orienting in Attention and Performance*" del 1984. Queste ricerche hanno apportato notevoli contributi allo studio dell'orientamento dell'attenzione, introducendo numerosi paradigmi e modalità di ricerca ancora ampiamente utilizzati nella letteratura contemporanea. Il paradigma sviluppato da Posner, conosciuto come Posner cueing task, nasce prima con l'obiettivo di indagare i processi di orientamento attentivo attraverso la misurazione dei tempi di reazione nella rilevazione di un semplice stimolo visivo. Questo compito di base prevedeva tre tipi di suggerimenti:

1. Suggerimento valido: il suggerimento indica correttamente la posizione in cui apparirà il bersaglio.
2. Suggerimento invalido: il suggerimento indica erroneamente una posizione diversa da quella in cui apparirà il bersaglio.
3. Suggerimento neutro: il segnale non indica alcuna posizione specifica.

I tempi di reazione di base venivano calcolati quando l'individuo doveva segnalare di aver visto il bersaglio dopo un suggerimento neutro. È importante notare che i tempi di suggerimento possono variare tra suggerimenti centrali e suggerimenti periferici, stimolando due diverse modalità di orientamento dell'attenzione. Nel primo caso, solitamente viene presentata una freccia al centro che indica il punto in cui potrebbe comparire lo stimolo, spingendo il soggetto a spostare intenzionalmente l'attenzione verso l'area del bersaglio. Nel secondo caso, invece, si verifica un cambiamento percettivo, come variazioni nell'illuminazione, direttamente nella zona suggerita, catturando l'attenzione del soggetto.

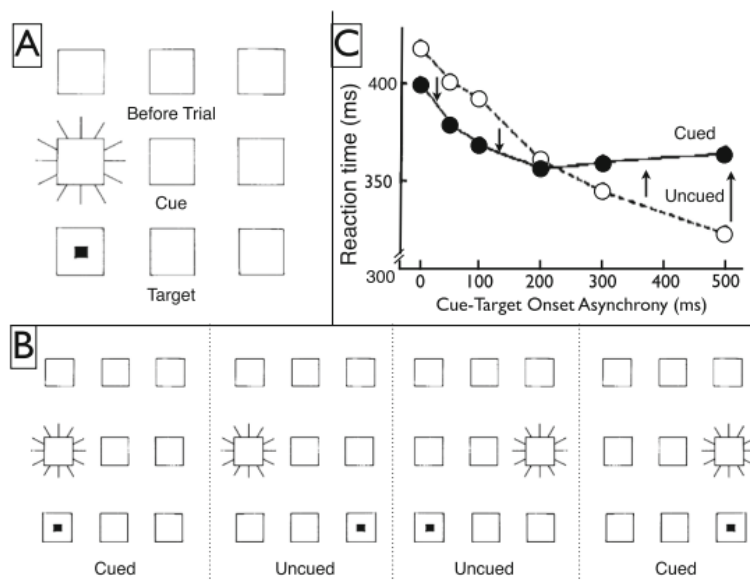


Figura 7. (A) La prima fase del paradigma inizia con una fila di tre quadrati, uno centrale e due periferici. In seguito, vi è la presentazione di uno stimolo suggerimento (seconda fila). E infine vi è la presentazione del target (terza fascia). (B) Nella figura B vengono mostrate le quattro possibili combinazioni suggerimento-target che danno origine a due condizioni: cued e uncued (prove valide e prove invalide). (C) Schema generico dei risultati: da questo grafico si nota l'inversione dell'effetto a seconda del CTOA, da effetto facilitatorio ad effetto inibitorio (immagine tratta da: Dukewich, Klein 2015).

Le prime applicazioni di questo paradigma si sono concentrate sull'indagine dell'effetto facilitatorio dell'attenzione nella rilevazione degli stimoli percettivi presentati nella stessa zona suggerita. Nel 1984, Posner e Cohen, attraverso l'applicazione di questo paradigma, hanno osservato che un aumento dell'CTOA (Cue-Target Onset Asynchrony), ovvero dell'intervallo di tempo tra il suggerimento e la presentazione del bersaglio, porta a un'inversione dell'effetto, che diventa inibitorio anziché facilitatorio (Bashinski, Bacharach, 1980; Cheal, Lyon, Hubbard, 1991; Eriksen, Collins, 1969; Posner, 1980).

Questa scoperta ha suscitato un grande interesse e ha aperto nuove strade di ricerca nel campo dell'orientamento dell'attenzione. Questa "Inibizione Ipsi-Laterale" (Berlucchi, Tassinari, Marzi, Di Stefano, 1989), "Effetto inibitorio alternativo" (Tassinari, Aglioti, Chelazzi, Marzi, Berlucchi, 1987) diventata nota con il nome di "Inibizione di Ritorno" è stata oggetto di approfondite indagini per comprendere i suoi meccanismi neurali, le sue influenze sul comportamento e le sue implicazioni per la percezione e l'elaborazione delle informazioni visive. In questi particolari esperimenti il paradigma classico viene leggermente modificato. Attraverso l'utilizzo di alcune procedure si cerca di scoraggiare l'individuo a mantenere l'attenzione sulla zona suggerita, in modo da poter favorire il disingaggio. Le metodologie utilizzate per ottenere questa facilitazione del disingaggio sono principalmente due: la prima consiste nel presentare lo stimolo bersaglio la maggior parte delle volte nella posizione centrale in modo da allenare l'individuo a spostare l'attenzione verso il centro quando viene presentato un suggerimento che indica una zona periferica; la seconda consiste nel far lampeggiare lo stimolo di fissazione (nella zona centrale) in modo da spingere uno spostamento dell'attenzione verso le zona di fissazione. Il punto di svolta nel mostrare l'effetto inibitorio piuttosto che il classico cueing effect di facilitazione è compreso fra i 200 e 300 millisecondi di CTOA. Quando infatti viene prolungato il tempo fra il suggerimento ed il bersaglio oltre questa soglia si manifestava l'IOR in tutte le prove sperimentali (Posner, 1984). Risultati simili, infatti, sono stati ottenuti utilizzando suggerimenti periferici di diversa natura percettiva (ad esempio l'oscuramento della zona selezionata piuttosto che l'illuminazione della stessa), suggerendo come questo fenomeno sia indipendente dalle modalità sensoriali di presentazione del suggerimento. Un elemento di fondamentale importanza emerso nei cinque esperimenti condotti in questo studio è l'esclusiva presenza dell'effetto di inibizione di ritorno con suggerimenti periferici. I cues centrali, infatti, non presentavano

alcun effetto (la presenza dell'IOR con suggerimenti centrali è stata ampiamente dimostrata in studi successivi), suggerendo il ruolo centrale dell'attenzione automatica (riflessiva o esogena) nell'IOR. Negli esperimenti di Posner e in successive ricerche è emerso come il meccanismo alla base dell'inibizione di ritorno opera su coordinate ambientali (legate all'ambiente esterno) e non su coordinate retinotopiche (legate alle posizioni dell'ambiente nella retina). L'effetto inibitorio, infatti, risulta presente anche quando vi è uno spostamento saccadico, ovvero, la zona in cui si manifesta questo effetto permane anche se cambiano le coordinate retinotopiche (dato dal movimento oculare; Posner and Cohen, 1984; Maylor and Hockey, 1985). Una conclusione teorica data dagli autori riguardava la concezione dello switching dell'effetto. Secondo Posner, infatti, non vi è un cambiamento del fenomeno ma piuttosto una sovrapposizione dell'effetto di facilitazione e di inibizione, entrambi i fenomeni compresenti. L'effetto cueing, però, risulta essere l'effetto principale in quanto presente sia in prove con suggerimento centrale (in cui viene stimolato lo spostamento intenzionale dell'attenzione) sia in prove con suggerimento periferico (orientamento involontario), occupandosi, quindi, della facilitazione del sottoprocesso dell'ancoraggio attentivo. L'effetto inibitorio, invece, presente solo in prove con suggerimento periferico punterebbe alla facilitazione del sottoprocesso di disancoraggio attentivo. La sovrapposizione dei due fenomeni, quindi, comporta l'impossibilità dell'effetto di inibizione di manifestarsi comportamentalmente, fino a CTOA in cui l'effetto facilitatorio non scompare. In seguito, le ricerche di Rafael e collaboratori dimostrano l'importanza del meccanismo di programmazione oculomotorio nella manifestazione del fenomeno dell'inibizione di ritorno. In una variante dell'esperimento di Posner sono stati utilizzate due tipologie di suggerimento (centrale e periferico). La differenziazione del paradigma consiste nelle richieste. Infatti, i soggetti venivano istruiti a compiere diverse risposte a seconda della condizione. Veniva richiesto di preparare un movimento saccadico, di compiere il movimento saccadico o semplicemente di rilevare il target. I risultati hanno dimostrato che nelle condizioni in cui è coinvolto il movimento oculomotorio (preparazione ed esecuzione della saccade) il fenomeno dell'inibizione di ritorno si manifesta anche con cue centrali, a differenza del semplice compito di detenzione del segnale che presenta l'effetto esclusivamente per l'orientamento automatico dell'attenzione.

Teoria della facilitazione nella ricerca

Una teoria successiva, proposta da Klein e collaboratori, propone l'effetto dell'inibizione di ritorno come un effetto "facilitatore del foraggio", ovvero, un processo, implicato in compiti di ricerca più complessi, che permette una ricerca visuo-spaziale più efficiente.

In linea con il pensiero di Treisman e Gelade, Klein propone un disegno sperimentale (Figura 8) in cui viene studiato l'IOR su compiti di ricerca visiva: ricerca seriale e ricerca pop-out. Secondo gli autori, infatti, le due tipologie di ricerca venivano compiute attraverso processi pre-attentivi (nella ricerca pop-out) e attentivi (nella ricerca seriale).

L'obiettivo di Klein, in questo particolare esperimento, è quello di sondare il ruolo del disimpegno dell'attenzione nell'IOR. Durante la ricerca seriale, infatti, l'attenzione si sposta sequenzialmente in un continuo processo di disancoraggio, spostamento e ancoraggio.

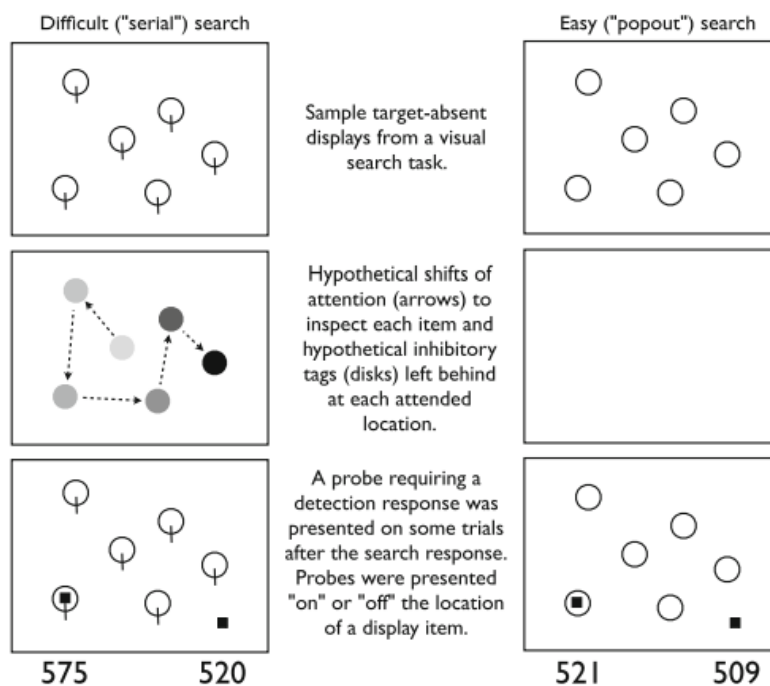


Figura 8. Nella prima colonna sono rappresentate le fasi di un compito di ricerca seriale mentre nella seconda un compito di ricerca pop-out. Nella prima scena di entrambe le colonne vengono mostrate le zone oggetti di ricerca, nella scena centrale viene ipotizzato il movimento dell'attenzione nello spostarsi da una zona all'altra (solo per la ricerca seriale), infine, nell'ultima fascia vengono mostrati i risultati con i tempi di reazione, che presentano una differenza significativa per la ricerca seriale e una differenza non significativa per la ricerca pop-out (immagine tratta da: Dukewich, Klein 2015).

Presentando il bersaglio target in una delle posizioni oggetto della ricerca seriale dovrebbe, quindi, manifestarsi un effetto inibitorio nelle prove con bersagli in posizioni segnalate in precedenza. Questo non dovrebbe succedere nel compito di ricerca pop-out, in quanto task non attentivo.

I risultati dello studio in questione sono in linea con le ipotesi iniziali. L'effetto inibitorio è stato osservato solamente nella condizione di ricerca seriale. Quindi, considerando come assodato che l'attenzione procede in maniera seriale, analizzando ogni distrattore nel momento in cui la ricerca risulta complessa (ad esempio quando la differenza fra target e distrattore si basa su caratteristiche non salienti), è possibile affermare che la difficoltà del compito è un fattore che influenza l'effetto inibitorio. In altre parole, la teoria di Klein suggerisce che l'inibizione del ritorno funziona come un meccanismo per migliorare l'efficienza della ricerca visiva. Questo effetto è stato osservato solo nei compiti di ricerca seriale che richiedono l'uso dell'attenzione per analizzare ogni elemento uno per uno. In questi compiti, l'attenzione si sposta continuamente tra gli elementi e il bersaglio target viene presentato con un effetto inibitorio rispetto ai bersagli nelle posizioni non segnalate in precedenza. Tuttavia, questo effetto non è stato osservato nei compiti di ricerca pop-out che non richiedono l'uso dell'attenzione per individuare il bersaglio target. In sintesi, i risultati suggeriscono che la difficoltà del compito influisce sull'effetto inibitorio dell'inibizione del ritorno.

Disancoraggio dell'attenzione e inibizione di ritorno

Nonostante la considerazione teorica generale sul fine dell'inibizione di ritorno sia stata accolta da gran parte della letteratura scientifica, le cause di questo fenomeno sono state oggetto di numerose critiche e dibattiti. Un nodo molto importante all'interno del progresso scientifico nella comprensione dell'IOR è stato la considerazione del ruolo del dis-ingaggio dell'attenzione come processo necessario e sufficiente per ottenere il fenomeno inibitorio. Nei quattro esperimenti dell'articolo del 2005 pubblicato da Berger e collaboratori in cui l'obiettivo era quello di studiare la relazione fra l'orientamento riflessivo e l'orientamento volontario dell'attenzione visuo-spaziale, sono state tratte

conclusioni di notevole importanza per quanto riguarda il ruolo delle due modalità di orientamento attentivo nel fenomeno inibitorio (Berger et al. 2005). Nel primo esperimento dei quattro presentati, un cue endogeno (una freccia che indicava la presentazione del target) era seguito da un cue esogeno (periferico). Il suggerimento centrale era predittivo della comparsa del target mentre il suggerimento periferico era non predittivo. Questo particolare paradigma permetteva di studiare i possibili effetti di interazione fra le due modalità di orientamento attentivo. I risultati presentavano i classici effetti studiati fino a quel momento: l'effetto di facilitazione nella posizione predetta e l'effetto di costo nella posizione non predetta da parte del suggerimento centrale, e l'effetto di facilitazione ed effetto di inibizione (a seconda del tempo fra il cue ed il target) nella posizione non predetta dei cue periferici. I risultati più importanti, però, riguardavano la relazione fra i due tipi di segnali. Gli effetti dell'orientamento endogeno e dell'orientamento esogeno erano completamente indipendenti gli uni dagli altri. Un ulteriore risultato emerso dalle analisi del compito contenente entrambe le modalità di orientamento attentivo consisteva nella manifestazione dell'effetto inibitorio nelle prove valide segnalate dal suggerimento centrale. Ovvero, l'IOR era presente nella zona in cui l'attenzione non era stata disingaggiata ma era ancorata alla posizione del target.

Numerosi studi successivi hanno presentato risultati molto simili, rivelando la manifestazione dell'effetto inibitorio in zone in cui l'attenzione, orientata selettivamente, non è stata disingaggiata (Maylor e Hockey, 1985; Pratt e Fischer, 2002; Chica, Lupiáñez 2004; Lupiáñez et al., 2004). In sintesi, questi studi suggeriscono che il dis-ingaggio dell'attenzione non è un processo necessario per ottenere il fenomeno inibitorio dell'inibizione di ritorno. Inoltre, gli effetti dell'orientamento endogeno ed esogeno sono stati trovati indipendenti gli uni dagli altri e l'effetto inibitorio è stato osservato anche quando l'attenzione non è stata disimpegnata ma era ancorata alla posizione del target.

Una seconda condizione riguardante il ruolo del disimpegno dell'attenzione nella manifestazione del fenomeno dell'inibizione di ritorno riguarda il concetto di sufficienza. Una volta dimostrato come il disimpegno possa essere non necessario affinché si manifesti l'IOR, è importante capire se esso sia quantomeno sufficiente, ovvero, se ogni volta che l'attenzione viene spostata dalla zona suggerita vi è un conseguente fenomeno inibitorio. In altre parole, la domanda è se il disimpegno dell'attenzione sia una condizione sufficiente per la manifestazione dell'IOR. Questo è un argomento importante

per comprendere appieno il ruolo del disimpegno dell'attenzione nell'IOR. Uno studio cardine nel determinare il dis-ingaggio come condizione sufficiente è stato condotto da Danzinger e Kingstone nel 1999. Questa ricerca ha portato risultati controversi ma con un grande impatto nella comprensione dell'IOR. Considerando l'ipotesi iniziale in cui l'effetto inibitorio era generato da un'inibizione dell'attenzione nel ritornare in posizioni precedentemente esplorate, una modalità per conoscere la connessione tra il distacco dell'attenzione e l'IOR consiste nel costringere l'attenzione a spostarsi rapidamente dalla zona suggerita in modo da rendere i CTOA, necessari al passaggio da un effetto facilitatorio all'effetto inibitorio, più brevi.

Il paradigma ideato da Danzinger e Kingstone, infatti, cercava di anticipare il disimpegno dall'area precedentemente oggetto dell'attenzione. Esso presentava quattro zone in cui poteva apparire lo stimolo target o suggerimento, disposti come da figura x. In una delle tre condizioni presenti nel primo esperimento, un suggerimento periferico in uno dei quattro riquadri risultava per la maggior parte delle volte predittivo della comparsa del target nel riquadro successivo (in senso orario). È stato riscontrato un effetto facilitazione nella zona predetta, che dimostrava come l'attenzione si fosse rapidamente disimpegnata dalla zona in cui era apparso il suggerimento per orientarsi nella zona predetta, e un effetto di costo presente quando il target appariva in riquadri diversi da quelli in cui vi era il suggerimento o la predizione del target. Un risultato molto importante riguarda la manifestazione dell'effetto dell'inibizione di ritorno nei riquadri in cui era stato presentato il cue. L'IOR era presente con intervalli di 50 millisecondi. Questi risultati sono stati molto spesso utilizzati per sottolineare l'importanza del processo di disimpegno attentivo nella manifestazione dell'IOR. In altre parole, lo studio di Danzinger e Kingstone ha dimostrato che quando l'attenzione viene rapidamente dis-ingaggiata dalla posizione segnalata, l'IOR può essere osservato anche a intervalli molto brevi (fino a 50 ms). Questo suggerisce che il disimpegno dell'attenzione sia un fattore importante per la manifestazione dell'IOR. Tuttavia, è importante notare che questi risultati sono stati oggetto di dibattito e controversie nella letteratura scientifica e ulteriori ricerche sono necessarie per comprendere appieno il ruolo del disimpegno dell'attenzione nell'IOR.

Risultati contrastanti, tuttavia, emergono nel secondo esperimento. Viene ripetuto lo stesso paradigma, con la stessa modalità di suggerimento (il cue periferico predice la comparsa del target nella posizione adiacente in senso orario). La differenza riguarda la

tipologia di compito. Nel primo esperimento il task riguardava un semplice compito di detenzione del segnale, nel secondo esperimento, invece, il task consisteva in un compito di discriminazione (discriminare tra lettere maiuscole presentate con l'orientamento standard e le stesse lettere ruotate di 180 gradi). I risultati hanno mostrato un effetto di facilitazione in entrambe le posizioni, quella suggerita dalla comparsa del cue periferico e quella predetta in quanto successiva alla posizione del cue. Questo dimostra evidenti differenze nella manifestazione dell'IOR a seconda della natura e complessità del compito. Questa discrepanza di risultati porta alla considerazione del disancoraggio come condizione non necessaria e nemmeno sufficiente per la manifestazione dell'IOR.

Inibizione di input e di output come due fenomeni distinti

Nel tentativo di spiegare la dissociazione dell'effetto dell'IOR presente solo in prove con stimolo di suggerimento esogeno, quindi periferico, in alcuni esperimenti e la presenza dell'effetto di inibizione con orientamento endogeno dell'attenzione in altri esperimenti, Taylor e Klein propongono una concezione dell'effetto IOR come un insieme di effetti diversi. Una problematica riscontrata nella letteratura in merito al fenomeno dell'IOR è emersa nello studio del 2015 di Dukewich e Klein. Attraverso la somministrazione di questionari a ricercatori impegnati nello studio del fenomeno inibitorio dell'attenzione è apparsa una chiara concezione frammentata delle cause sottostanti all'IOR. I risultati del questionario condotto da Dukewich e colleghi, infatti, presenta una notevole discrepanza fra la visione del costrutto teorico come un insieme di diversi fenomeni e la concezione dell'IOR come fenomeno unitario presente nelle numerose ricerche. Circa l'80 %, ha affermato di aver letto un articolo che discuteva l'effetto dell'inibizione di ritorno visuo-spaziale e di non considerare il fenomeno presentato come IOR. Una delle possibili interpretazioni che gli autori attribuiscono a questa discordanza tra i ricercatori risiede nel pensiero comune che l'inibizione di ritorno sia un fenomeno unico. Infatti, l'aver studiato l'IOR senza considerare la composizione stessa multifattoriale del fenomeno ha portato a numerosi risultati contrastanti.

Secondo Taylor e Klein, infatti esistono due forme di inibizione che agirebbero in due momenti distinti dell'elaborazione delle informazioni. La forma inibitoria di input agisce sulla qualità degli input processati, la forma inibitoria di output, invece, agirebbe in una

fase post-percettiva dell'elaborazione dell'informazione. La chiave nel comprendere la differenza tra quello che gli autori identificano come effetto di input e come effetto di output, secondo Taylor e Klein, risiede nel movimento oculomotorio. Secondo gli autori Redden e Hurst (2018), la tipologia del fenomeno di inibizione del ritorno che si presenta dipende dallo stato di attivazione del sistema oculomotore riflessivo. In particolare, i dati provenienti da ricerche che utilizzano paradigmi che indicano bersagli di luminosità periferici suggeriscono la forma di inibizione di input (di tipo attentivo/percettivo). D'altra parte, i dati provenienti da paradigmi con stimolazione dell'orientamento endogeno dell'attenzione attraverso cues centrali indicano una forma di inibizione di output (a un livello di risposta/decisione). Diversi studi forniscono prove a sostegno dell'ipotesi dissociativa dell'inibizione del ritorno, suggerendo che esistono due forme distinte di inibizione del ritorno che dipendono dallo stato di attivazione del sistema oculomotore riflessivo (Hilchey et al 2016).

Correlati neurofisiologici dell'inibizione di ritorno

Dalla scoperta del fenomeno dell'inibizione di ritorno (Posner, Cohen 1984), sono state condotte numerose ricerche attraverso diverse tecniche di neuroimmagine al fine di scoprire i correlati di questo fenomeno. Per una migliore esplorazione delle basi neurali dell'IOR sono stati impiegati un'ampia gamma di strumenti neuroscientifici.

I primi risultati sono stati ottenuti attraverso lo studio di pazienti con lesioni cerebrali. Infatti, ricerche su soggetti con danni sottocorticali dovuti a “paralisi sopra-nucleare progressiva” e soggetti con differenti lesioni corticali hanno dimostrato l'importanza dei sistemi sottocorticali, a discapito di sistemi corticali, nella manifestazione dell'inibizione di ritorno. Tra le strutture sottocorticali è stata ampiamente confermata l'importanza del collicolo superiore (SC), struttura al centro del controllo del sistema oculomotore. Grazie a studi di pazienti con lesioni localizzate al collicolo superiore, questa struttura è stata confermata come determinante nell'espressione dell'inibizione. Successivamente è stata dimostrato il coinvolgimento di alcune regioni corticali nella codifica delle coordinate ambientali e dell'oggetto. Nello specifico un esperimento di Sapir e colleghi del 2004 ha dimostrato l'importanza dei circuiti parietali nella conservazione delle coordinate ambientali delle etichette inibitorie stabilite dal collicolo superiore.

Un interessante risultato negli studi riportati da Bourgeois propone una dissociazione nella modalità di risposta. Pazienti con lesioni all'emisfero destro sono stati sottoposti ad un compito di attenzione visuo-spaziale con il classico paradigma *target-target* (ogni stimolo presentato deve essere riportato, quindi, non esiste uno stimolo suggerimento ma solo stimoli bersaglio) con due condizioni, la prima richiedeva risposte saccadiche e la seconda risposte manuali. Nella prima condizione è stato osservato l'IOR. Quando veniva richiesto dal compito di effettuare risposte manuali, invece, i soggetti non presentavano inibizione ma solo effetti di facilitazione. Un'interpretazione derivante da questi risultati, richiama la dissociazione dei correlati neurali a seconda della richiesta del compito. L'effetto inibitorio, quindi, sarebbe connesso ad un circuito corticale (fronto-parietale dell'emisfero destro) quando è richiesta una risposta manuale, mentre, dipenderebbe da circuiti sottocorticali nel caso di risposte saccadiche (Bourgeois et al. 2012). Oltre a studi su soggetti con lesioni cerebrali sono stati molto comuni anche studi di *brain imaging*, nello specifico le analisi più presenti in letteratura riguardano lo studio sui potenziali evento-relati. Martin-Arevalo e colleghi hanno esaminato l'uso dei potenziali evento-relati (ERP) per studiare l'IOR e hanno concluso che non esiste un ERP specifico che possa essere utilizzato come marcatore dell'IOR (Martin-Arevalo 2016).

Una delle difficoltà nell'utilizzo degli ERP per studiare l'IOR è la contaminazione dei dati da parte del movimento degli occhi. Molti studi sono stati scartati a causa di questa contaminazione. Questa problematica diventa ancora più rilevante quando si considera la dissociazione dell'effetto inibitorio. Eliminare la possibilità di studiare i potenziali di attivazione dell'IOR utilizzando anche il movimento oculare la considerazione dell'IOR esclusivamente come effetto di input (Satel, Wilson 2019). In sintesi, l'utilizzo degli ERP per studiare l'IOR presenta alcune difficoltà, in particolare la contaminazione dei dati da parte del movimento degli occhi. Tuttavia, la ricerca continua per comprendere meglio questo fenomeno e trovare modi per superare queste difficoltà. Una componente successiva dell'ERP, chiamato Nd, che si verifica tra 220-300 ms dopo il segnale, risulta essere spesso influenzata dalla segnalazione nei paradigmi IOR (Martín-Arévalo). Questo effetto è stato osservato non solo quando i movimenti degli occhi sono proibiti e c'è una stimolazione periferica ripetuta, ma anche quando i movimenti degli occhi sono stati fatti in risposta a frecce centrali (Satel, et al 2014). In altri esperimenti (McDonald et al. 2009; Pierce et al. 2017) è stato esplorato come l'effetto di inibizione di ritorno moduli

l'ampiezza della componente N2pc. È stato dimostrato, infatti, che la manifestazione del fenomeno inibitorio comporti una riduzione dell'ampiezza della componente N2pc in compiti di detenzione semplice o di ricerca visiva. I risultati delle numerose ricerche risultano, tuttavia, ancora inconsistenti

Capitolo III: Descrizione mvpalab toolbox

Come già ampiamente illustrato in precedenza, l'analisi dei pattern multivariati fornisce strumenti che permettono di aumentare la sensibilità e l'efficienza di analisi dei dati di neuroimaging. Con l'aumento della considerazione da parte della letteratura per il decoding multivariato sono nati diversi strumenti per semplificare il processo di analisi. La costruzione di una pipeline di MVPA senza l'aiuto di toolbox risulta spesso complessa soprattutto per i ricercatori con poche conoscenze di programmazione. La difficoltà nella costruzione di analisi, come già affermato, è stata una delle barriere che ha portato un rallentamento nell'utilizzo sempre più comune delle tecniche di decodifica multivariata. Negli ultimi anni, sono stati creati molti toolbox MVPA, come "SPM" (Penny, Friston 2011), "The Decoding Toolbox" (TDT; Hebart et al. 2015), "Pattern Recognition for Neuroimaging Toolbox" (PRoNT; Schrouff, Rosa 2013), specificamente progettati per studi fMRI. Un'importante applicazione del decoding multivariato sviluppatasi negli ultimi anni riguarda la Time-resolved MVPA, ovvero l'analisi multivariata dei pattern a risoluzione temporale. Come già descritta nei capitoli precedenti, la Time-resolved MVPA è un tipo di analisi possibile grazie all'utilizzo di strumenti di neuroimaging ad alta risoluzione temporale, quali MEG e EEG. Come per la risonanza magnetica funzionale, anche per l'analisi a risoluzione temporale è stata favorita dall'avvento di numerosi toolbox specifici. Alcuni esempi includono: "The Amsterdam Decoding and Modeling Toolbox" (ADAM; Fahrenfort, Driel 2018), "CoSMoMVPA" (Oosterhof, Connolly 2016), "MVPA-light" (Treder 2020), "The Decision Decoding Toolbox (DDTBOX; Bode et al. 2019)". Oltre alla nascita di nuovi toolbox nell'ambiente applicativo di Matlab sono nati anche alternative open source in Python come: MNE-Python (Gramfort et al, 2013), Nilearn (Abraham et al. 2014) e PyMVPA (Hanke et al. 2009).

3.1 MvpaLab

Nonostante l'avvento dei toolbox, l'analisi di decodifica, per quanto notevolmente semplificata, continua a risultare ostica per studenti e/o ricercatori con poca esperienza negli ambienti di Python o Matlab. Un toolbox che permette di ovviare a questa problematica è "Mvpalab" (López-García, Peñalver 2021). Grazie all'implementazione

di un interfaccia grafica (GUI), infatti, il toolbox offre la possibilità di costruire una pipeline per l'analisi di decodifica, senza dover utilizzare alcun linguaggio di programmazione. L'interfaccia risulta molto semplice da utilizzare con un design intuitivo, permettendo l'implementazione di diverse analisi di decoding. La semplicità nella programmazione della pipeline di ricerca e l'intuitività dei comandi nella schermata di interfaccia differenziano questo toolbox da tutti gli altri. Mvpalab permette l'implementazione di diverse funzionalità di decodifica, tra cui la Time-resolved MVPA. All'interno della GUI è possibile selezionare diverse caratteristiche, nella prima sezione della schermata di configurazione dell'MVPA c'è la possibilità di poter selezionare le configurazioni generali.

Nella prima parte è possibile inserire i dati divisi per classe e per soggetto (nel caso in cui si voglia procedere per un analisi between subject). Un elemento molto importante riguarda la possibilità di convertire i dati ERP provenienti da toolbox quali EEGLab. Per quanto concerne la compatibilità dei dati provenienti da altri strumenti di indagine, all'interno della wiki, è presente una guida dettagliata per poter convertire i dati in file compatibili con il toolbox. Anche qui, la semplicità di conversione dei dati e la chiarezza delle istruzioni definiscono Mvpalab uno dei migliori toolbox per neofiti nel campo della programmazione. Una volta inseriti i dati, l'interfaccia grafica presenta diversi riquadri che permettono le applicazioni di differenti paradigmi di analisi.

I riquadri presentati riguardano:

- Una configurazione generale dell'analisi
- Specifiche per la Temporal Generalization
- Specifiche del modello di classificazione
- Calcolo delle statistiche di significatività
- Filtri

3.2 Configurazione Generale

Nella primo riquadro dell'interfaccia vengono configurati gli aspetti generali dell'analisi. In questa sezione risulta importante specificare il tempo di inizio e di fine e soprattutto il "passo" dell'analisi. Di default, mvpalab, esegue l'analisi per ogni specifico istante di tempo. Questa modalità, tuttavia, oltre che apportare un peso computazionale consistente è spesso superflua. La scelta del passo di analisi è strettamente legata alla domanda di ricerca e al processo cognitivo studiato, ma molto spesso si rivela inutile studiare cambiamenti dell'ordine dei millisecondi in quanto risulta essere un periodo eccessivamente breve. Nella sezione successiva invece è possibile configurare l'estrazione di features e l'averaging dei trials. Qui, infatti, è possibile selezionare il numero di trials da mediare nel tentativo di ridurre il rumore presente nei dati. Come già descritto nei capitoli precedenti l'averaging permette una riduzione del rumore ma al costo di una riduzione del numero di trials.

The image shows a software interface for configuring analysis parameters. It has a top navigation bar with tabs: 'General configuration', 'Temporal generalization', 'Classification model', 'Statistics', 'Sliding filter', and 'Configure analysis'. The 'Configure analysis' tab is active. Below the tabs are six numbered sections:

- 1. Analysis timing:** Start time: 0 ms, Final time: 0 ms, Step size: 1 timepoints.
- 2. Feature extraction:** Features: Raw voltage (dropdown). Trial average and balanced datasets: Number trials to average: 0 (spinner), Randomize trials, Match class size, Match kfold.
- 3. Dimensionality reduction:** Method: None (dropdown), No. of components: 0 (spinner).
- 4. Data smoothing:** Method: No smoothing (dropdown), Length: 1 timepoints.
- 5. Data normalization:** Normalization method: None (dropdown).
- 6. Electrode selection:** Indexes: default: all electrodes selected (text input), *Type the indexes of the selected electrodes (e.g., 1 3 5 7 9 11 13)*

Figura 9. (immagine estratta dalla wiki del toolbox: <https://github.com/dlopezg/mvpalab/wiki>)

Oltre a questo, è possibile decidere se randomizzare il numero di trials per ogni soggetto e bilanciare le classi. È ormai noto, infatti, quanto classi sbilanciate (quando nei dati vi sono molte più prove di una condizione rispetto ad un'altra) comportino risultati distorti. Il bilanciamento dei dati avviene attraverso la riduzione della categoria con più trials in

modo da riportare un numero equilibrato di prove per condizione. È possibile, inoltre, bilanciare i singoli fold della cross-validation.

Una sezione molto importante nella configurazione generale riguarda la riduzione della dimensionalità. Pratica ormai molto diffusa nel machine learning e nelle neuroscienze, la riduzione delle dimensioni permette numerosi vantaggi (vedi capitolo 1) tra cui una riduzione dei tempi computazionali, eliminazione di dati superflui e migliore performance del modello. Tuttavia, come già specificato nei capitoli precedenti l'utilizzo di tecniche di riduzione delle features presenta un notevole costo sull'interpretabilità dei risultati. Nel toolbox `mvpalab` è disponibile l'utilizzo della Principal Component Analysis (PCA), per la riduzione della dimensionalità, specificando, inoltre, il numero di features risultanti. Nella sezione successiva, invece, è possibile configurare lo smoothing dei dati. Questa procedura permette di inserire nei dati del modello istanti di tempo che non sarebbero stati considerati per via del "passo" scelto dell'analisi. Nella Time-resolved MVPA, è possibile, infatti, non solo selezionare istanti di tempo con un passo selezionato, ma attraverso lo smoothing si può selezionare una finestra di tempo e attraverso la media dei valori riportarla ad un istante. La sezione dedicata allo smoothing permette di scegliere la grandezza della finestra di tempo da considerare e di scegliere anche la metodologia di smoothing ("time window", "Kernel Gaussiano"). In seguito, è possibile selezionare le quattro tipologie di normalizzazione dei dati: attraverso le features, istanti di tempo, trials, cross-validazione. Infine, l'ultima sezione permette di attuare una selezione degli elettrodi per l'analisi, ovvero, di scegliere una regione di interesse (ROI). L'opzione di default nella selezione degli elettrodi dell'interfaccia è quella di non attuare nessuna selezione.

3.3 Specifiche per la Temporal Generalization

Nel riquadro specifico per la Temporal Generalization analysis è possibile configurare gli istanti di tempo in modo da ottenere una matrice specifica per una determinata finestra temporale. L'estrazione della diagonale corrisponde alla distribuzione delle accuratèzze ottenute per l'analisi generale.

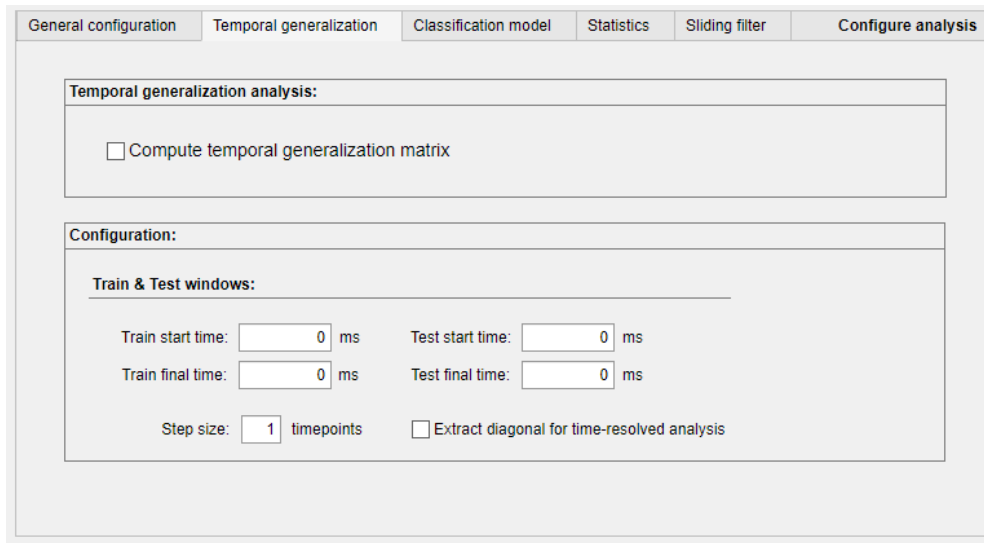


Figura 10. (immagine estratta dalla wiki del toolbox: <https://github.com/dlopezg/mvpalab/wiki>)

3.4 Specifiche del modello di classificazione

Il riquadro contenente le specifiche sul modello di classificazione permette di personalizzare non solo la scelta del classificatore ma numerosi parametri relativi alla convalida incrociata. Nella prima sezione è possibile, infatti, scegliere tra due algoritmi di classificazione: La support vector machine e la linear discriminant analysis (vedi capitolo 1). Oltre a scegliere l'algoritmo di classificazione, nella stessa sezione è possibile cambiare il kernel. La scelta del kernel presenta un elemento fondamentale nel contesto dell'interpretazione dei risultati, come specificato nel primo capitolo. In questa tipologia di analisi, infatti, è sempre indicato utilizzare un kernel lineare in modo da mantenere intatte le potenzialità interpretative dell'analisi. Nelle specifiche riguardanti la cross-validation, invece, è possibile selezionare una modalità k-fold o leave one out (vedi capitolo 1). Nella procedura K-fold è richiesta la specifica dei gruppi in cui verranno suddivisi i dati nella convalida incrociata. Per la procedura leave-one-out, invece, il numero di fold sarà automaticamente identificato con il numero di trials presenti. Nella sezione riguardo l'ottimizzazione è possibile includere una modalità di configurazione di default del toolbox nel tentativo di migliorare gli Iper-parametri del modello (nel caso della SVM lineare, "c" è l'unico parametro e consiste nel "costo dell'errore", più il parametro è basso e meno sarà il costo dell'errore di classificazione nella fase di training, quindi ci sarà meno modellazione del classificatore sui dati di addestramento e meno

probabilità di overfitting). La sezione degli output permette un'ampia scelta nelle metriche relative alla performance del modello. La metrica di default è “l'accuratezza media” che consiste nel rapporto fra il numero di successi e numero di fallimenti del modello nella fase di test (la performance può anche essere calcolata nella fase di training per valutare possibili effetti di overfitting). Questa metrica risulta molto semplice ed intuitiva ed è la misura della performance più utilizzata nel campo delle neuroscienze cognitive. Tuttavia, dati non bilanciati comportano prestazioni distorte, l'accuratezza media per quanto molto efficace non permette un'ottimale comprensione del fenomeno in queste situazioni. Una misura molto utile da affiancare all'accuratezza media in caso di problemi di distribuzione fra classi è la “matrice di confusione”.

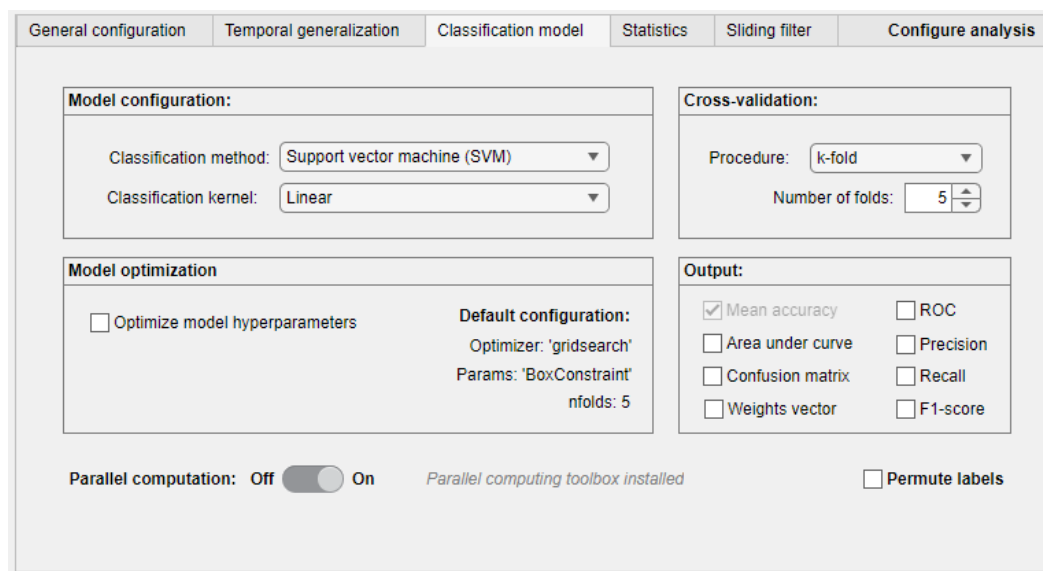


Figura 11. (immagine estratta dalla wiki del toolbox: <https://github.com/dlopezg/mvpalab/wiki>)

Questa particolare metrica permette di confrontare il numero di previsioni corrette ed errate per ogni classe. In una matrice di confusione sono presenti quattro valori: veri positivi, veri negativi, falsi positivi e falsi negativi. Dalla presentazione della matrice, l'informazione più importante non riguarda la semplice performance che racchiude gli errori di classificazione ma la tipologia di errore. Ogni colonna della matrice rappresenta i valori predetti, mentre ogni riga rappresenta i valori attuali. Nel confronto fra queste due misure è possibile quindi comprendere la confusione presente nelle classificazioni spesso

sbilanciate. Tra le altre metriche vi sono: Precisione (Precision), una misura della percentuale di previsioni positive corrette fatte dal modello rispetto a tutte le previsioni positive. Punteggio F1 (F1 Score), una misura che combina precisione e richiamo (recall) per fornire un equilibrio tra le due, Curva ROC (Receiver Operating Characteristic) una rappresentazione grafica delle prestazioni di un modello di classificazione binaria. Una curva ROC ideale si avvicina al punto in alto a sinistra del grafico, indicando una sensibilità elevata e un basso tasso di falsi positivi e, infine, Area Under the Curve (AUC), una metrica utilizzata per valutare le prestazioni di un modello di classificazione binaria basato sulla curva ROC.

Un'ultima metrica di analisi disponibile nel riquadro differisce da tutte le altre, in quanto non connessa alla performance del soggetto. La metrica "weights vector" permette di salvare per ogni istante il vettore dei pesi del classificatore fondamentale per l'analisi della proiezione dei pesi (vedi capitolo 1). Il vettore dei pesi per ogni istante di tempo è il risultato della media dei pesi raccolti per ogni cross-validazione. Una 3 fold cross-validation, ad esempio comporta l'addestramento di 5 modelli differenti per ogni rimescolamento dei gruppi di dati. Il toolbox `mvpalab` raccoglie il vettore dei pesi di ognuno di questi modelli per poi ottenere il vettore risultante ottenuto dalla media delle cinque ripetizioni.

Come descritto nel capitolo 1, il vettore dei pesi presenta una scarsa capacità interpretativa. Infatti, per poter ottenere risultati attendibili riguardo le attivazioni delle features bisogna trasformare il vettore dei pesi grezzi in "pesi corretti" (vedi capitolo 1, Haufe 2014). L'opzione "Weights vector" di `mvpalab` permette di ottenere al termine dell'analisi due file contenenti il vettore dei pesi grezzi per ogni istante di tempo e il vettore connesso dei pesi corretti. Questa funzione, assieme alle altre, permette una tangibile semplificazione dell'intero processo di analisi. Un'opzione molto utile riguarda la possibilità di conseguire operazioni computazionali in parallelo. Questa opzione permette di ridurre drasticamente i tempi di computazione affiancandosi ad un toolbox preesistente ("Parallel computation toolbox")

3.5 Calcolo delle statistiche di significatività

Per quanto riguarda il riquadro delle statistiche invece, si preoccupa della parte dell'analisi relativa al calcolo della significatività. Permette di scegliere il tipo di ipotesi (a una coda o a due code) e il numero di permutazioni per ogni soggetto o gruppo. Infine, permette di scegliere una soglia di significatività (a seconda dell'alfa scelto).

Il calcolo della significatività all'interno del toolbox `mvpalab` avviene attraverso un approccio di permutazione non parametrico basato su cluster. Gli autori si sono ispirati al lavoro di Stelzer e collaboratori nel calcolo della significatività per dati di risonanza magnetica (Stelzer et al. 2013) adattandolo su dati M\EEG. Attraverso un numero definito di permutazioni (100 permutazioni di default) viene calcolata l'accuratezza nulla (chance level) per ogni soggetto ed istante di tempo. Successivamente, viene estratta, in maniera casuale, una delle mappe di accuratezze permutate per ogni soggetto. Queste mappe vengono mediate tra i soggetti. Questa procedura viene ripetuta per 105 volte in modo da ottenere 105 mappe di accuratezza di gruppo. In seguito, per ogni istante, viene calcolata la distribuzione casuale dei valori e determinate le soglie/a (a seconda dell'ipotesi e dell'intervallo di confidenza scelto). Infine, vengono raccolti i cluster di punti temporali che superano la soglia e sottoposti ad un metodo di correzione, nello specifico, il "False Discovery Rate" (FDR). Il toolbox `mvpalab` presenta, inoltre, diverse altre funzioni tra cui una sezione per il filtraggio dei dati M\EEG e altre tipologie di analisi come la MultiVariate Cross-Classification che in quanto non utilizzate per l'analisi non verranno approfondite in questo testo. Un'ultima funzione che offre questo toolbox è una seconda interfaccia grafica che segue gli stessi principi di intuitività e semplicità di utilizzo applicata alla rappresentazione grafica dei risultati. Nello specifico dell'analisi trattata in questo elaborato, è stata utilizzata la funzione "`mvpalab_plottempogen()`" per la creazione della matrice grafica della temporal generalization analysis.

La rappresentazione grafica della distribuzione delle accuratezze è stata implementata su python attraverso le librerie "`matplotlib`" e "`seaborn`". La rappresentazione topografica dei pesi del classificatore, invece, è stata ottenuta attraverso la funzione "`topoplot()`" del toolbox `EEGLab`. In conclusione, il toolbox `mvpalab` si presenta come uno strumento intuitivo e flessibile, permettendo l'applicazione di una diversa gamma di analisi all'interno del decoding multivariato.

The image shows the 'Configure analysis' tab of the mvpalab toolbox. It features a navigation bar at the top with tabs for 'General configuration', 'Temporal generalization', 'Classification model', 'Statistics', 'Sliding filter', and 'Configure analysis'. The main content area is divided into three sections:

- Cluster-based permutation analysis:** Contains a checkbox labeled 'Compute cluster-based permutation analysis' and a dropdown menu labeled 'Select analysis:' with the selected option 'Above and below chance - Two tails'.
- Permuted maps:** Contains two spinners: 'Number of permuted maps (per subject):' set to 100 and 'Number of permuted maps (group level):' set to 1e+05. It also has a checkbox labeled 'Plot null distribution'.
- Percentile thresholds:** Contains two percentage input fields: 'Group significance level:' set to 95% and 'Cluster significance level:' set to 95%.

Figura 12. (immagine estratta dalla wiki del toolbox: <https://github.com/dlopezg/mvpalab/wiki>)

La presenza di un ambiente flessibile permette di massimizzare l'utilità del toolbox consentendo un semplice utilizzo a studenti e neofiti attraverso l'utilizzo dell'interfaccia grafica ma offrendo numerosi strumenti avanzati per esperti attraverso l'implementazione della pipeline attraverso la configurazione manuale. Inoltre, l'applicazione di strumenti complementari (grazie al contributo di toolbox esterni) quali analisi di frequenza e computazioni in parallelo rendono mvpalab un ottimo ed efficace strumento per le analisi di decoding multivariato.

Capitolo IV: Applicazione della multivariate pattern analysis

L'analisi trattata in questo elaborato di tesi riguarda l'applicazione della decodifica multivariata su dati ottenuti da un disegno sperimentale condotto dalla ricercatrice Fu Wanlu (Fu, 2023). In questo capitolo verranno descritte le metodologie e i paradigmi utilizzati per studiare l'inibizione di ritorno e i risultati ottenuti attraverso un'analisi univariata dei dati comportamentali e dei potenziali ERP, riportati in Fu 2023 (Fu, 2023). In seguito, verrà presentato nel dettaglio il percorso di analisi di decodifica partendo dai dati dell'esperimento di Fu. L'analisi dei pattern multivariati presentata in questo elaborato di tesi è stata effettuata utilizzando due sistemi diversi: Python e Matlab. In Python, sono state utilizzate diverse librerie, tra cui numpy, pandas, matplotlib, seaborn e sklearn, per eseguire l'analisi. In Matlab, invece, è stato utilizzato il toolbox "Mvpalab" per lo stesso scopo.

4.1 Descrizione dell'esperimento

L'esperimento è tratto dal progetto di dottorato eseguito dalla dottoressa Fu Wanlu e collaboratori, condotto nel Dipartimento di Psicologia Generale dell'Università di Padova (DPG). L'oggetto di questo studio consiste nell'indagare il fenomeno dell'inibizione di ritorno in relazione ad altre abilità quali: l'abilità di calcolo e l'abilità di lettura. Tuttavia, le relazioni del fenomeno dell'IOR con le altre abilità non sono state considerate nell'analisi multivariata dei pattern e, di conseguenza, non verranno approfondite in questo elaborato.

Metodologia

In questo studio, 30 partecipanti italiani, di cui 10 maschi e 20 femmine, con un'età compresa tra i 18 e i 29 anni, sono stati sottoposti a diverse valutazioni per testare le loro abilità cognitive. In primo luogo, sono state valutate le loro abilità di lettura e di calcolo attraverso dei test cartacei. Questi test hanno fornito una misura di base delle capacità dei partecipanti in queste aree. Successivamente, i partecipanti sono stati sottoposti al paradigma di cueing spaziale di Posner (vedi paragrafi precedenti). Durante questo compito, sono state rilevate le registrazioni elettroencefalografiche (EEG) per monitorare

l'attività cerebrale dei partecipanti. Inoltre, i partecipanti hanno eseguito un compito numerico all'inizio del periodo di registrazione EEG ed un compito di tipo fonologico al termine della sessione. Questi compiti aggiuntivi hanno fornito ulteriori informazioni sulle capacità cognitive dei partecipanti

Descrizione del paradigma sperimentale

In questo esperimento, il fenomeno dell'inibizione di ritorno è stato misurato attraverso l'applicazione del paradigma "spatial cueing task" (Posner & Cohen, 1984), nello specifico è stata utilizzata la versione verticale (figura 13). Questo compito è stato presentato attraverso uno schermo da 19 pollici con una risoluzione di 1024 x 768, funzionante a 60 Hz. La distanza presente fra lo schermo, attraverso cui veniva sottoposto il compito e ogni soggetto era di 62 cm. Il programma del paradigma è stato interamente scritto in Python.

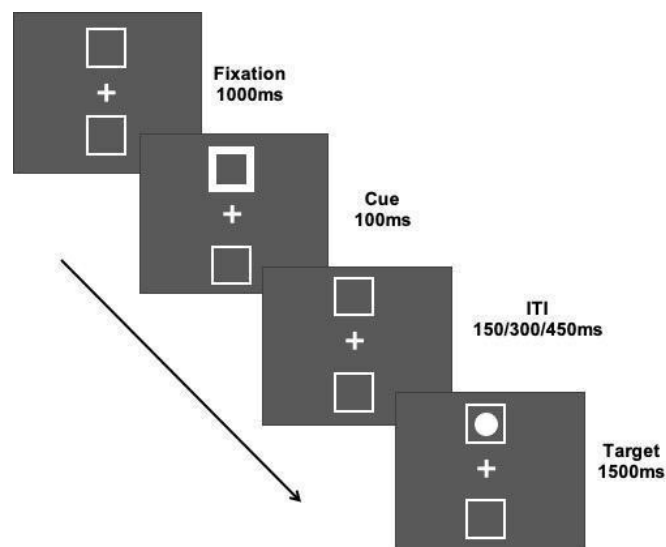


Figura 13. Descrizione del paradigma sperimentale dell'esperimento di Fu 2023. Nella primo frame sono presenti i due riquadri posizionati verticalmente e il punto di fissazione centrale. In seguito, vi è la comparsa dello stimolo suggerimento periferico, successivamente vi è un intervallo e nel quarto frame, infine, il target compare, in questo caso nella posizione suggerita. (immagine tratta da Fu 2023)

La versione verticale del paradigma di Posner consiste nella disposizione delle due caselle rispettivamente una all'interno dell'emi-campo superiore e l'altra all'interno dell'emi-

campo inferiore. Come punto di fissazione centrale è stata disposta una croce grigia al centro dello schermo. La distanza fra il punto di fissazione centrale e le due caselle era di nove gradi. Nel paradigma presentato, quindi, una volta esposta la croce del punto di fissazione centrale per 1000 millisecondi, appariva il suggerimento in uno dei due riquadri (superiore o inferiore), per circa 100 millisecondi. In seguito, il bersaglio appariva con un intervallo (CTOA) diverso a seconda della condizione. Le tre condizioni presentavano un intervallo di 150 ms, 300 ms e 450 ms, dalla scomparsa dello stimolo bersaglio. È importante sottolineare che le tre diverse condizioni erano sottoposte in maniera casuale; quindi, gli intervalli tra lo stimolo di suggerimento e lo stimolo target seguivano un ordine casuale per l'intera durata dell'esperimento. La durata di presentazione dello stimolo target poteva scomparire una volta data la risposta o permanere fino ad una scadenza di 1500 millisecondi. Il suggerimento era un cue di tipo periferico, volto all'attivazione dell'orientamento involontario o riflessivo dell'attenzione. Il cue, infatti, consisteva nell'illuminarsi o lampeggiare del bordo di uno dei due riquadri. Lo stimolo target, invece, corrispondeva ad un cerchio bianco pieno.

La risposta allo stimolo, nel compito di detenzione del segnale, avveniva attraverso l'utilizzo della barra spaziatrice in una semplice tastiera QWERTY. Le condizioni presenti nel paradigma erano di due tipi:

- Prove invalide, quando lo stimolo target appariva in una casella diversa da quella segnalata dallo stimolo di suggerimento
- Prove valide, quando lo stimolo target veniva presentato nella stessa posizione dello stimolo di suggerimento

Il numero di condizioni (valide e invalide) era bilanciato per ogni partecipante. Il numero totale di prove quindi era di 280, di cui, 40 prove senza stimolo target, al fine di scoraggiare tentativi di risposte anticipatorie e 120 prove per ciascuna condizione (Valida o Invalida).

Registrazione dei dati EEG

La registrazione dei segnali elettroencefalografici è avvenuta all'interno di una stanza schermata elettricamente e con una acustica ridotta. È stato utilizzato una cuffia elastica (actiCAP; Brain Products, Gilching, Germania) con 64 elettrodi preamplificati montati secondo il sistema internazionale 10-20. I dati sono stati archiviati utilizzando il sistema BrainVision Recorder (Brain Products GmbH, Germania). La frequenza di campionamento utilizzata era di 100 Hz, l'impedenza, invece, è stata mantenuta sotto i 5 k Ω . Gli elettrodi corticali erano online riferiti a FCz.

Analisi dei dati comportamentali

Per quanto riguarda l'analisi dell'effetto comportamentale dell'inibizione di ritorno è stata prima svolta una pre-elaborazione dei dati. È stato controllato il tasso di falsi allarmi, ovvero, il tasso di risposte in prove in cui non vi era il bersaglio (2,15%) e il tasso di prove mancanti, ovvero, quando non vi era alcuna risposta in prove con lo stimolo target presente (0,05%). I tempi di reazione, in seguito, sono stati calcolate le prestazioni medie di ogni soggetto nel compito di detenzione del segnale e catalogate per ogni condizione (dipendente dal CTOA) e per ogni prova (valida e invalida). È stata, successivamente condotta un'analisi preliminare attraverso l'utilizzo dell'ANOVA within-subject fra le condizioni di CTOA e le condizioni di suggerimento, nel tentativo di individuare il cueing effect e la sua interazione rispetto al CTOA. Una volta ottenuta l'evidenza di un'interazione significativa fra tipologia di suggerimento e lunghezza del CTOA, i tempi di reazione per ogni suggerimento sono stati confrontati separatamente in ogni condizione CTOA (250 ms, 400 ms, 550 ms) per un ulteriore approfondimento nel comprendere l'effetto cueing. L'IOR è stato infine calcolato con il rapporto dei tempi di reazione fra le prove valide e le prove invalide. Su questo è stata applicata nuovamente una ANOVA al fine di indagare l'IOR nei tre CTOA. L'effetto IOR è stato calcolato dividendo i tempi di reazione tra le condizioni di suggerimento valide e non valide. Infine, è stata adottata un'ANOVA a una via per indagare l'effetto IOR nei tre CTOA.

Analisi dei potenziali evento relati (ERP)

Per quanto concerne i dati evento relati target-onset (il punto zero coincide con il punto in cui viene mostrato il target) è stata compiuta una segmentazione dei segnali in epoche da 900 millisecondi, con un punto di inizio a 50 ms prima della comparsa del target fino a 850 millisecondi dopo la comparsa del target. È stata effettuata una *baseline correction* utilizzando i 50 millisecondi precedenti alla presentazione dello stimolo. Sono state, successivamente, respinte le epoche con un'ampiezza superiore a $\pm 75\mu\text{V}$ e applicato un filtro a basse frequenze a 30 Hz. Attraverso queste procedure preliminari sono state scartate circa il 4 % delle prove. Le epoche sono state, in seguito, mediate tra i partecipanti per ogni condizione e per emi-campo visivo. Sulla base di studi precedenti che attraverso l'utilizzo di un paradigma simile hanno riscontrato attivazioni nella regione occipitale, sono state identificate due regioni di interesse (ROI) che includono gli elettrodi parietali e occipitali quali: P7, P5, P3, P1, PO7, PO3, O1 nell'emisfero sinistro e P8, P6, P4, P2, PO8, PO4, O2 nell'emisfero destro. Nel particolare è stata indagata l'interazione di effetti del suggerimento con le attivazioni dei due emisferi, a tal proposito è stata condotta un ANOVA within subject a modello misto tra le condizioni di suggerimento e gli emisferi.

Risultati comportamentali

Una volta calcolati i tempi di reazione medi per ciascuna condizione e CTOA è stata condotta una ANOVA within subject a misure ripetute. L'analisi ha mostrato come l'interazione bidirezionale tra condizione di suggerimento e condizioni CTOA risulti significativa (vedi tabella). Come anche gli effetti principali del suggerimento nelle condizioni di cueing (vedi tabella). Il confronto post-hoc ha rivelato l'esistenza dell'effetto inibitorio in tutte e tre le condizioni CTOA. I soggetti, infatti, presentano tempi di reazione più brevi nelle prove invalide rispetto alle prove valide.

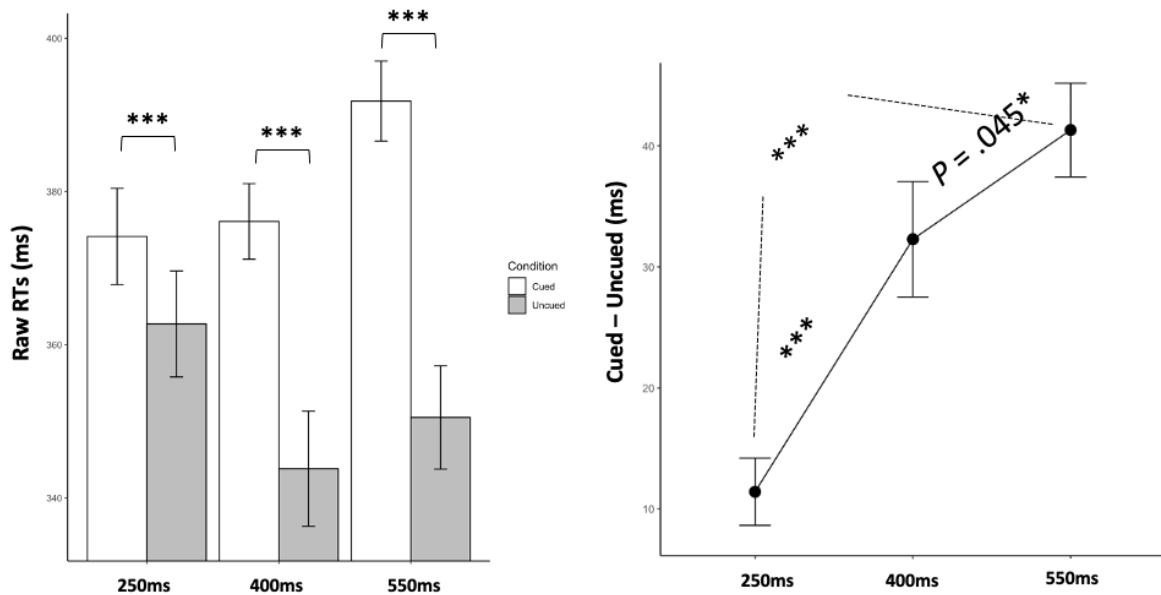


Figura 14 Nel primo grafico vengono presentati i tempi di reazione confrontando le prove valide con le prove invalide, nel secondo invece vengono confrontati i tempi di reazione medi per ogni condizione CTOA (immagine tratta da Fu 2023).

	<u>Mean RTs</u> <u>(Uncued)</u>	<u>Mean RTs</u> <u>(Cued)</u>	<i>t</i>	<u>Cohen's d</u>
250 ms	362.70 (7.03)	374.13 (6.38)	4.05***	0.74
400 ms	343.80 (7.64)	376.09 (5.02)	6.67***	1.22
550 ms	350.50 (6.85)	391.80 (5.31)	10.49***	1.92

Tabella 1. Risultati del t-test per dati appaiati tra i tempi di reazione medi di prove valide e invalide per ogni condizione CTOA di ogni soggetto (tabella tratta da Fu 2023).

È risultato evidente anche come l'effetto inibitorio, ovvero, la presenza di tempi di reazione più lunghi nelle prove Invalide, aumentano in funzione della lunghezza dell'intervallo di tempo tra lo stimolo di suggerimento e lo stimolo target. Per comprendere meglio la relazione fra l'aumento dell'effetto inibitorio e l'aumento dei CTOA sono stati calcolati i tempi di reazione medi per ogni condizione CTOA ed è stata confrontata la differenza fra i tempi di reazione delle prove valide e i tempi di reazione delle prove invalide. Su questi dati è stata condotta una ANOVA unidirezionale che ha

mostrato effetti del CTOA sul fenomeno inibitorio. Inoltre, l'applicazione di un paired t-test ha portato alla consapevolezza di una maggiore dimensione dell'effetto nella condizione CTOA 550 ms rispetto alla condizione CTOA 400 ms che a sua volta ha mostrato una maggiore dimensione dell'effetto rispetto alla condizione CTOA 250 ms. Da questi dati è possibile conseguire che l'aumento dell'intervallo di tempo che intercorre fra la presentazione del suggerimento e quella del target è correlata positivamente alla differenza fra i tempi di reazione di prove congruenti e tempi di reazione di prove incongruenti.

Risultati segnali ERP

Per quanto riguarda l'analisi dei potenziali evento-relati sono stati identificati sette elettrodi per ogni emisfero in modo da formare due diverse regioni di interesse per emisfero. Gli elettrodi all'interno del ROI riguardavano le zone occipitali e parietali.

Nel tentativo di verificare l'IOR delle condizioni di suggerimento o la possibile lateralizzazione dell'effetto, è stata applicata una ANOVA within-subject a disegno misto per ogni campo visivo in ogni condizione di suggerimento e per ogni regione di interesse (quindi per l'emisfero destro e l'emisfero sinistro). Nello specifico sono state selezionate diverse componenti, tra cui: P170 (160-180 ms), N250 (210-230 ms), N300 (270-310 ms), P300 (325-345 ms), N400 (350-370 ms), P400 (410-430 ms) e un'onda lenta (450-500 ms). I risultati hanno evidenziato una maggiore ampiezza per entrambe le regioni di interesse nelle prove invalide rispetto alle prove valide della componente positiva P170. Questa componente risulta essere la prima differenza riscontrabile nell'effetto di inibizione di ritorno. In seguito, è stata rilevata un'interazione significativa tra gli emisferi nella componente negativa N250 che presenta anche essa una maggiore ampiezza nelle prove invalide rispetto alle prove valide, solamente per l'emisfero destro. Anche per la componente N300 l'analisi della varianza ha mostrato un effetto di interazione tra l'ampiezza del picco negativo e gli emisferi, con una maggiore ampiezza per le prove valide rispetto alle prove invalide.

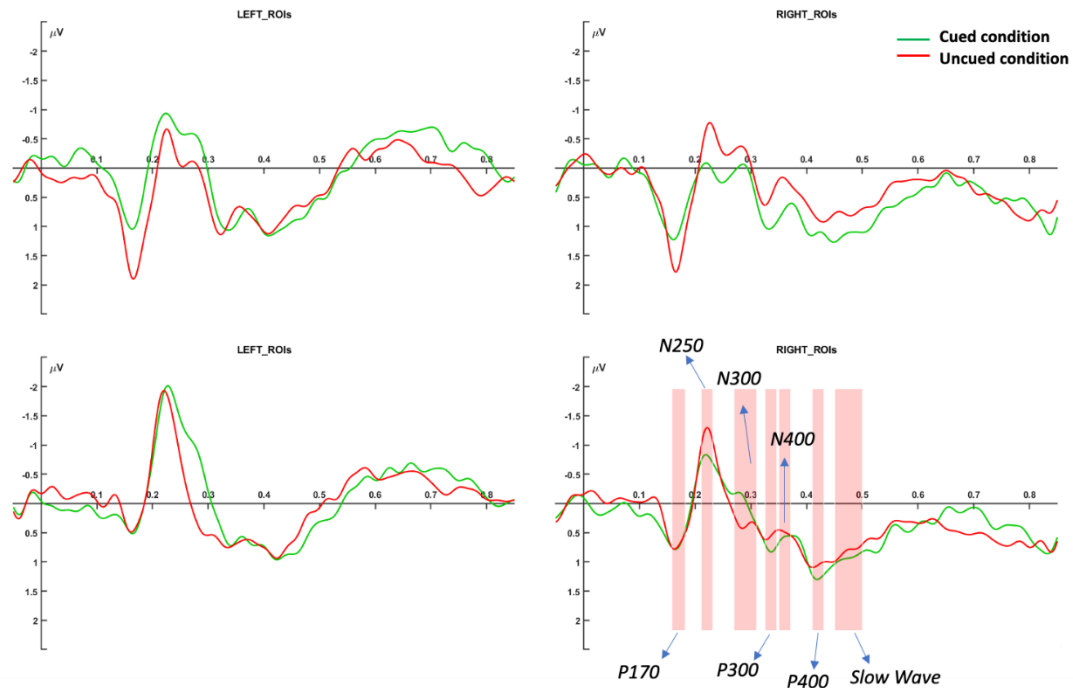


Figura 15. Forme d'onda ERP target-locked per target apparsi nel campo visivo superiore (prima riga) e target apparsi nel campo visivo inferiore (seconda riga) relativi alla ROI nell'emisfero sinistro (prima colonna) e alla ROI nell'emisfero destro (seconda colonna). I rettangoli rossi indicano le finestre di tempo in cui vi sono differenze (immagine tratta da Fu 2023).

	df	P170		N250		N300 Peak Amplitude		N300 Peak Latency	
		F	η^2	F	η^2	F	η^2	F	η^2
Upper visual field									
Cueing	1, 29	19.15***	0.40	0.13	0.00	0.02	0.00	0.03	0.00
Hemisphere	1, 29	0.00	0.00	3.18	0.10	0.06	0.00	1.91	0.06
Cueing * Hemisphere	1, 29	2.57	0.08	17.23***	0.37	8.76**	0.23	1.25	0.04
Lower visual field									
Cueing	1, 29	0.00	0.00	2.78	0.09	10.54**	0.27	0.97	0.03
Hemisphere	1, 29	3.73	0.11	8.50**	0.23	3.37	0.10	2.39	0.08
Cueing * Hemisphere	1, 29	0.08	0.00	6.10*	0.17	15.76***	0.35	4.29*	0.13

Tabella 2. Risultati dell'ANOVA per misure ripetute a due vie di ampiezza media per gli effetti di cueing e gli emisferi su ciascun componente sia per il campo visivo superiore che per quello inferiore (immagine tratta da Fu 2023).

Per quanto riguarda gli effetti delle componenti relativi agli emi-campi visivi, la componente N250 presenta una lateralizzazione destra significativa per entrambi i campi visivi superiore e inferiore. Per quanto riguarda la componente N300 essa presenta una significatività solamente nell'emisfero sinistro e nel campo visivo inferiore, non presentando alcuna differenza significativa nel campo visivo superiore. Anche per la

componente P170 non vi è alcun effetto nelle prove con presentazione dello stimolo target nel campo visivo inferiore.

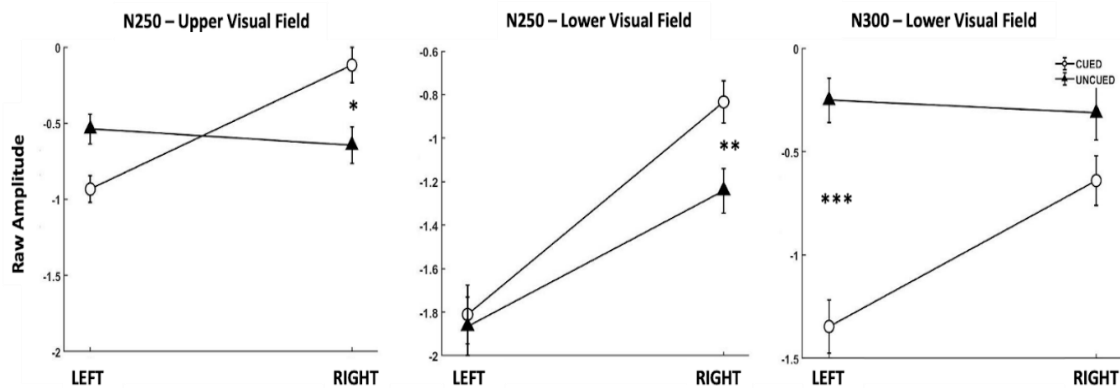


Figura 16. Risultati dell'ampiezza media sugli emisferi sinistro e destro per il componente N250 nel campo visivo superiore e inferiore e N300 nel campo visivo inferiore. La barra di errore rappresenta l'errore standard della media (immagine tratta da Fu 2023).

4.2 Descrizione della pipeline

Per quanto riguarda l'analisi multivariata dei dati ERP oggetto di questo elaborato, sono state implementate diverse pipeline in modo da poter confrontare l'effetto di alcune metodologie sulla performance ed interpretabilità dei dati. Attraverso l'utilizzo del toolbox `mvpalab` nell'ambiente MATLAB è stata implementata una pipeline di un'analisi MVPA time resolved single-subject a trial singolo. Nel corso dell'analisi sono state applicate diverse metodologie per verificare la robustezza dei risultati. Sono stati confrontati i diversi risultati dell'analisi (in termini di distribuzione di accuratezze) applicando la riduzione della dimensionalità dei dati (PCA a 3, 8, 16 e 32 componenti principali), utilizzando uno smoothing spaziale con una finestra temporale di 5 millisecondi e, infine, utilizzando due differenti algoritmi di classificazione (SVM, LDM). Per quanto riguarda la pipeline implementata su python, oltre all'analisi single trial single-subject, è stata implementata un'analisi across-subject con due trial per ogni soggetto (un trial per condizione). L'implementazione delle pipeline nell'ambiente di programmazione Python è avvenuta attraverso la creazione di dataframe `MultiIndex` contenente l'intero campione di dati suddiviso per le tre condizioni CTOA (250 ms, 400 ms, 550 ms) attraverso le funzioni presenti nella libreria `Pandas`. Una volta ottenuto il dataframe è stato implementato un modello di classificazione quale Support vector machine attraverso la funzione `svm.SVC` di `sklearn`. È stata condotta, in seguito, un

ottimizzazione dei parametri attraverso un ciclo for. Per il calcolo delle accuratezze nulle è stato utilizzato un procedimento diverso rispetto all'implementazione su MATLAB.

Per quanto riguarda le analisi across-subjects è stata ottenuta una serie di accuratezze nulle per soggetto attraverso l'addestramento del classificatore con le lable randomizzate e applicando una Repeated Stradified K Fold cross validation con venti ripetizioni. Per l'analisi single-subject, single trial, invece, è stata calcolata un accuratezza nulla per ogni soggetto attraverso le permutazioni ottenendo un vettore di 30 accuratezze nulle, una per soggetto. La significatività oltre il livello del caso è stata calcolata attraverso la somministrazione del test statistico Wilcoxon Signed-Rank Test. Nell'analisi accross-subject per ogni istante di tempo è stata confrontata la distribuzione delle accuratezze ripetute e delle accuratezze nulle ripetute (repeated stratified k fold), nell'analisi single subject, invece, è stata confrontata la distribuzione delle accuratezze ottenute da ogni soggetto per istante di tempo con le accuratezze nulle anche qui ottenute per singolo soggetto e istante di tempo. In seguito, è stata applicata una correzione per comparazioni multiple (FDR) e in seguito ottenuti gli istanti di significatività. Le analisi aggiuntive sono state eseguite tramite il toolbox mvpalab, di conseguenza solo sulla pipeline single-subject, single trial.

4.3 Risultati e discussione

Analisi delle accuratezze per istante di tempo

Per l'analisi across subject con due pseudotrial per soggetto (una per condizione Valida e una per condizione Invalida) le accuratezze ottenute nella fase di training sono risultate molto alte rispetto alle accuratezze ottenute nella fase di test suggerendo che i modelli fossero andati in overfitting. L'ottimizzazione degli iper-parametri non ha prodotto miglioramenti, nello specifico, l'unico parametro ottimizzabile per una SVM con kernel lineare è il parametro "C" che indica il costo per gli errori di classificazione.

Le analisi single subject e single trial hanno ottenuto risultati pressoché identici nelle due implementazioni (mvpalab e python).

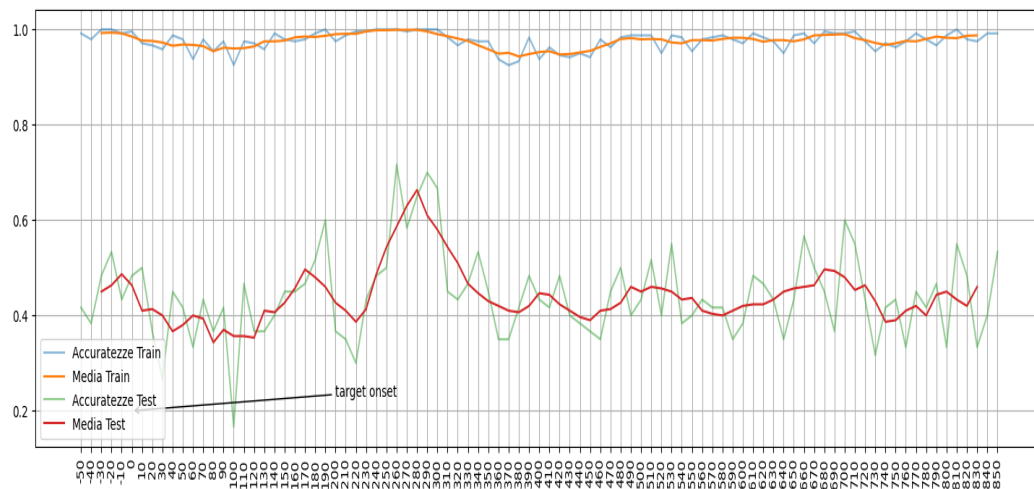


Figura 17. Rappresentazione grafica della distribuzione di accuratze per istante di tempo per la fase di training (linea blu) e per la fase di test (linea verde). Per ogni distribuzione è stata calcolata la media mobile (linea arancione e linea rossa).

Condizione 250

Per quanto riguarda la condizione 250 è stato ottenuto un periodo di significatività oltre il caso nella finestra di tempo da 170 ms e 380 ms (ad esclusione dell'istante 220 ms). Il picco, ottenuto a 260 ms, presenta un'accuratezza di 0,58 .Come in articoli precedenti che applicavano la time resolved MVPA su dati EEG sono stati eliminati periodi di significatività con meno di tre istanti consecutivi. È stata calcolata una media mobile in modo da rendere la rappresentazione grafica della distribuzione più chiara. Il calcolo della significatività è stato ottenuto attraverso l'applicazione del procedimento implementato nel toolbox mvpalab (vedi capitolo 3 paragrafo 1), individuando i cluster di istanti temporali che superano il livello del caso, ovvero oltre un'accuratezza di 0,518. L'intervallo di tempo significativo, nell'analisi temporale di decodifica multivariata, risulta congruente con i risultati delle analisi univariate condotte da Fu. Il primo picco, infatti, che presenta differenze nelle due condizioni è la P170 che coincide con il punto di inizio della finestra di significatività dell'analisi MVPA. Un'altra interessante considerazione riguarda il picco di accuratezza nella finestra significativa individuata a 260 ms. Una possibile interpretazioni di questo picco può essere data considerando la componente N250 che mostra una consistente differenza di ampiezza nelle due

condizioni. Il periodo di significatività delle accuratèzze si estende fino a 390 millisecondi, comprendendo anche differenze significative nelle componenti P300 e N400.

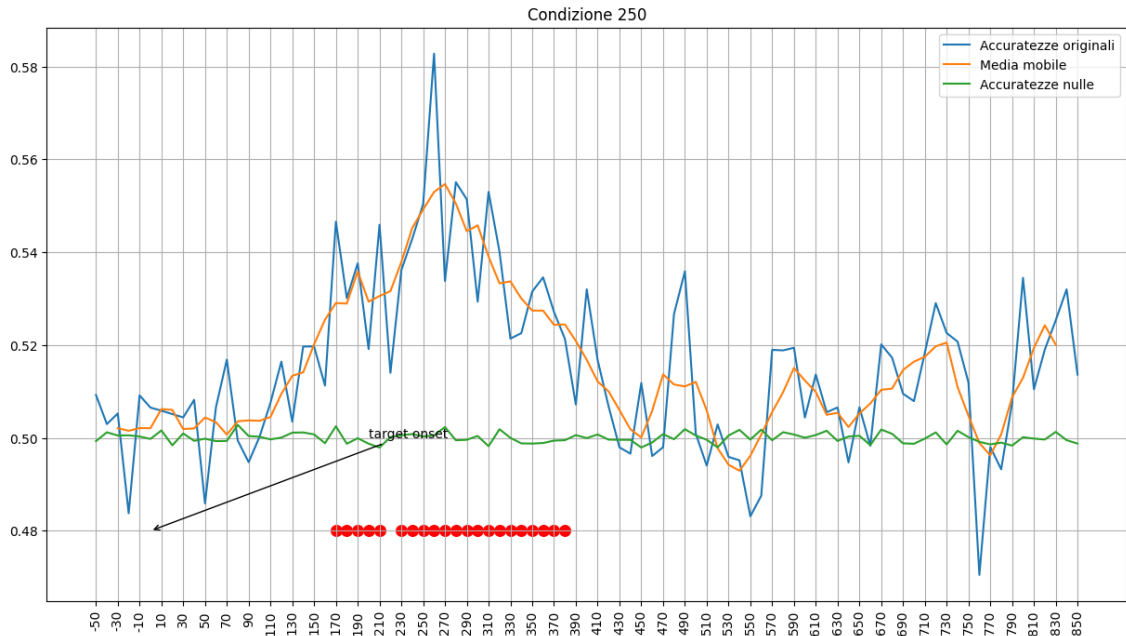


Figura 18. Rappresentazione grafica per la condizione con un CTOA DI 250 millisecondi: della distribuzione di accuratèzze per istanti di tempo (linea blu), della media mobile (linea arancione) e delle accuratèzze nulle (linea verde). In basso sono stati evidenziati i singoli punti risultati significativi.

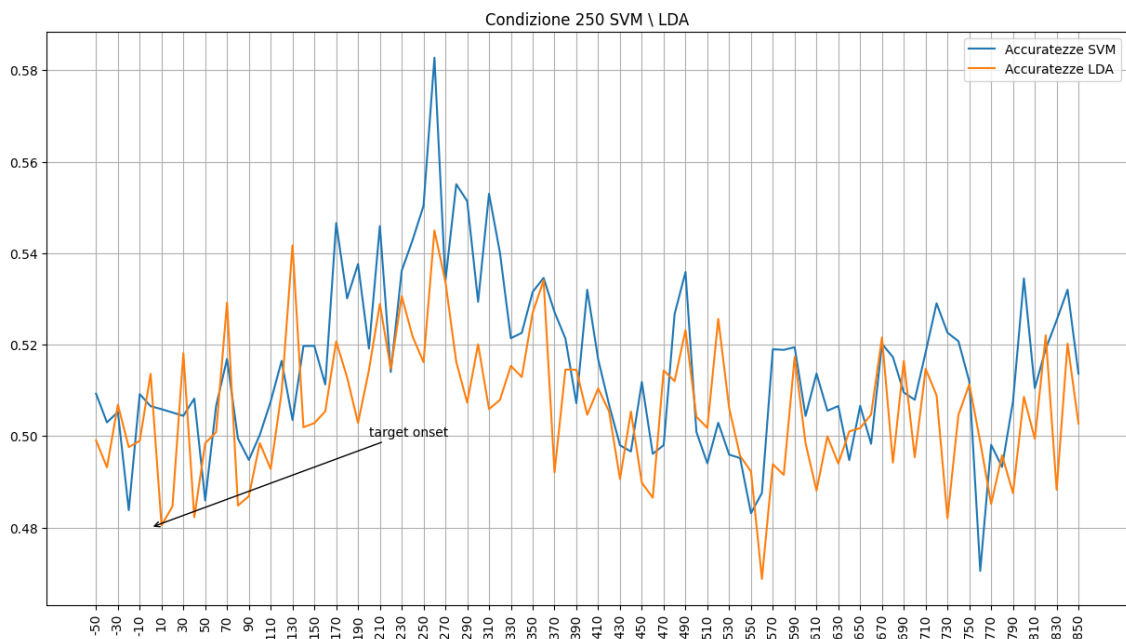


Figura 19. Rappresentazione grafica, per la condizione con un CTOA di 250 millisecondi della distribuzione di accuratèzze calcolata attraverso l'uso del classificatore SVM (blu) e del classificatore LDA (arancione).

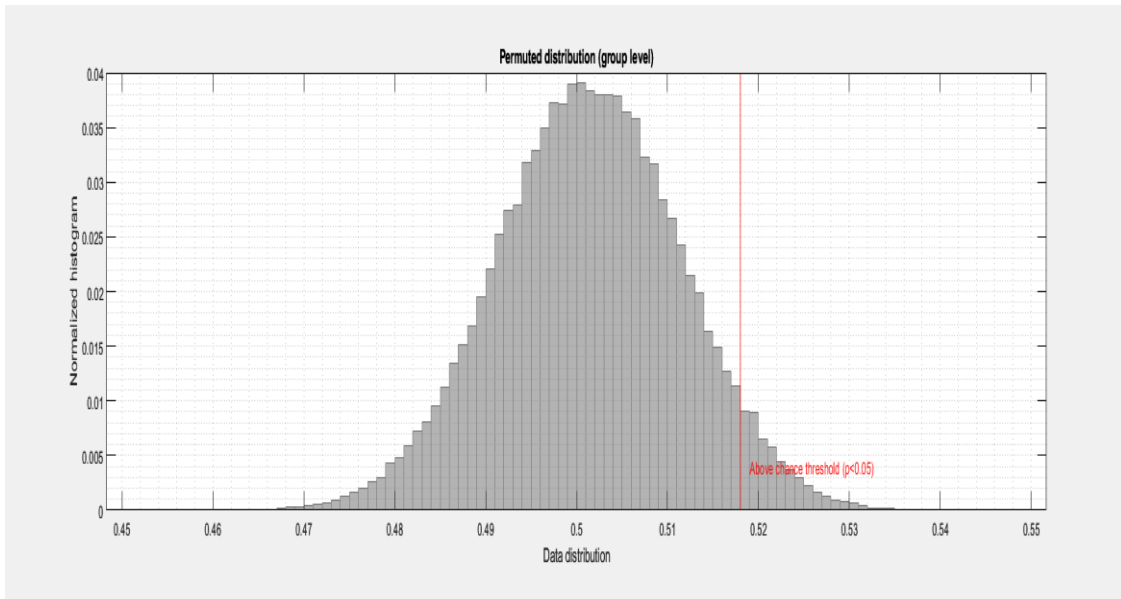


Figura 20. Mappa delle distribuzioni di accuratze permutate e calcolo del threshold ($\alpha = 0,05$)

Come descritto in precedenza, sono state indagate differenti alternative nel tentativo di testare la robustezza dei risultati. Nel confronto con l'algoritmo di classificazione LDA la Support Vector Machine presenta una distribuzione in media più performante. Un ulteriore confronto è stato fra i vettori di accuratezza ottenuti con lo stessa pipeline ma con riduzione della dimensionalità del campione (PCA; 3, 8,16, 32).

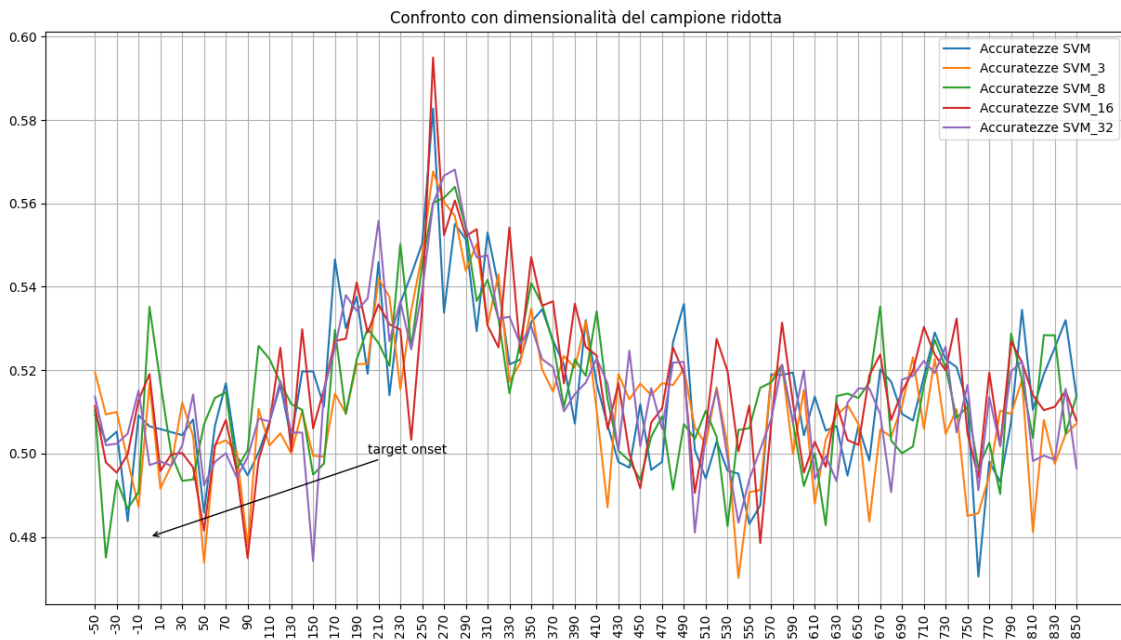


Figura 21. Distribuzione delle accuratze per istanti di tempo per il confronto di implementazioni con diverse riduzioni della dimensionalità.

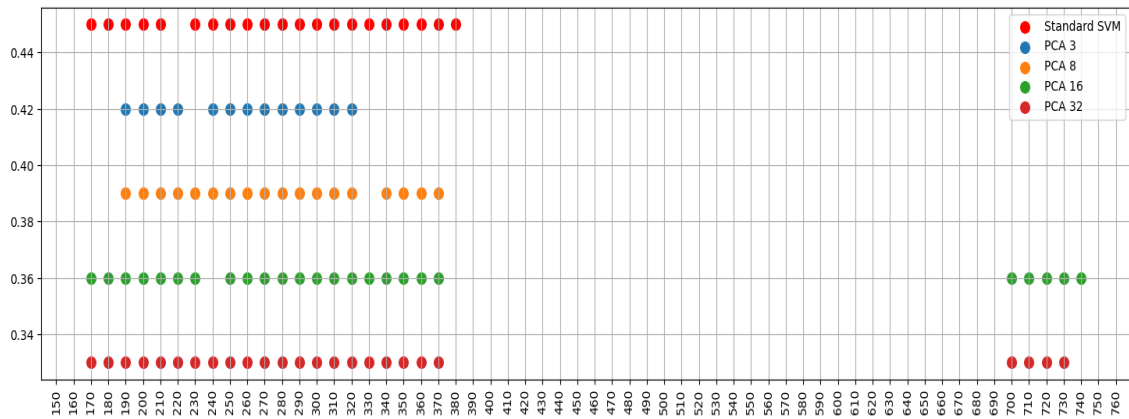


Figura 22. Rappresentazione degli istanti significativi per ogni distribuzione

Le distribuzioni delle accuratèzze ottenute tramite l'applicazione della PCA con diverse componenti principali hanno seguito un andamento simile alla distribuzione di accuratèzze ottenute senza alcuna riduzione della dimensione. Per la riduzione a tre componenti principali il picco si presenta nell'istante 260 ms con un'accuratèzza pari a 0,56 circa. A otto componenti il picco si sposta a 280 ms con un'accuratèzza di 0,56 circa. L'analisi con una PCA a 16 componenti presenta il picco piú alto a 260 ms con una accuratèzza di 0,59 circa. Per l'analisi a 32 componenti il picco emerge a 280 ms con una accuratèzza di 0,56 circa. Anche il periodo di significatività presenta similitudini presentando in molte analisi un periodo nella finestra di tempo 170 ms – 380 ms.

Condizione 400

Per la condizione CTOA 400 ms il picco nella distribuzione delle accuratèzze si presenta all'istante 260 ms con un'accuratèzza di 0,56. Il periodo di significatività ricopre la finestra di tempo 170 ms – 330 ms (escluso l'istante di 240 ms). Anche qui, il punto di inizio della finestra di tempo significativa a 170 ms potrebbe essere dovuto alla differenza significativa della componente P170, rilevata nelle analisi univariate. Il picco di accuratèzza a 240 millisecondi potrebbe anch'esso riflettere differenze nella componente N250.

L'intervallo di istanti significativi, nella condizione 400, termina prima, di circa 50 millisecondi, suggerendo un minore impatto informativo delle componenti successive come l'N400.

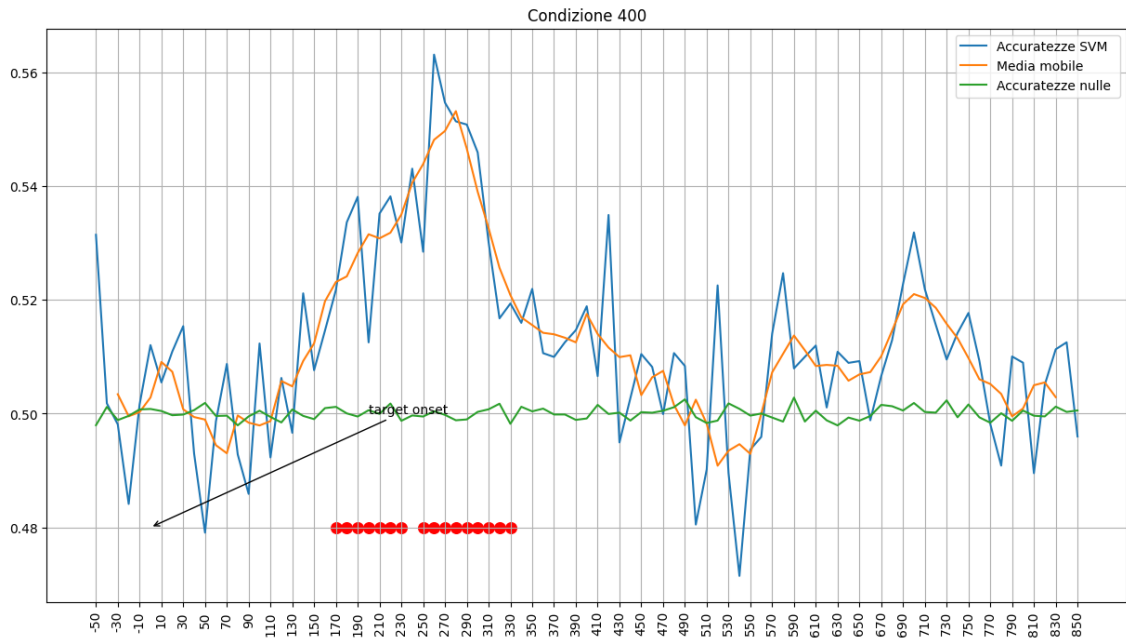


Figura 23. Rappresentazione grafica per la condizione con un CTOA DI 400 millisecondi: della distribuzione di accuratezze per istanti di tempo (linea blu), della media mobile (linea arancione) e delle accuratezze nulle (linea verde). In basso sono stati evidenziati i singoli punti risultati significativi.

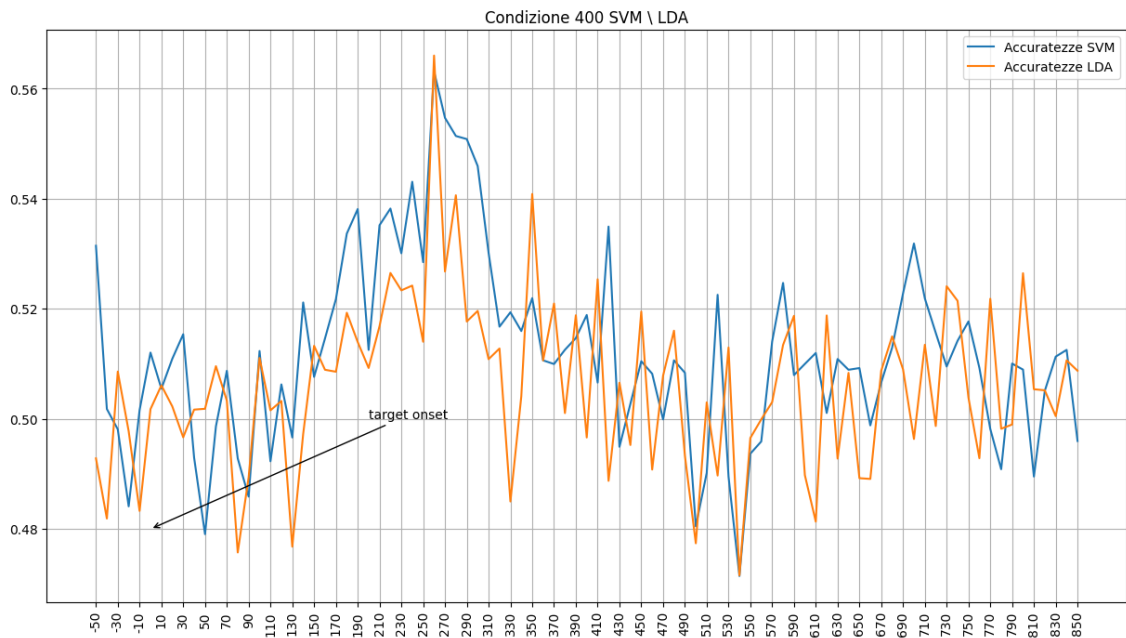


Figura 24. Rappresentazione grafica, per la condizione con un CTOA di 250 millisecondi della distribuzione di accuratezze calcolata attraverso l'utilizzo del classificatore SVM (blu) e del classificatore LDA (arancione).

Il confronto tra i due classificatori ha riportato, anche in questa condizione, la superiorità di performance della Support Vector Machine rispetto alla LDA. Inoltre, i risultati con analisi con dimensionalità ridotte ha dimostrato la robustezza dell'analisi in quanto le performance del modello in tutte queste condizioni risulta molto simile.

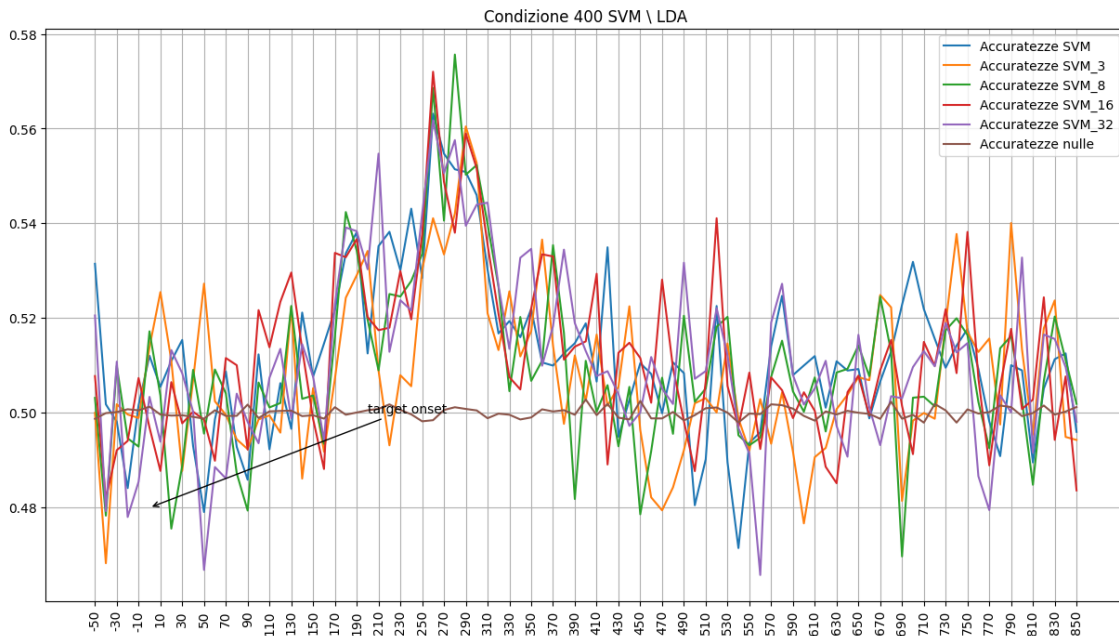


Figura 25. Distribuzione delle accurattee per istanti di tempo per il confronto di implementazioni con diverse riduzioni della dimensionalità.

Condizione 550

Nella condizione CTOA 550 ms il picco massimo è a 260 ms con un accuratezza di 0,57 circa. La finestra di tempo in cui le accurattee risultano significative è 180 ms – 320 ms. Anche qui la finestra di tempo risultata significativa si presenta molto simile con un inizio a 180 millisecondi. Nella condizione 550, come nella condizione precedente, l'intervallo di istanti significativi risulta essere più corto, interrompendosi a 320 ms. Questo suggerisce una interessante considerazione sulle componenti tardive nella discriminazione dell'effetto di inibizione a seconda dell'intervallo presente fra lo stimolo suggerimento e lo stimolo target. Inoltre la presenza di intervalli simili permette di inferire

un'attività neuronale distribuita molto simile tra le condizioni, soprattutto per le componenti precoci come la P170 e la N250.

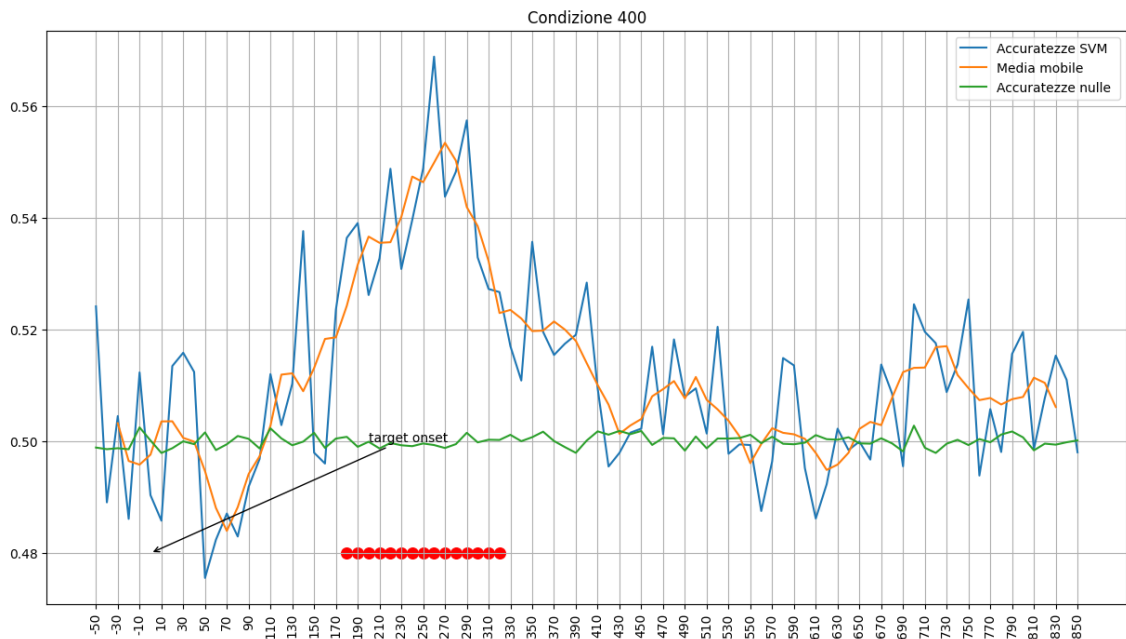


Figura 27. Rappresentazione grafica per la condizione con un CTOA DI 400 millisecondi: della distribuzione di accuratze per istanti di tempo (linea blu), della media mobile (linea arancione) e delle accuratze nulle (linea verde). In basso sono stati evidenziati i singoli punti risultati significativi.

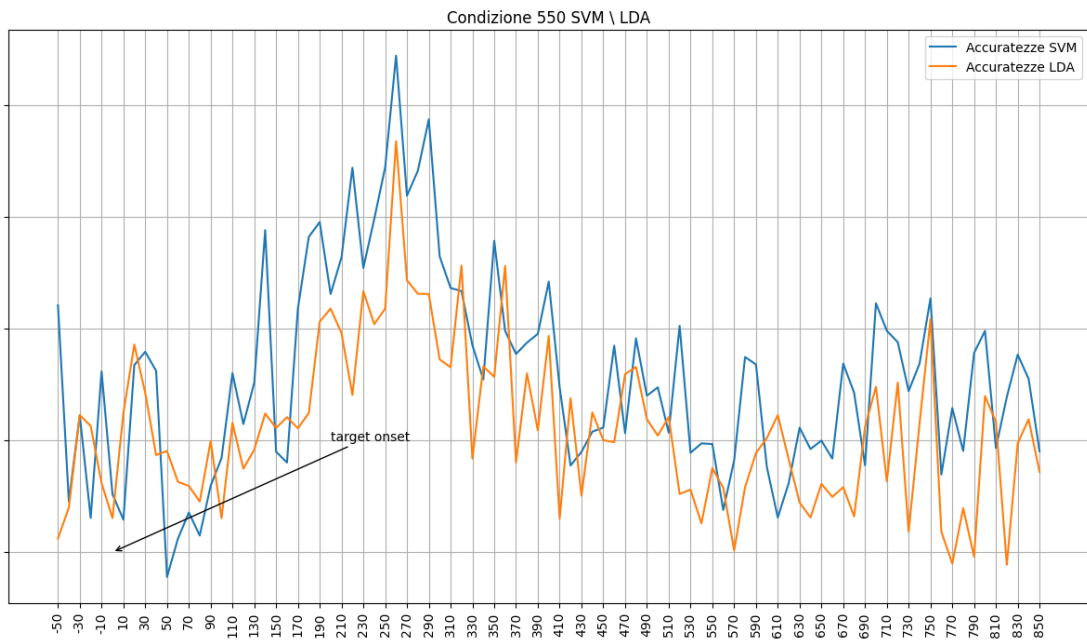


Figura 26. Rappresentazione grafica, per la condizione con un CTOA di 250 millisecondi della distribuzione di accuratze calcolata attraverso l'utilizzo del classificatore SVM (blu) e del classificatore LDA (arancione).

Anche qui, come per le condizioni precedenti c'è stata una migliore performance per il modello SVM rispetto alla LDA. Allo stesso modo, le prove di analisi con l'applicazione della PCA per la riduzione del campione hanno mostrato risultati molto simili fra loro.

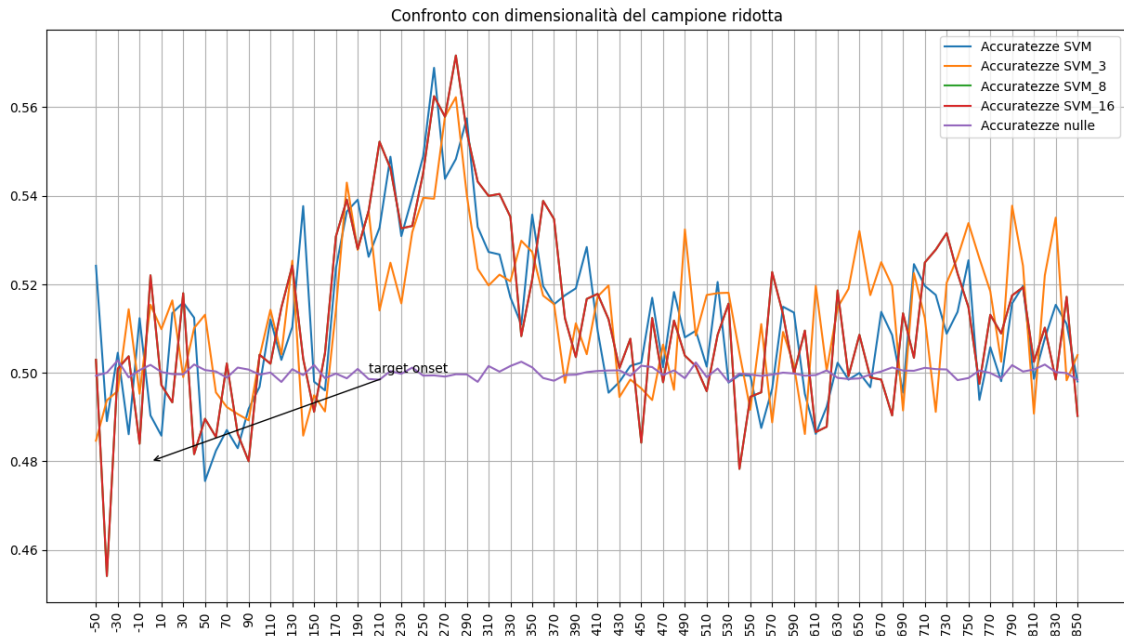


Figura 28. Distribuzione delle accuratze per istanti di tempo per il confronto di implementazioni con diverse riduzioni della dimensionalità.

Temporal Generalization

Come già descritto nei capitoli precedenti, attraverso la matrice di generalizzazione temporale, è possibile comprendere non solo la presenza di informazioni rilevanti, ma anche l'evoluzione di queste informazioni nel tempo.

Nella condizione CTOA 250 ms all'interno della matrice è possibile notare accuratze più alte nell'intervallo 240 ms - 320 ms, questo indica una generalizzazione dei pattern in quest'area. Questo intervallo, racchiude anche il picco di accuratezza a 260 millisecondi. È possibile affermare infatti che negli istanti successivi i patterns presenti permangono anche negli istanti successivi. Nella condizione CTOA 400 ms la generalizzazione delle accuratze è maggiore nell'intervallo 260 ms – 340 ms. Mentre per la condizione CTOA 550 ms vi è una più marcata generalizzazione nell'intervallo 240 ms – 340 ms. Da questo si evince come alcuni dei pattern discriminanti per le condizioni

non siano circoscritti ad un solo istante di tempo ma abbiano una forma distribuita anche nell'intervallo temporale.

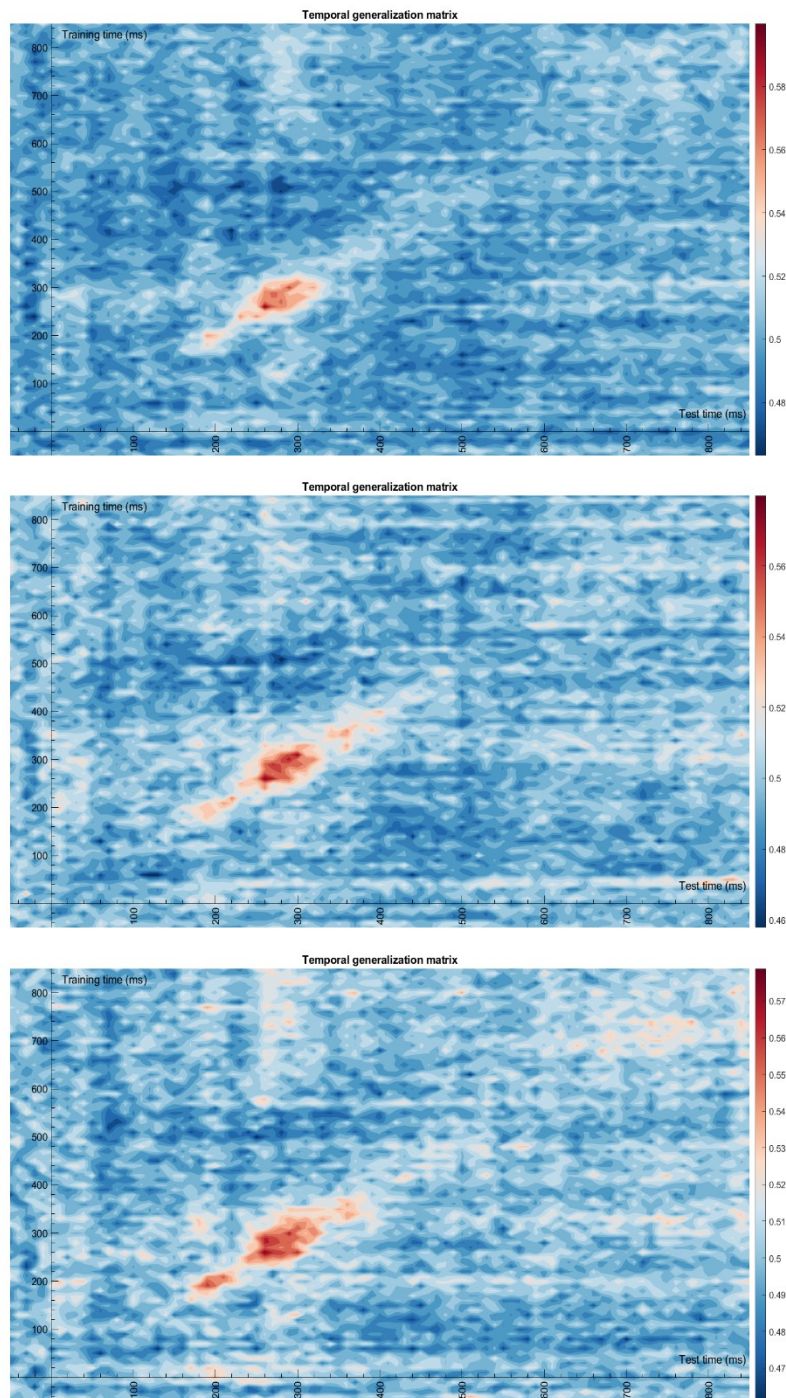


Figura 29. Matrici della generalizzazione temporale. Nella prima figura è rappresentata la condizione con CTOA di 250 ms, nella figura centrale vi è la condizione con un CTOA di 400 ms e, infine, la condizione con CTOA di 550 ms

210 ms l'attivazione risulta meno localizzata e più diffusa nelle aree fronto-temporali e occipitale destra, con attivazioni negative (n.b. negativo e positivo non indicano la polarizzazione dell'attivazione elettrica) nella zona centro-parietale sinistra. L'intervallo 270 ms – 320 ms presenta un pattern di attivazione molto simile all'intervallo 170-210. Infine, nell'ultima finestra di tempo è possibile notare un'attivazione maggiore delle aree occipitali destre. In conclusione anche la mappa della distribuzione dei pesi sottolinea l'importanza della finestra temporale 220-260 mostrando attivazioni più localizzate e fortemente lateralizzate.

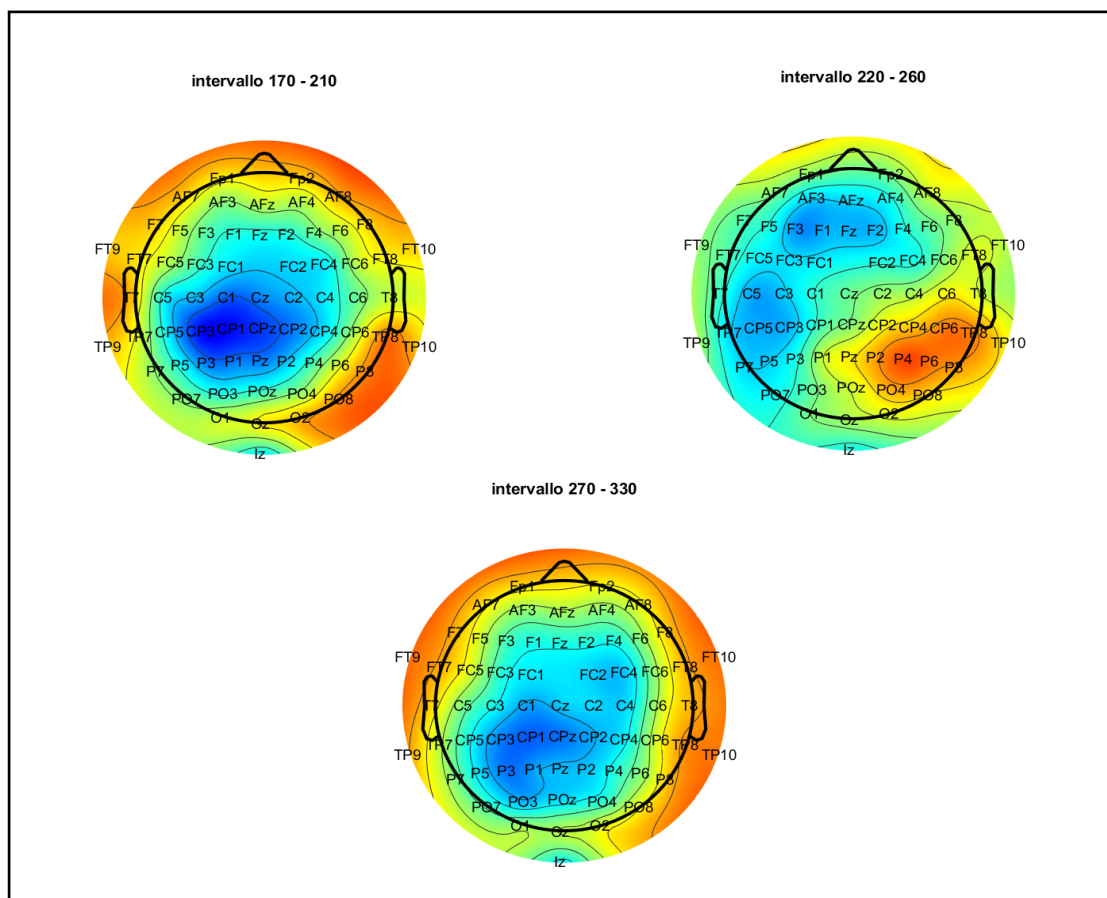


Figura 31. Rappresentazione topografica dei pesi del classificatore corretta con la metodologia di Haufe (Haufe, 2014) mediata per quattro intervalli temporali nella condizione con CTOA pari a 400 millisecondi

La condizione 400 millisecondi presenta pattern di attivazioni molto simili. Nel primo intervallo è possibile notare un'attivazione negativa centro parietale. Il periodo seguente, 220 ms – 260 ms presenta attivazioni molto localizzate, con il coinvolgimento specifico

degli elettrodi P4, P2 E P6. Nell'ultimo periodo 270 ms 330 ms l'attivazione ritorna ad essere meno localizzata con attivazioni fronto-temporali di entrambi gli emisferi e attivazioni negative nella zona centro parietale sinistra.

Anche la condizione 550 dimostra un pattern di attivazione simile alle due condizioni precedenti. L'intervallo 180 ms – 210 ms dimostra un'attivazione fronto-temporale sinistra meno marcata. L'intervallo 220 ms – 260 ms presenta anche esso attivazioni localizzate alla regione parietale destra con un coinvolgimento specifico degli elettrodi P2, P6 E P2. Una differenza individuabile risiede nel maggiore coinvolgimento di elettrodi presenti nella zona temporo-parietale destra con gli elettrodi TP8 e TP10.

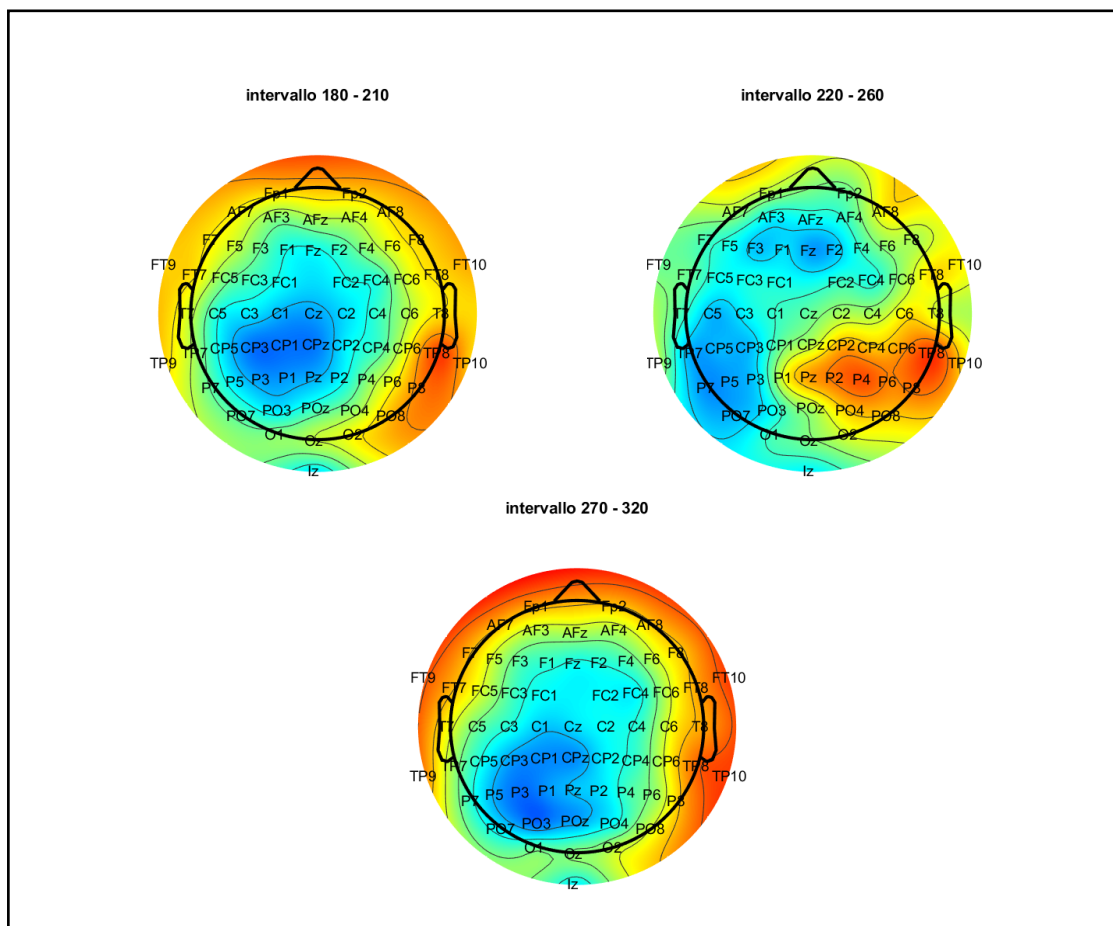


Figura 32. Rappresentazione topografica dei pesi del classificatore corretta con la metodologia di Haufe (Haufe, 2014) mediata per quattro intervalli temporali nella condizione con CTOA pari a 550 millisecondi

L'intervallo seguente 270 ms – 320 ms presenta un'attivazione meno localizzata con attivazioni positive nelle regioni fronto-temporali e una attivazione negativa maggiormente localizzata nella zona parieto-occipitale sinistra.

In conclusione è possibile riscontrare un'informazione più completa nell'analisi di decodifica multivariata, in quanto è possibile risalire ad informazioni riguardo ad elettrodi non presenti nel ROI dell'analisi univariata. Infine, attivazioni simili in tutte e tre le condizioni suggeriscono un'indipendenza dei correlati neurofisiologici dal CTOA. Ovvero, per quanto l'intervallo di tempo fra la presentazione dello stimolo suggerimento e la presentazione dello stimolo target comporti differenze significative nelle manifestazioni comportamentali non sono collegati ad una conseguente differenza nei patterns di attivazione.

Capitolo V: Conclusioni

In questo capitolo verranno brevemente esposte alcune considerazioni sui limiti che presenta la multivariate pattern analysis e anche dei dubbi etici che i progressi di questa tecnica hanno portato.

3.6 Limiti dell'MVPA

Nonostante i vantaggi della multivariate pattern analysis sono stati ampiamente discussi nei capitoli precedenti, tuttavia, risulta naturale pensare che una tecnica innovativa per quanto apporti numerosi benefici nella sua applicazione abbia comunque anche diversi svantaggi. Nell'articolo "*Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience*" di Ritchie, Kaplan e Klein del 2019, infatti, vengono illustrate alcune problematiche che accompagnano l'applicazione del multivariate decoding. In particolare, viene discusso il principio base su cui poggia, secondo gli autori, il "neural decoding". Se un classificatore riesce a discriminare le diverse condizioni sulla base dei dati è implicito pensare che esso stia valutando le stesse informazioni relative ai correlati neurali delle condizioni di partenza (Ritchie, Kaplan e Klein, 2019). Questo principio, infatti, poggia sul fatto che le informazioni utilizzate da un algoritmo di classificazione, nell'analisi di decodifica, rispecchino le rappresentazioni neurali di una determinata condizione. Il problema principale di questo assunto riguarda il fatto che difficilmente è possibile conoscere quali informazioni l'algoritmo di classificazione utilizza per discriminare le condizioni. È importante considerare, infatti, che le rappresentazioni neurali presenti nel cervello potrebbero non coincidere con le informazioni estratte dal classificatore considerando anche che queste vengono raccolte con tecniche di neuroimaging imperfette. Questo pensiero, secondo gli autori, poggia su presupposti errati, tra cui, l'estrema fiducia nelle tecniche di neuroimaging che possono, in realtà, generare artefatti che guidano la classificazione. Gli strumenti di "brain scanning" infatti non presentano una qualità spazio-temporale adeguata a fornire questo grado di sicurezza. Un'altra problematica di grande rilevanza riguarda i diversi algoritmi di classificazione. Come descritto in precedenza, diversi autori, nel corso dello sviluppo delle tecniche di MVPA si sono accorti di quanto siano poco indicati algoritmi di classificazione non lineari per condurre un'analisi di decodifica. Essi, infatti, risultano

eccessivamente complessi e potrebbero estrarre informazioni che persino il cervello stesso non sarebbe in grado di utilizzare (Kamitani, Tong 2005; Naselaris et al., 2011). I classificatori lineari risulterebbero invece più indicati in quanto “biologicamente plausibili” ovvero in grado di decodificare informazioni interpretabili dal cervello. Ciò che non convince, secondo gli autori, è la considerazione semplicistica del classificatore lineare. Infatti, il considerare i classificatori lineari come biologicamente plausibile non indica comunque che essi non possano utilizzare artefatti per la discriminazione. La visione errata, secondo gli autori, parte dall’idea di plausibilità biologica degli algoritmi di classificazione lineare che portano a concepire il modello come un surrogato del cervello. Ovvero, il classificatore sostituisce il cervello nel compito di decodifica, mostrando le informazioni implicite dell’attività neuronale. Gli algoritmi di classificazione lineare, infatti, possono essere sorprendentemente efficaci nel rilevare informazioni utili alla discriminazione ma non all’interpretazione. È importante sottolineare che l’obiettivo degli autori, in questo articolo, non è condannare in toto l’analisi di decodifica multivariata, bensì arginare le concezioni fortemente ottimistiche che la considerano come “mind-reading”. Infatti, l’applicazione della MVPA presenta una validità attendibile nel momento in cui viene utilizzata per la predizione di comportamenti sulla base di modelli o teorie psicologiche preesistenti.

3.7 Questioni etiche

Nel percorso di affermazione del multivariate decoding nell’ambiente neuroscientifico, sono nate diverse preoccupazioni etiche nello sviluppo di questi strumenti innovativi (Tong, Pratte 2012). Negli anni più recenti, infatti, la neuroetica ha avuto notevoli sviluppi, stimolata dall’evoluzione delle informazioni disponibili da tecniche di brain imaging. La principale preoccupazione di questa branca è dovuta agli sviluppi di tecniche di decodifica che comportano una performance sempre più alta. Diversi studi di multivariate decoding hanno dimostrato un’enorme efficacia nel distinguere se un individuo fosse affetto da schizofrenia (Nenadic et al. 2009), o depressione (Craddock et al. 2009), con livelli di accuratezza molto alti (fra l’85% e il 90%). Altre applicazioni hanno dimostrato una sorprendente capacità del decoding multivariato nel discriminare, ad esempio, la presenza di menzogne (Bizzi et al. 2009).

Un'importante conseguenza riguarda la considerazione della privacy mentale, in quanto queste tecnologie permettono una rilevazione di dinamiche personali con performance molto elevate. Portando a nuovissime questioni neuroetiche dal momento che non è esistito alcun precedente, fino ad ora, che permette di scrutare la mente di un altro essere umano.

Conclusioni

La multivariate pattern analysis si conferma quindi come una metodologia di indagine indubbiamente avanzata, apportando numerosissimi vantaggi rispetto ad un tipo di analisi tradizionale. L'utilizzo di algoritmi lineari applicati a disegni sperimentali permette di estrarre molte informazioni utili all'interpretazione dei correlati neurofisiologici dei processi cognitivi. Nello specifico, l'MVPA, è risultata molto utile per l'analisi dei dati riguardanti il fenomeno dell'IOR ampliando le conoscenze ottenute dall'analisi univariata precedentemente compiuta. Un fenomeno così complesso come l'inibizione di ritorno che presenta diversi risultati contrastanti anche per lo studio dei correlati neurofisiologici ad esso associati, si presta più che bene ad un'analisi di tipo multivariato che permette di condurre una ricerca che coinvolge l'intero sistema senza una selezione di determinate regioni di interesse dovute a teorizzazioni precedenti. Inoltre, il processo di semplificazione promosso dall'implementazione di toolbox sempre più completi porta l'analisi di decodifica ad essere uno strumento progressivamente più indispensabile per il ricercatore intento ad indagare fenomeni neurocognitivi.

References

- Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics, 8*. <https://doi.org/10.3389/fninf.2014.00014>
- Ashton, K., Zinszer, B., Cichy, R. M., Nelson, C. A., Aslin, R. N., & Bayet, L. (2022). Time-resolved multivariate pattern analysis of infant EEG data: A practical tutorial. *Developmental Cognitive Neuroscience, 54*, 101094. <https://doi.org/10.1016/j.dcn.2022.101094>
- Bandettini, P. A. (2009). What's New in Neuroimaging Methods? *Annals of the New York Academy of Sciences, 1156*(1), 260–293. <https://doi.org/10.1111/j.1749-6632.2009.04420.x>
- Bashinski, H. S., & Bacharach, V. R. (1980). Enhancement of perceptual sensitivity as the result of selectively attending to spatial locations. *Attention Perception & Psychophysics, 28*(3), 241–248. <https://doi.org/10.3758/bf03204380>
- Berger, A., Henik, A., & Rafal, R. D. (2005). Competition Between Endogenous and Exogenous Orienting of Visual Attention. *Journal of Experimental Psychology: General, 134*(2), 207–221. <https://doi.org/10.1037/0096-3445.134.2.207>
- Berlucchi, G. (2006). Inhibition of return: A phenomenon in search of a mechanism and a better name. *Cognitive Neuropsychology, 23*(7), 1065–1074. <https://doi.org/10.1080/02643290600588426>
- Berlucchi, G., Tassinari, G., Marzi, C. A., & Di Stefano, M. (1989). Spatial distribution of the inhibitory effect of peripheral non-informative cues on simple reaction time to non-fixated visual targets. *Neuropsychologia, 27*(2), 201–221. [https://doi.org/10.1016/0028-3932\(89\)90172-3](https://doi.org/10.1016/0028-3932(89)90172-3)
- Bhavsar, H., & Panchal, M. H. (2012). A review on support vector machine for data classification. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1*(10), 185-189
- Blankertz, B., Lemm, S., Treder, M. S., Haufe, S., & Müller, K. (2011). Single-trial analysis and classification of ERP components — A tutorial. *NeuroImage, 56*(2), 814–825. <https://doi.org/10.1016/j.neuroimage.2010.06.048>
- Bizzi, E., Hyman, S. E., Raichle, M. E., Kanwisher, N., Phelps, E., Morse, S. J., Sinnot-Armstrong, W., Rakoff, J. S., & Greely, H. T. (2009). Using imaging to detect deceit. *American Academy of Arts and Sciences*.
- Blei, D. M., & Smyth, P. (2017). Science and data science. *Proceedings of the National Academy of Sciences of the United States of America, 114*(33), 8689–8692. <https://doi.org/10.1073/pnas.1702076114>
- Bode, S., Feuerriegel, D., Bennett, D. L., & Alday, P. M. (2018). The Decision Decoding ToolBOX (DDTBOX) – A Multivariate Pattern Analysis Toolbox for Event-Related Potentials. *Neuroinformatics, 17*(1), 27–42. <https://doi.org/10.1007/s12021-018-9375-z>

- Bode, S., Sewell, D. K., Lilburn, S. D., Forte, J. D., Smith, P. L., & Stahl, J. (2012). Predicting Perceptual Decision Biases from Early Brain Activity. *The Journal of Neuroscience*, *32*(36), 12488–12498. <https://doi.org/10.1523/jneurosci.1708-12.2012>
- Bourgeois, A., Chica, A. B., Migliaccio, R., De Schotten, M. T., & Bartolomeo, P. (2012). Cortical control of inhibition of return: Evidence from patients with inferior parietal damage and visual neglect. *Neuropsychologia*, *50*(5), 800–809. <https://doi.org/10.1016/j.neuropsychologia.2012.01.014>
- Breiman, L. (2001). Statistical Modeling: The Two Cultures (with comments and a rejoinder by the author). *Statistical Science*, *16*(3). <https://doi.org/10.1214/ss/1009213726>
- Bzdok, D. (2017). Classical Statistics and Statistical Learning in Imaging Neuroscience. *Frontiers in Neuroscience*, *11*. <https://doi.org/10.3389/fnins.2017.00543>
- Bzdok, D., & Yeo, B. T. (2017). Inference in the age of big data: Future perspectives on neuroscience. *NeuroImage*, *155*, 549–564. <https://doi.org/10.1016/j.neuroimage.2017.04.061>
- Carlson, T. A., Hogendoorn, H., Kanai, R., Mesik, J., & Turret, J. (2011). High temporal resolution decoding of object position and category. *Journal of Vision*, *11*(10), 9. <https://doi.org/10.1167/11.10.9>
- Carlson, T. A., Schrater, P., & He, S. (2003). Patterns of Activity in the Categorical Representations of Objects. *Journal of Cognitive Neuroscience*, *15*(5), 704–717. <https://doi.org/10.1162/jocn.2003.15.5.704>
- Carlson, T. A., Tovar, D. A., Alink, A., & Kriegeskorte, N. (2013). Representational dynamics of object vision: The first 1000 ms. *Journal of Vision*, *13*(10), 1. <https://doi.org/10.1167/13.10.1>
- Carlson, T. A., & Wardle, S. G. (2015). Sensible decoding. *NeuroImage*, *110*, 217–218. <https://doi.org/10.1016/j.neuroimage.2015.02.009>
- Cauchoux, M., Barragan-Jason, G., Serre, T., & Barbeau, E. J. (2014). The Neural Dynamics of Face Detection in the Wild Revealed by MVPA. *The Journal of Neuroscience*, *34*(3), 846–854. <https://doi.org/10.1523/jneurosci.3030-13.2014>
- Chambers, J. C. (1993). Greater or lesser statistics: a choice for future research. *Statistics and Computing*, *3*(4), 182–184. <https://doi.org/10.1007/bf00141776>
- Chan, A. C. L., Halgren, E., Marinkovic, K., & Halgren, E. (2011). Decoding word and category-specific spatiotemporal representations from MEG and EEG. *NeuroImage*, *54*(4), 3028–3039. <https://doi.org/10.1016/j.neuroimage.2010.10.073>
- Cheal, M., Lyon, D. W., & Hubbard, D. A. (1991). Does Attention Have Different Effects on Line Orientation and Line Arrangement Discrimination? *The Quarterly Journal of Experimental Psychology*, *43*(4), 825–857. <https://doi.org/10.1080/14640749108400959>
- Chen, G., Adleman, N. E., Saad, Z. S., Leibenluft, E., & Cox, R. A. (2014). Applications of multivariate modeling to neuroimaging group analysis: A comprehensive alternative to univariate general linear model. *NeuroImage*, *99*, 571–588. <https://doi.org/10.1016/j.neuroimage.2014.06.027>

- Chica, A. B., & Lupiáñez, J. (2004). Inhibition of return without return attention. *Psicothema*, *16*(2), 248–254.
- Cichy, R. M., & Pantazis, D. (2017). Multivariate pattern analysis of MEG and EEG: A comparison of representational structure in time and space. *NeuroImage*, *158*, 441–454. <https://doi.org/10.1016/j.neuroimage.2017.07.023>
- Cichy, R. M., Pantazis, D., & Oliva, A. (2014). Resolving human object recognition in space and time. *Nature Neuroscience*, *17*(3), 455–462. <https://doi.org/10.1038/nn.3635>
- Cichy, R. M., Ramirez, F., & Pantazis, D. (2015). Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *NeuroImage*, *121*, 193–204. <https://doi.org/10.1016/j.neuroimage.2015.07.011>
- Cleveland, W. S. (2001). Data Science: an Action Plan for Expanding the Technical Areas of the Field of Statistics. *International Statistical Review*, *69*(1), 21–26. <https://doi.org/10.1111/j.1751-5823.2001.tb00477.x>
- Cowles, M., & Davis, C. (1982). On the origins of the .05 level of statistical significance. *American Psychologist*, *37*(5), 553–558. <https://doi.org/10.1037/0003-066x.37.5.553>
- Cox, D. G., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) “brain reading”: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *NeuroImage*, *19*(2), 261–270. [https://doi.org/10.1016/s1053-8119\(03\)00049-1](https://doi.org/10.1016/s1053-8119(03)00049-1)
- Craddock, R. C., Holtzheimer, P. E., Hu, X., & Mayberg, H. S. (2009). Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine*, *62*(6), 1619–1628. <https://doi.org/10.1002/mrm.22159>
- Curran, E., & Arden, N. K. (2003). Learning to control brain activity: A review of the production and control of EEG components for driving brain–computer interface (BCI) systems. *Brain and Cognition*, *51*(3), 326–336. [https://doi.org/10.1016/s0278-2626\(03\)00036-8](https://doi.org/10.1016/s0278-2626(03)00036-8)
- Danziger, S., & Kingstone, A. (1999). Unmasking the inhibition of return phenomenon. *Percept Psychophys*, *61*(6), 1024–1037. <https://doi.org/10.3758/bf03207610>
- Donoho, D. L. (2017). 50 Years of Data Science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. <https://doi.org/10.1080/10618600.2017.1384734>
- Dukewich, K. R., & Klein, R. M. (2015). Inhibition of return: A phenomenon in search of a definition and a theoretical framework. *Atten Percept Psychophys*, *77*(5), 1647–1658. <https://doi.org/10.3758/s13414-015-0835-3>
- Efron, B., & Hastie, T. (2016). *Computer Age Statistical Inference: Algorithms, Evidence, and Data Science*. <http://dx.doi.org/10.1017/cbo9781316576533>
- Efron, B., & Tibshirani, R. (1991). Statistical Data Analysis in the Computer Age. *Science*, *253*(5018), 390–395. <https://doi.org/10.1126/science.253.5018.390>
- Efron, B., & Tibshirani, R. (1994). *An Introduction to the Bootstrap*. CRC Press.
- Eriksen, C. W., & Collins, J. J. (1969). Temporal course of selective attention. *Journal of Experimental Psychology*, *80*(2, Pt.1), 254–261. <https://doi.org/10.1037/h0027268>

- Fahrenfort, J. J., Van Driel, J., Van Gaal, S., & Olivers, C. N. L. (2018). From ERPs to MVPA Using the Amsterdam Decoding and Modeling Toolbox (ADAM). *Frontiers in Neuroscience*, *12*. <https://doi.org/10.3389/fnins.2018.00368>
- Friedman, J. H. (2001). The Role of Statistics in the Data Revolution? *International Statistical Review*, *69*(1), 5–10. <https://doi.org/10.1111/j.1751-5823.2001.tb00474.x>
- Friston, K. J., Chu, C., Mourao-Miranda, J., Hulme, O. J., Rees, G., Penny, W. D., & Ashburner, J. (2008). Bayesian decoding of brain images. *NeuroImage*, *39*(1), 181–205. <https://doi.org/10.1016/j.neuroimage.2007.08.013>
- Fu, W. L. (2023, March 26). *Cross-Cultural Investigation of The Relation Between Attentional Disengagement and Reading Performance: Behavioral and Electrophysiological Evidence from Chinese and Italian*. <https://hdl.handle.net/11577/3474172>
- Gessell, B., Stanley, M. L., & De Brigard, F. (2021). Multivariate pattern analysis and the search for neural representations. *Synthese*, *199*(5–6), 12869–12889. <https://doi.org/10.1007/s11229-021-03358-3>
- Gigerenzer, G. (2004). Mindless statistics. *J. Soc. Econ*, *33*(5), 587–606. <https://doi.org/10.1016/j.socec.2004.09.033>
- Gramfort, A. (2013). MEG and EEG data analysis with MNE-Python. *Frontiers in Neuroscience*, *7*. <https://doi.org/10.3389/fnins.2013.00267>
- Groetswagers, T., Wardle, S. G., & Carlson, T. A. (2016). Decoding Dynamic Brain Patterns from Evoked Responses: A Tutorial on Multivariate Pattern Analysis Applied to Time Series Neuroimaging Data. *Journal of Cognitive Neuroscience*, *29*(4), 677–697. https://doi.org/10.1162/jocn_a_01068
- Hanke, M., Halchenko, Y. O., Sederberg, P. B., Olivetti, E., Fründ, I., Rieger, J. W., Herrmann, C., Haxby, J. V., Hanson, S. R., & Pollmann, S. (2009). PyMVPA: a unifying approach to the analysis of neuroscientific data. *Frontiers in Neuroinformatics*, *3*. <https://doi.org/10.3389/neuro.11.003.2009>
- Hanson, S. R., Matsuka, T., & Haxby, J. V. (2004). Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: is there a “face” area? *NeuroImage*, *23*(1), 156–166. <https://doi.org/10.1016/j.neuroimage.2004.05.020>
- Haufe, S., Meinecke, F. C., Görgen, K., Dähne, S., Haynes, J., Blankertz, B., & Bießmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *NeuroImage*, *87*, 96–110. <https://doi.org/10.1016/j.neuroimage.2013.10.067>
- Haxby, J. V. (2012). Multivariate pattern analysis of fMRI: The early beginnings. *NeuroImage*, *62*(2), 852–855. <https://doi.org/10.1016/j.neuroimage.2012.03.016>
- Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. <https://doi.org/10.1126/science.1063736>
- Haynes, J. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, *87*(2), 257–270. <https://doi.org/10.1016/j.neuron.2015.05.025>

- Haynes, J., & Rees, G. (2005). Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nature Neuroscience*, *8*(5), 686–691. <https://doi.org/10.1038/nn1445>
- Haynes, J., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. <https://doi.org/10.1038/nrn1931>
- Haynes, J., Sakai, K., Rees, G., Gilbert, S. J., Frith, C. D., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology*, *17*(4), 323–328. <https://doi.org/10.1016/j.cub.2006.11.072>
- Hebart, M. N., & Baker, C. I. (2018). Deconstructing multivariate decoding for the study of brain function. *NeuroImage*, *180*, 4–18. <https://doi.org/10.1016/j.neuroimage.2017.08.005>
- Hebart, M. N., Görden, K., & Haynes, J. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, *8*. <https://doi.org/10.3389/fninf.2014.00088>
- Hilchey, M. D., Dohmen, D., Crowder, N. A., & Klein, R. M. (2016). When is inhibition of return input- or output-based? It depends on how you look at it. *Canadian Journal of Experimental Psychology*, *70*(4), 325–334. <https://doi.org/10.1037/cep0000075>
- Hogendoorn, H., Verstraten, F. a. J., & Cavanagh, P. (2015). Strikingly rapid neural basis of motion-induced position shifts revealed by high temporal-resolution EEG pattern classification. *Vision Research*, *113*, 1–10. <https://doi.org/10.1016/j.visres.2015.05.005>
- Hong, X., Bo, K., Meyyappan, S., Tong, S., & Ding, M. (2020). Decoding attention control and selection in visual spatial attention. *Human Brain Mapping*, *41*(14), 3900–3921. <https://doi.org/10.1002/hbm.25094>
- Huth, A. G., Nishimoto, S., Vu, A. T., & Gallant, J. L. (2012). A Continuous Semantic Space Describes the Representation of Thousands of Object and Action Categories across the Human Brain. *Neuron*, *76*(6), 1210–1224. <https://doi.org/10.1016/j.neuron.2012.10.014>
- Isik, L., Meyers, E. M., Leibo, J. Z., & Poggio, T. (2014). The dynamics of invariant object recognition in the human visual system. *Journal of Neurophysiology*, *111*(1), 91–102. <https://doi.org/10.1152/jn.00394.2013>
- Johnson, M. R., & Johnson, M. K. (2014). Decoding individual natural scene representations during perception and imagery. *Frontiers in Human Neuroscience*, *8*. <https://doi.org/10.3389/fnhum.2014.00059>
- Jordan, M. I., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, *349*(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Kaiser, D. P., Azzalini, D., & Peelen, M. V. (2016). Shape-independent object category responses revealed by MEG and fMRI decoding. *Journal of Neurophysiology*, *115*(4), 2246–2250. <https://doi.org/10.1152/jn.01074.2015>
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685. <https://doi.org/10.1038/nn1444>

- Kanwisher, N., McDermott, J. H., & Chun, M. M. (1997). The Fusiform Face Area: A Module in Human Extrastriate Cortex Specialized for Face Perception. *The Journal of Neuroscience*, *17*(11), 4302–4311. <https://doi.org/10.1523/jneurosci.17-11-04302.1997>
- King, J., & Dehaene, S. (2014). Characterizing the dynamics of mental representations: the temporal generalization method. *Trends in Cognitive Sciences*, *18*(4), 203–210. <https://doi.org/10.1016/j.tics.2014.01.002>
- Klein, R. M. (2000). Inhibition of return. *Trends in Cognitive Sciences*, *4*(4), 138–147. [https://doi.org/10.1016/s1364-6613\(00\)01452-2](https://doi.org/10.1016/s1364-6613(00)01452-2)
- Klein, R. M., & MacInnes, W. J. (1999). Inhibition of Return is a Foraging Facilitator in Visual Search. *Psychological Science*, *10*(4), 346–352. <https://doi.org/10.1111/1467-9280.00166>
- Kriegeskorte, N. (2011). Pattern-information analysis: From stimulus decoding to computational-model testing. *NeuroImage*, *56*(2), 411–421. <https://doi.org/10.1016/j.neuroimage.2011.01.061>
- Kriegeskorte, N. (2015). Crossvalidation in brain imaging analysis. *bioRxiv*. <https://doi.org/10.1101/017418>
- Kriegeskorte, N., & Bandettini, P. A. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649–662. <https://doi.org/10.1016/j.neuroimage.2007.02.022>
- Kriegeskorte, N., & Diedrichsen, J. (2019). Peeling the Onion of Brain Representations. *Annual Review of Neuroscience*, *42*(1), 407–432. <https://doi.org/10.1146/annurev-neuro-080317-061906>
- Kriegeskorte, N., & Douglas, P. K. (2019). Interpreting encoding and decoding models. *Current Opinion in Neurobiology*, *55*, 167–179. <https://doi.org/10.1016/j.conb.2019.04.002>
- Kriegeskorte, N., Goebel, R., & Bandettini, P. A. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(10), 3863–3868. <https://doi.org/10.1073/pnas.0600244103>
- Kübler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychological Bulletin*, *127*(3), 358–375. <https://doi.org/10.1037/0033-2909.127.3.358>
- Linn, K. A., Gaonkar, B., Doshi, J., Davatzikos, C., & Shinohara, R. T. (2016). Addressing Confounding in Predictive Models with an Application to Neuroimaging. *The International Journal of Biostatistics*, *12*(1), 31–44. <https://doi.org/10.1515/ijb-2015-0030>
- López-García, D., Peñalver, J., Górriz, J. M., & Ruz, M. (2022). MVPAlab: A machine learning decoding toolbox for multidimensional electroencephalography data. *Computer Methods and Programs in Biomedicine*, *214*, 106549. <https://doi.org/10.1016/j.cmpb.2021.106549>
- Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based brain–computer interfaces. *Journal of Neural Engineering*, *4*(2), R1–R13. <https://doi.org/10.1088/1741-2560/4/2/r01>

- Lupiáñez, J., Decaix, C., Siéoff, E., Chokron, S., Milliken, B., & Bartolomeo, P. (2004). Independent effects of endogenous and exogenous spatial cueing: inhibition of return at endogenously attended target locations. *Experimental Brain Research*, *159*(4), 447–457. <https://doi.org/10.1007/s00221-004-1963-5>
- Martín-Arévalo, E., Chica, A. B., & Lupiáñez, J. (2016). No single electrophysiological marker for facilitation and inhibition of return: A review. *Behavioural Brain Research*, *300*, 1–10. <https://doi.org/10.1016/j.bbr.2015.11.030>
- Maylor, E. A., & Hockey, R. (1985). Inhibitory component of externally controlled covert orienting in visual space. *Journal of Experimental Psychology: Human Perception and Performance*, *11*(6), 777–787. <https://doi.org/10.1037/0096-1523.11.6.777>
- McCarthy, G., Puce, A., Gore, J. C., & Allison, T. (1997). Face-Specific Processing in the Human Fusiform Gyrus. *Journal of Cognitive Neuroscience*, *9*(5), 605–610. <https://doi.org/10.1162/jocn.1997.9.5.605>
- Pierce, A. K., Crouse, M. D., & Green, J. L. (2017). Evidence for an attentional component of inhibition of return in visual search. *Psychophysiology*, *54*(11), 1676–1685. <https://doi.org/10.1111/psyp.12905>
- Misaki, M., Kim, Y. H., Bandettini, P. A., & Kriegeskorte, N. (2010). Comparison of multivariate classifiers and response normalizations for pattern-information fMRI. *NeuroImage*, *53*(1), 103–118. <https://doi.org/10.1016/j.neuroimage.2010.05.051>
- Mitchell, T. M., Shinkareva, S. V., Carlson, A. P., Chang, K., Malave, V. L., Mason, R. J., & Just, M. A. (2008). Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science*, *320*(5880), 1191–1195. <https://doi.org/10.1126/science.1152876>
- Müller, K., Tangermann, M., Dornhege, G., Krauledat, M., Curio, G., & Blankertz, B. (2008). Machine learning for real-time single-trial EEG-analysis: From brain–computer interfacing to mental state monitoring. *Journal of Neuroscience Methods*, *167*(1), 82–90. <https://doi.org/10.1016/j.jneumeth.2007.09.022>
- Mur, M., Bandettini, P. A., & Kriegeskorte, N. (2009). Revealing representational content with pattern-information fMRI—an introductory guide. *Social Cognitive and Affective Neuroscience*, *4*(1), 101–109. <https://doi.org/10.1093/scan/nsn044>
- Murphy, B. R., Poesio, M., Bovolo, F., Bruzzone, L., Dalponte, M., & Lakany, H. (2011). EEG decoding of semantic category reveals distributed representations for single concepts. *Brain and Language*, *117*(1), 12–22. <https://doi.org/10.1016/j.bandl.2010.09.013>
- Naselaris, T., & Kay, K. (2015). Resolving Ambiguities of MVPA Using Explicit Models of Representation. *Trends in Cognitive Sciences*, *19*(10), 551–554. <https://doi.org/10.1016/j.tics.2015.07.005>
- Naselaris, T., Kay, K., Nishimoto, S., & Gallant, J. L. (2011). Encoding and decoding in fMRI. *NeuroImage*, *56*(2), 400–410. <https://doi.org/10.1016/j.neuroimage.2010.07.073>

- Nenadic, I., Sauer, H., & Gaser, C. (2010). Distinct pattern of brain structural deficits in subsyndromes of schizophrenia delineated by psychopathology. *NeuroImage*, *49*(2), 1153–1160. <https://doi.org/10.1016/j.neuroimage.2009.10.014>
- Neyman, J., and Pearson, E. S. On the problem of the most efficient tests for statistical hypotheses. (1933). *Philosophical Transactions of the Royal Society of London*, *231*(694–706), 289–337. <https://doi.org/10.1098/rsta.1933.0009>
- Ng, B., Reh, R. K., & Mostafavi, S. (2022). A practical guide to applying machine learning to infant EEG data. *Developmental Cognitive Neuroscience*, *54*, 101096. <https://doi.org/10.1016/j.dcn.2022.101096>
- Noble, W. S. (2006). What is a support vector machine? *Nature Biotechnology*, *24*(12), 1565–1567. <https://doi.org/10.1038/nbt1206-1565>
- Norman, K. A., Polyn, S. M., Detre, G., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>
- Norman, K. A., Quamme, J. R., & Newman, E. L. (2009). Multivariate methods for tracking cognitive states. In *Oxford University Press eBooks* (pp. 299–330). <https://doi.org/10.1093/acprof:oso/9780199217298.003.0017>
- Oosterhof, N. N., Connolly, A. J., & Haxby, J. V. (2016). CoSMoMVPA: Multi-Modal Multivariate Pattern Analysis of Neuroimaging Data in Matlab/GNU Octave. *Frontiers in Neuroinformatics*, *10*. <https://doi.org/10.3389/fninf.2016.00027>
- O’Toole, A. J., Jiang, F., Abdi, H., & Haxby, J. V. (2005). Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. *Journal of Cognitive Neuroscience*, *17*(4), 580–590. <https://doi.org/10.1162/0898929053467550>
- Peelen, M. V., Atkinson, A. B., & Vuilleumier, P. (2010). Supramodal Representations of Perceived Emotions in the Human Brain. *The Journal of Neuroscience*, *30*(30), 10127–10134. <https://doi.org/10.1523/jneurosci.2161-10.2010>
- Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier.
- Pereira, F. C., Mitchell, T. M., & Botvinick, M. (2009). Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage*, *45*(1), S199–S209. <https://doi.org/10.1016/j.neuroimage.2008.11.007>
- Pierce, A. K., Crouse, M. D., & Green, J. L. (2017). Evidence for an attentional component of inhibition of return in visual search. *Psychophysiology*, *54*(11), 1676–1685. <https://doi.org/10.1111/psyp.12905>
- Poline, J., & Brett, M. (2012). The general linear model and fMRI: Does love last forever? *NeuroImage*, *62*(2), 871–880. <https://doi.org/10.1016/j.neuroimage.2012.01.133>
- Posner, M. I. (1984). Components of visual orienting. *Attention and Performance*, *10*, 531–556. <http://ci.nii.ac.jp/naid/10015676922>

- Posner, M. I., Rafal, R. D., Choate, L. S., & Vaughan, J. (1985). Inhibition of return: Neural basis and function. *Cognitive Neuropsychology*, *2*(3), 211–228. <https://doi.org/10.1080/02643298508252866>
- Posner, M. I., Snyder, C. L., & Davidson, B. R. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General*, *109*(2), 160–174. <https://doi.org/10.1037/0096-3445.109.2.160>
- Pratt, J., & Fischer, M. R. (2002). Examining the role of the fixation cue in inhibition of return. *Canadian Journal of Experimental Psychology*, *56*(4), 294–301. <https://doi.org/10.1037/h0087405>
- Puce, A., Allison, T., Asgari, M. M., Gore, J. C., & McCarthy, G. (1996). Differential Sensitivity of Human Visual Cortex to Faces, Letterstrings, and Textures: A Functional Magnetic Resonance Imaging Study. *The Journal of Neuroscience*, *16*(16), 5205–5215. <https://doi.org/10.1523/jneurosci.16-16-05205.1996>
- Redden, R. S., Hurst, A., & Klein, R. M. (2018). Dissociating two forms of inhibition of return using temporal order judgments. *Journal of Vision*. <https://doi.org/10.1167/18.10.1183>
- Ritchie, J. L., Kaplan, D. L., & Klein, C. (2017). Decoding the Brain: Neural Representation and the Limits of Multivariate Pattern Analysis in Cognitive Neuroscience. *The British Journal for the Philosophy of Science*, *70*(2), 581–607. <https://doi.org/10.1093/bjps/axx023>
- Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. <https://doi.org/10.1037/h0042519>
- Russell, S., Russell, S. J., Norvig, P., & Davis, E. (2010). *Artificial Intelligence: A Modern Approach*. Prentice Hall.
- Sapir, A., Hayes, A. R., Henik, A., Danziger, S., & Rafal, R. D. (2004). Parietal Lobe Lesions Disrupt Saccadic Remapping of Inhibitory Location Tagging. *Journal of Cognitive Neuroscience*, *16*(4), 503–509. <https://doi.org/10.1162/089892904323057245>
- Satel, J., Hilchey, M. D., Wang, Z., Reiss, C. W., & Klein, R. M. (2014). In search of a reliable electrophysiological marker of oculomotor inhibition of return. *Psychophysiology*, *51*(10), 1037–1045. <https://doi.org/10.1111/psyp.12245>
- Satel, J., Wilson, N., & Klein, R. M. (2019). What Neuroscientific Studies Tell Us about Inhibition of Return. *Vision*, *3*(4), 58. <https://doi.org/10.3390/vision3040058>
- Scheffé, H. (1999). *The Analysis of Variance*. John Wiley & Sons.
- Schrouff, J., & Phillips, C. (2012). Multivariate Pattern Recognition Analysis: Brain Decoding. In *Coma and Disorders of Consciousness*. https://doi.org/10.1007/978-1-4471-2440-5_4
- Schrouff, J., Rosa, M. J., Rondina, J. M., Marquand, A. F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., & Mourao-Miranda, J. (2013). PRoNTO: Pattern Recognition for Neuroimaging Toolbox. *Neuroinformatics*, *11*(3), 319–337. <https://doi.org/10.1007/s12021-013-9178-1>

- Schwarzkopf, D. S., & Rees, G. (2011). Pattern classification using functional magnetic resonance imaging. *Wiley Interdisciplinary Reviews: Cognitive Science*, 2(5), 568–579. <https://doi.org/10.1002/wcs.141>
- Serences, J. T., & Boynton, G. M. (2007). The Representation of Behavioral Choice for Motion in Human Visual Cortex. *The Journal of Neuroscience*, 27(47), 12893–12899. <https://doi.org/10.1523/jneurosci.4021-07.2007>
- Simanova, I., Van Gerven, M. a. J., Oostenveld, R., & Hagoort, P. (2010). Identifying Object Categories from Event-Related EEG: Toward Decoding of Conceptual Representations. *PLOS ONE*, 5(12), e14465. <https://doi.org/10.1371/journal.pone.0014465>
- Simanova, I., Van Gerven, M. a. J., Oostenveld, R., & Hagoort, P. (2015). Predicting the Semantic Category of Internally Generated Words from Neuromagnetic Recordings. *Journal of Cognitive Neuroscience*. https://doi.org/10.1162/jocn_a_00690
- Sona, D., Veeramachaneni, S., Olivetti, E., & Avesani, P. (2007). Inferring Cognition from fMRI Brain Images. In *Lecture Notes in Computer Science* (pp. 869–878). Springer Science+Business Media. https://doi.org/10.1007/978-3-540-74695-9_89
- Stokes, M. A., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. S. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron*, 78(2), 364–375. <https://doi.org/10.1016/j.neuron.2013.01.039>
- Tassinari, G., Aglioti, S. M., Chelazzi, L., Marzi, C. A., & Berlucchi, G. (1987). Distribution in the visual field of the costs of voluntarily allocated attention and of the inhibitory after-effects of covert orienting. *Neuropsychologia*, 25(1), 55–71. [https://doi.org/10.1016/0028-3932\(87\)90043-1](https://doi.org/10.1016/0028-3932(87)90043-1)
- The Tumultuous History of the Search for Artificial Intelligence, Daniel Crevier. 1993. Basic Books, New York, NY. 432 pages. ISBN: 0-465-02997-3. >7.50. (1994). *Bulletin of Science, Technology & Society*, 14(4), 224. <https://doi.org/10.1177/027046769401400414>
- Tong, F., & Pratte, M. S. (2012). Decoding Patterns of Human Brain Activity. *Annual Review of Psychology*, 63(1), 483–509. <https://doi.org/10.1146/annurev-psych-120710-100412>
- Treder, M. S. (2020). MVPA-Light: A Classification and Regression Toolbox for Multi-Dimensional Data. *Frontiers in Neuroscience*, 14. <https://doi.org/10.3389/fnins.2020.00289>
- Treisman, A., & Gelade, G. A. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)
- Tukey, J. W. (1962). The Future of Data Analysis. *Annals of Mathematical Statistics*, 33(1), 1–67. <https://doi.org/10.1214/aoms/1177704711>
- Van Gerven, M. a. J., Maris, E., Sperling, M. R., Sharan, A., Litt, B., Anderson, C., Baltuch, G. H., & Jacobs, J. J. (2013). Decoding the memorization of individual stimuli with direct human brain recordings. *NeuroImage*, 70, 223–232. <https://doi.org/10.1016/j.neuroimage.2012.12.059>
- Vapnik, V., & Chervonenkis, A. Y. (1982). Necessary and Sufficient Conditions for the Uniform Convergence of Means to their Expectations. *Theory of Probability and Its Applications*, 26(3), 532–553. <https://doi.org/10.1137/1126059>

- Vidal, J. J. (1973). Toward direct brain-computer communication. *Annual review of Biophysics and Bioengineering*, 2(1), 157-180.
- Wang, J., Baucom, L. B., & Shinkareva, S. V. (2012). Decoding abstract and concrete concept representations based on single-trial fMRI data. *Human Brain Mapping*, 34(5), 1133–1147. <https://doi.org/10.1002/hbm.21498>
- Wardle, S. G., Kriegeskorte, N., Grootswagers, T., Khaligh-Razavi, S., & Carlson, T. A. (2016). Perceptual similarity of visual patterns predicts dynamic neural activation patterns measured with MEG. *NeuroImage*, 132, 59–70. <https://doi.org/10.1016/j.neuroimage.2016.02.019>
- What's on your mind? (2006). *Nature Neuroscience*, 9(8), 981. <https://doi.org/10.1038/nn0806-981>
- Wolff, M. J., Ding, J., Myers, N. D., & Stokes, M. A. (2015). Revealing hidden states in visual working memory using electroencephalography. *Frontiers in Systems Neuroscience*, 9. <https://doi.org/10.3389/fnsys.2015.00123>
- Wolpaw, J. R., Birbaumer, N., McFarland, D. J., Pfurtscheller, G., & Vaughan, T. M. (2002). Brain–computer interfaces for communication and control. *Clinical Neurophysiology*, 113(6), 767–791. [https://doi.org/10.1016/s1388-2457\(02\)00057-3](https://doi.org/10.1016/s1388-2457(02)00057-3)
- Wolpaw, J., & Wolpaw, E. W. (2012). *Brain-Computer Interfaces: Principles and Practice*. Oxford University Press.
- Wolpert, D. H. (1996). The Lack of A Priori Distinctions Between Learning Algorithms. *Neural Computation*, 8(7), 1341–1390. <https://doi.org/10.1162/neco.1996.8.7.1341>
- Xanthopoulos, P., Pardalos, P. M., & Trafalis, T. B. (2012). Linear Discriminant Analysis. In *Springer eBooks* (pp. 27–33). https://doi.org/10.1007/978-1-4419-9878-1_4
- Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics*, 1168, 022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>