

**POLITECNICO DI MILANO**  
**Facoltà di Ingegneria dei Sistemi**  
**Corso di Studi in INGEGNERIA MATEMATICA**



**TESI DI LAUREA DI SECONDO LIVELLO**

**Analisi statistica dei segnali elettrici cerebrali in  
pazienti in stato vegetativo e di minima coscienza**

**RELATORI:**

**Dott. Simone VANTINI**

**Dott.ssa Laura M. SANGALLI**

**Candidato:**  
**Marco ANTONINI**  
**Matricola 755168**

**ANNO ACCADEMICO 2011-2012**

## Sommario

Questo lavoro di tesi si inserisce nell'ambito di una collaborazione scientifica tra l'Unità di Ingegneria Biomedica dell'Istituto Neurologico "Carlo Besta" di Milano e il Laboratorio di Modellistica e Calcolo Scientifico (MOX) del Dipartimento di Matematica "F. Brioschi" del Politecnico di Milano, all'interno del progetto di ricerca *Coma Research Center*. I dati analizzati hanno riguardato EEG in sonno e in veglia di 19 pazienti in stato vegetativo o di minima coscienza, registrati per mezzo di 19 elettrodi applicati sullo scalpo secondo il Sistema Internazionale 10-20 e riferiti ai lobi auricolari. Un'analisi multivariata delle componenti principali è stata effettuata sui dati ricostruiti via wavelets e standardizzati, al fine di indagare la struttura di correlazione dei 19 canali di acquisizione del segnale EEG. Ciò ha permesso di trovare, per tutti gli EEG in esame, spazi tridimensionali su cui proiettare i dati, individuati da componenti dallo stesso significato fisico, ma con importanti differenze per quanto riguarda i loadings ad esse associati. L'analisi dei loadings ha inoltre messo in evidenza per alcuni soggetti la presenza di aree spente di cervello. Tali zone (non direttamente osservabili esaminando i tracciati dei segnali registrati dai 19 canali) potrebbero essere la chiave per una corretta distinzione dei pazienti in relazione al loro stato di coscienza.

## **Abstract**

This thesis is part of a scientific collaboration between the Biomedical Engineering Unit of the Neurological Institute “Carlo Besta” in Milan and the Laboratory for Modeling and Scientific Computing (MOX) of the Department of Mathematics “F. Brioschi” at Politecnico di Milano, within the *Coma Research Center* project. The analyzed data concerned sleep and waking EEGs of 19 patients in a vegetative state or in a minimally conscious state, recorded from 19 electrodes placed on the scalp according to the International Federation 10-20 system and referred to the ear lobes. A multivariate principal component analysis was performed on the data reconstructed by wavelets and then standardized, in order to investigate the correlation structure of the 19 channels of EEG signal. This has allowed to find, for all the studied EEGs, three-dimensional spaces on which to project the data, identified by components with the same physical meaning, but with important differences as regards the loadings associated with them. The analysis of loadings has also highlighted for some subjects the presence of switched off areas of the brain. These areas (not directly observable by examining the graphs of the signals recorded from 19 channels) may be the key to a proper distinction of patients with regard to their state of consciousness.

# Indice

<b>Introduzione</b>	<b>1</b>
<b>1 Disordini di coscienza e ElettroEncefaloGrafia</b>	<b>5</b>
1.1 Disordini di coscienza . . . . .	5
1.1.1 Coma . . . . .	8
1.1.2 Stato vegetativo . . . . .	8
1.1.3 Stato di minima coscienza . . . . .	10
1.1.4 Locked-in Syndrome . . . . .	11
1.1.5 Diagnosi dei disordini di coscienza . . . . .	11
1.1.5.1 Coma Recovery Scale-Revised (CRS-R) . . . . .	13
1.1.5.2 Strumenti complementari di diagnosi . . . . .	13
1.2 ElettroEncefaloGrafia . . . . .	14
1.2.1 Metodi di registrazione . . . . .	15
1.2.2 La struttura del sonno . . . . .	18
<b>2 Metodi</b>	<b>21</b>
2.1 La trasformata discreta di Fourier . . . . .	22
2.2 Le wavelets . . . . .	24
2.2.1 Le wavelets ortogonali di Daubechies . . . . .	26
2.2.2 Stimatore wavelet non lineare di un dato funzionale . . . . .	27
2.2.3 Estensione al caso di funzioni multidimensionali . . . . .	29
2.2.3.1 Errori correlati nelle p dimensioni . . . . .	31
2.3 Analisi delle Componenti Principali (PCA) . . . . .	32

---

2.3.1	Componenti Principali per variabili standardizzate . . .	34
2.4	Cluster Analysis . . . . .	35
2.4.1	Algoritmi gerarchici . . . . .	36
2.4.1.1	Metodi gerarchici agglomerativi . . . . .	37
<b>3</b>	<b>Analisi</b>	<b>40</b>
3.1	Smoothing dei dati . . . . .	40
3.2	Standardizzazione dei dati e PCA . . . . .	46
3.3	Analisi spettrale . . . . .	50
3.4	Cluster Analysis . . . . .	55
<b>4</b>	<b>Conclusioni</b>	<b>57</b>
<b>A</b>	<b>Figure e Tabelle del Capitolo 3</b>	<b>59</b>
	<b>Bibliografia</b>	<b>94</b>

# Elenco delle figure

1.1	(Da [12]). Illustrazione delle due principali componenti della coscienza in stati fisiologici normali e patologici. Nel primo caso le due componenti sono in genere positivamente correlate.	7
1.2	(Da [12]). Danni cerebrali acuti possono comportare l'entrata in coma. In genere il coma dura pochi giorni, e una volta che il paziente apre gli occhi entra in uno stato vegetativo. Dopo aver mostrato i primi segni di coscienza il paziente entra in stato di minima coscienza, e alla fine recupera piena coscienza. In rari casi una persona può essere affetta da locked-in syndrome.	8
1.3	Sistema Internazionale 10-20. A ogni elettrodo è associata una sigla che ne identifica in modo univoco la posizione sullo scalpo. Le lettere identificano la regione della corteccia esplorata (Fp: frontopolare; F: frontale; C: centrale; P: parietale; T: temporale; O: occipitale), mentre i numeri (o la lettera z) indicano l'emisfero (numeri dispari: sinistra; numeri pari: destra; z: linea mediana).	16
1.4	Diversi tipi di onde elettroencefalografiche normali.	17
1.5	All'apertura degli occhi il ritmo alfa viene sostituito da onde asincrone di basso voltaggio.	18
1.6	Variazioni del segnale EEG in veglia e in sonno.	20

2.1	(Da [22]). Esempio in tre dimensioni della procedura di soft-thresholding descritta dalla (2.15): lo shrinkage agisce solo sulla norma del vettore $\mathbf{d}_{j,k}$ , preservandone la direzione. . . .	31
3.1	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, ottenuto interpolando linearmente le istanze di campionamento. . . . .	42
3.1	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets. . . . .	43
3.2	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, ottenuto interpolando linearmente le istanze di campionamento. . . . .	44
3.2	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets. . . . .	45
3.3	Autovalori delle matrici di correlazione dei canali relative agli EEG registrati in sonno. . . . .	47
3.4	Autovalori delle matrici di correlazione dei canali relative agli EEG registrati in veglia. . . . .	48
3.5	Loadings delle Componenti A, B e C relative agli EEG in sonno e in veglia del Paziente 10. I numeri in alto a sinistra fanno riferimento alle quote di variabilità spiegata dalle tre componenti. I canali delle componenti B e C sono stati riordinati per evidenziare i contrasti. . . . .	49
3.6	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG in sonno e in veglia del Paziente 10. . . . .	50
3.7	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets e la standardizzazione globale dei canali. . . . .	51
3.7	Ricostruzione del segnale raffigurato in Figura 3.7.(a) utilizzando le Componenti A, B e C. . . . .	52
3.8	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets e la standardizzazione globale dei canali. . . . .	53

---

3.8	Ricostruzione del segnale raffigurato in Figura 3.8.(a) utilizzando le Componenti A, B e C. . . . .	54
3.9	Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Paziente 10) . . . . .	55
A.1	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, ottenuto interpolando linearmente le istanze di campionamento. . . . .	60
A.1	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets. . . . .	61
A.1	Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets e la standardizzazione globale dei canali. . . . .	62
A.1	Ricostruzione del segnale raffigurato in Figura A.1.(c) utilizzando le Componenti A, B e C. . . . .	63
A.2	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, ottenuto interpolando linearmente le istanze di campionamento. . . . .	64
A.2	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets. . . . .	65
A.2	Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets e la standardizzazione globale dei canali. . . . .	66
A.2	Ricostruzione del segnale raffigurato in Figura A.2.(c) utilizzando le Componenti A, B e C. . . . .	67
A.3	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 1 a 7) . . . .	68



A.3	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 8 a 14) . . .	69
A.3	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti (Pazienti da 15 a 19) . . .	70
A.4	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 1 a 7) . . .	71
A.4	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 8 a 14) . . .	72
A.4	Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 15 a 19) . . .	73
A.5	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 1 a 5) . . . . .	74
A.5	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 6 a 10) . . . . .	75
A.5	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 11 a 15) . . . . .	76

---

A.5	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 16 a 19) . . . . .	77
A.6	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 1 a 5) . . . . .	78
A.6	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 6 a 10) . . . . .	79
A.6	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 11 a 15) . . . . .	80
A.6	I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 16 a 19) . . . . .	81
A.7	Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 1 a 5) . . . . .	82
A.7	Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 6 a 10) . . . . .	83
A.7	Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 11 a 15) . . . . .	84

---

A.7	Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 16 a 19) . . . . .	85
A.8	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (0s,16s)) . . . . .	86
A.8	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (16s,32s)) . . . . .	87
A.8	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (32s,48s)) . . . . .	88
A.8	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (48s,64s)) . . . . .	89
A.9	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (0s,16s)) . . . . .	90

---

A.9	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (16s,32s)) . . . . .	91
A.9	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (32s,48s)) . . . . .	92
A.9	Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (48s,64s)) . . . . .	93

# Elenco delle tabelle

A.1	Identificativo, stato di coscienza e punteggio CRS-R dei pazienti esaminati. . . . .	59
-----	--	----

# Introduzione

Questo lavoro di tesi si inserisce nell'ambito di una collaborazione scientifica tra l'Unità di Ingegneria Biomedica dell'Istituto Neurologico "Carlo Besta" di Milano e il Laboratorio di Modellistica e Calcolo Scientifico (MOX) del Dipartimento di Matematica "F. Brioschi" del Politecnico di Milano, all'interno del progetto di ricerca *Coma Research Center (CRC)*. L'obiettivo del progetto è quello di sintetizzare l'informazione contenuta negli elettroencefalogrammi (EEG) dei pazienti in stato vegetativo (VS) e di minima coscienza (MCS), con lo scopo di trovare dei profili o degli indici che permettano di descrivere e discriminare i pazienti appartenenti a queste categorie di coscienza, e pertanto fornire un supporto di tipo quantitativo al clinico che sia di aiuto per uniformare i criteri di diagnosi di questi stati. In particolare, questo elaborato si propone come obiettivo principale quello di indagare la struttura di correlazione dei 19 canali di acquisizione dell'EEG osservati nel tempo, al fine di effettuare una riduzione dimensionale dei dati che consenta di ottenere informazioni non direttamente ricavabili esaminando i tracciati dei segnali elettroencefalografici.

Dal punto di vista neurologico, la coscienza è caratterizzata da due componenti: la vigilanza (*arousal*) e la consapevolezza (*awareness*) di sé e dell'ambiente. Lo stato di coscienza è stabilito dal buon funzionamento delle due componenti. Il coma è caratterizzato dall'assenza di vigilanza, e pertanto anche di consapevolezza, mentre una persona in stato vegetativo recupera vigilanza ma non mostra alcun segno di consapevolezza. I pazienti in stato di minima coscienza sono vigili e inoltre mostrano segni apprezzabili, seppur

fluttuanti, di consapevolezza.

Nella riabilitazione neurologica una corretta distinzione tra stato vegetativo e di minima coscienza è di grande importanza, sia per le implicazioni in termini di prognosi e scelte terapeutiche, sia a livello medico-legale ed etico. Infatti, nonostante l'introduzione di criteri diagnostici, sono ad oggi ancora frequenti i casi di diagnosi errate dovuti alla difficoltà di rilevare e interpretare correttamente i segni di coscienza, e ciò può compromettere il recupero potenziale del paziente.

Lo studio dei disordini di coscienza pone pertanto grandi sfide in materia di diagnosi, prognosi e cure quotidiane. I progressi tecnologici in materia di neuroimaging e di elettrofisiologia infatti hanno contribuito ad aumentare la comprensione dei meccanismi del cervello umano, ma ad oggi non esistono ancora adeguate evidenze che le metodiche neurofisiologiche permettano di produrre profili di risposta in grado di caratterizzare correttamente i pazienti in stato vegetativo o in stato di minima coscienza, e soprattutto che possano essere impiegate a livello individuale.

Premesso ciò, i dati analizzati nel lavoro di tesi riguardano elettroencefalogrammi in sonno e in veglia di 19 pazienti affetti da disordini di coscienza (9 in stato vegetativo e 10 in stato di minima coscienza), registrati per mezzo di 19 elettrodi applicati sullo scalpo in posizioni opportune secondo il Sistema Internazionale 10-20 e riferiti ai lobi auricolari; possiamo quindi interpretarli dal punto di vista statistico come funzioni aleatorie  $\mathbf{f} : \mathbb{R} \mapsto \mathbb{R}^p$ ,  $p = 19$ , osservate discretamente e affette da errore di misurazione. È pertanto necessario introdurre una base funzionale che permetta di ricostruire opportunamente il dato a partire dalle sue istanze di campionamento, effettuando un'operazione di *smoothing* tale da filtrare la parte di variabilità non rilevante ai fini dell'analisi (il rumore) e che consenta inoltre di sopperire alla mancanza di regolarità tipica delle osservazioni discrete.

Sono molte in letteratura le tecniche proposte per la stima dei dati funzionali. A tal proposito, in questo lavoro abbiamo scelto di utilizzare le wavelets ortogonali di Daubechies di ordine 10 (cioè con 10 momenti nulli) per via delle loro proprietà di regolarità. Lo stimatore utilizzato è quello proposto da Pigoli e Sangalli in [22] nel caso di errori normali multivariati i.i.d. di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $S$  generica e incognita; tale

stimatore infatti tiene conto della natura multidimensionale dei dati e della correlazione tra le componenti di errore nelle  $p$  direzioni tipica dei segnali biomedici tra cui l'EEG, e inoltre utilizza la stessa base wavelets per stimare tutte le coordinate del segnale.

La successiva analisi multivariata delle componenti principali, effettuata al fine di indagare la struttura di correlazione dei canali, ha permesso di trovare per tutti gli EEG in esame spazi tridimensionali su cui proiettare i dati dal chiaro significato fisico; come vedremo, infatti, la prima componente che individua tali spazi rappresenta una media dell'attività cerebrale registrata dai 19 canali, mentre la seconda e la terza contrastano rispettivamente l'attività registrata negli emisferi destro-sinistro e nelle parti anteriore-posteriore degli emisferi. I dati ricostruiti a partire dalle tre componenti individuate e i relativi spettri d'ampiezza confermano la bontà dell'approssimazione effettuata.

Infine, analizzando graficamente i loadings delle tre componenti principali, osserveremo come alcuni pazienti presentino delle zone spente di cervello. Il prossimo passo dell'analisi, che richiederà l'implementazione di tecniche specifiche e che è segnalato come sviluppo futuro del lavoro di tesi, sarà pertanto finalizzato alla ricerca di una metrica che permetta di raggruppare tra loro pazienti che manifestano la presenza di zone spente del cervello, non necessariamente le stesse.

La struttura della trattazione è sintetizzata di seguito:

- **Capitolo 1:** *Disordini di coscienza e ElettroEncefaloGrafia.*

In questo capitolo è fornita una panoramica dell'ambito clinico in cui si inserisce il lavoro di tesi, focalizzando l'attenzione sui disordini di coscienza e sull'ElettroEncefalografia. Vengono pertanto presentati i possibili disordini dello stato di coscienza che si possono presentare in seguito a danni cerebrali acuti e alcune delle più comuni tecniche di diagnosi. Sono inoltre descritti i metodi di registrazione del segnale elettroencefalografico, le attività ritmiche caratteristiche e la struttura del sonno.



- **Capitolo 2:** *Metodi.*

Vengono qui brevemente presentate le tecniche utilizzate nello studio del problema clinico in esame, focalizzando in particolar modo l'attenzione sui principali risultati e sulle proprietà di interesse per il lavoro di tesi: la trasformata discreta di Fourier per lo studio in frequenza del segnale, le wavelets per lo smoothing del dato funzionale multidimensionale, l'analisi delle componenti principali per la riduzione dimensionale del problema, e infine la cluster analysis per il raggruppamento dei soggetti.

- **Capitolo 3:** *Analisi.*

In questo capitolo vengono presentati in modo dettagliato i dati a disposizione, la loro modellizzazione in termini matematico-statistici, le analisi effettuate e i risultati ottenuti.

- **Capitolo 4:** *Conclusioni.*

In questo capitolo vengono riassunti i risultati ottenuti durante il lavoro di tesi. Sono inoltre citati alcuni possibili campi di sviluppo per il lavoro svolto.

- **Appendice A:** *Figure e Tabelle del Capitolo 3.*

Vengono qui riportate le figure del Capitolo 3, oltre che una tabella riassuntiva delle informazioni a disposizione sugli stati di coscienza dei pazienti in esame.

# Disordini di coscienza e ElettroEncefaloGrafia

In questo capitolo è fornita una panoramica dell'ambito clinico in cui si inserisce il lavoro di tesi, focalizzando l'attenzione sui disordini di coscienza e sull'ElettroEncefalografia. Il lettore interessato ad approfondimenti può consultare, ad esempio, [12],[9],[17],[18],[19],[11] o [27].

Dal punto di vista neurologico la coscienza è caratterizzata da due componenti: la vigilanza (*arousal*) e la consapevolezza (*awareness*) di sé e dell'ambiente. Lo stato di coscienza è stabilito dal buon funzionamento delle due componenti. Il coma è caratterizzato dall'assenza di vigilanza, e pertanto anche di consapevolezza, mentre una persona in stato vegetativo (VS) recupera vigilanza ma non mostra alcun segno di consapevolezza. Nello stato di minima coscienza (MCS) invece i pazienti sono vigili e inoltre mostrano segni apprezzabili, seppur fluttuanti, di consapevolezza.

L'ElettroEncefaloGrafia è la registrazione in modo continuo e non invasivo dell'attività elettrica cerebrale per mezzo di elettrodi posti sul cuoio capelluto.

## 1.1 Disordini di coscienza

In anni recenti le tecniche di rianimazione hanno portato ad un notevole aumento del numero di pazienti che sopravvivono a lesioni cerebrali gravi.

Alcuni pazienti recuperano coscienza nei primi giorni dopo l'evento, mentre altri muoiono rapidamente; altri pazienti invece la recuperano più lentamente, in parte o del tutto, attraverso diverse fasi. Lo studio dei disordini di coscienza pone pertanto grandi sfide in materia di diagnosi, prognosi e cure quotidiane. Infatti, i segni di coscienza non sono sempre facili da rilevare e, in alcuni casi, possono anche non essere notati; ciò può pertanto comportare prognosi e scelte terapeutiche sbagliate, e compromettere di conseguenza il recupero potenziale del paziente.

Ad oggi non esiste alcuna definizione universalmente condivisa di coscienza, e inoltre la coscienza non può essere misurata da alcuna indagine strumentale. La definizione più accettata è di tipo operativo e fa riferimento in particolare a due componenti: la vigilanza (*arousal*) e la consapevolezza (*awareness*).

La vigilanza è dedotta clinicamente osservando la presenza di apertura degli occhi da parte del paziente e, a livello neuro-anatomico, indagando la conservazione delle funzioni talamiche e tronco-encefaliche.

La consapevolezza fa riferimento invece alla percezione cosciente, comprendente la cognizione, le esperienze del passato e del presente, e le intenzioni. A livello clinico è principalmente dedotta per mezzo di comandi come la richiesta di muovere le dita o di chiudere gli occhi, mentre a livello neuro-anatomico è dedotta indagando l'attività della corteccia cerebrale e della rete fronto-parietale. La consapevolezza può essere ulteriormente suddivisa in consapevolezza dell'ambiente oppure di sé; la prima può essere definita come la percezione cosciente del proprio ambiente attraverso modalità di tipo sensoriale (visive, olfattive, uditive), mentre invece la seconda non è correlata a stimoli esterni e fa riferimento anche alla conoscenza della propria storia sociale e culturale oltre che della propria famiglia.

Per essere consapevoli bisogna essere svegli, ma quando si è svegli non necessariamente si è consapevoli; infatti, come accennato sopra, la coscienza dipende dall'interazione tra l'attività della corteccia cerebrale, del tronco cerebrale e del talamo, e pertanto risulta compromessa quando una di queste attività viene a mancare. Dunque, come mostrato in Figura 1.1, la coscienza è caratterizzata da diversi stati, i quali non sempre sono nettamente separati fra loro e spesso rappresentano stati progressivi di transizione.

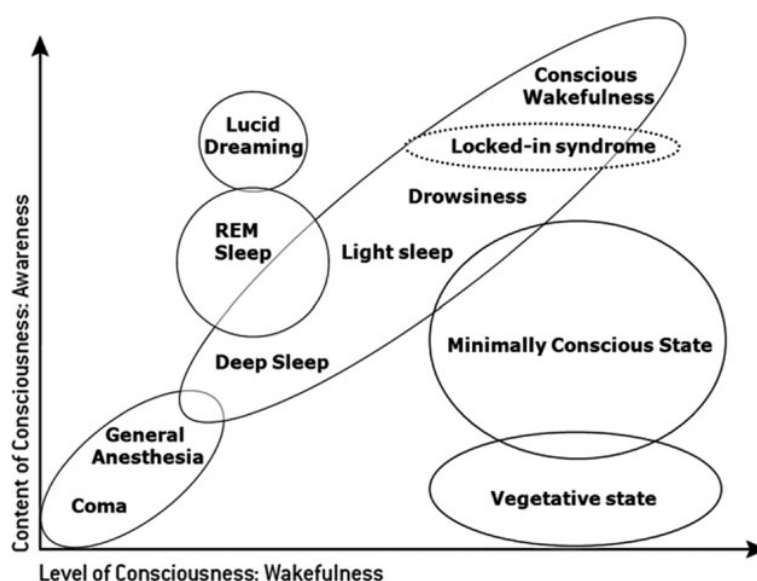


Figura 1.1: (Da [12]). Illustrazione delle due principali componenti della coscienza in stati fisiologici normali e patologici. Nel primo caso le due componenti sono in genere positivamente correlate.

I disordini di coscienza (DOC, dall'inglese Disorders Of Consciousness) si manifestano in seguito a danni cerebrali acuti, che possono essere sia di natura traumatica (incidente, caduta, ...) sia di tipo non traumatico (ictus, anossia, ...) e possono comportare la perdita di coscienza e l'entrata in coma. Alcuni pazienti recuperano coscienza nei primi giorni dopo l'evento, mentre nei casi più gravi il paziente muore nel giro di pochi giorni. Altri pazienti invece dal momento in cui aprono gli occhi entrano in uno stato vegetativo. Quando il paziente in stato vegetativo recupera gradualmente la consapevolezza entra in uno stato di minima coscienza. Tale stato è spesso seguito da un periodo di transitoria amnesia post-traumatica in cui il paziente rimane in stato confusionario; nella maggior parte dei casi il paziente recupera in poche settimane, ma in alcuni casi può rimanere in questo stato di minima coscienza per diversi mesi o addirittura anni. Un'altra condizione eccezionale è la "locked-in syndrome" (LIS), in cui il paziente è pienamente cosciente, ma non è in grado di comunicare se non con movimenti oculari (Figura 1.2).

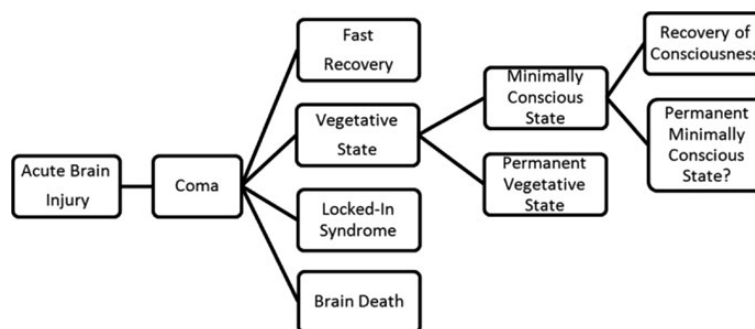


Figura 1.2: (Da [12]). Danni cerebrali acuti possono comportare l'entrata in coma. In genere il coma dura pochi giorni, e una volta che il paziente apre gli occhi entra in uno stato vegetativo. Dopo aver mostrato i primi segni di coscienza il paziente entra in stato di minima coscienza, e alla fine recupera piena coscienza. In rari casi una persona può essere affetta da locked-in syndrome.

Vengono ora presentati in modo più dettagliato i disordini dello stato di coscienza.

### 1.1.1 Coma

Il coma è un profondo stato d'incoscienza che fa seguito a molteplici tipi di danno cerebrale, strutturale o metabolico. Il paziente giace per lo più immobile, ad occhi chiusi, e non può essere risvegliato anche se stimolato in maniera intensa. I pazienti in coma sono caratterizzati dalla mancanza di cicli sonno-veglia, e dall'assenza di espressione e comprensione verbale, sebbene si possano ottenere risposte di tipo riflesso a stimoli dolorosi. In questi pazienti si ha assenza di consapevolezza di sé e dell'ambiente, le funzioni autonome come la termoregolazione corporea e la respirazione sono ridotte ed è pertanto necessaria un'assistenza respiratoria.

### 1.1.2 Stato vegetativo

Lo stato vegetativo è una condizione clinica che può insorgere dopo il coma. Secondo la Multi-Society Task Force (1994) lo stato vegetativo è caratterizzato da:

- intermittente vigilanza che si manifesta con la presenza di cicli sonno-veglia (ad esempio periodi di apertura spontanea degli occhi);
- sufficiente conservazione delle funzioni autonome tale da permettere la sopravvivenza con adeguate cure mediche;
- mancata coscienza di sé e mancata consapevolezza dell'ambiente circostante;
- assenza di qualunque gesto volontario e finalizzato di tipo spontaneo e di risposte motorie, verbali e comportamentali a stimoli di diversa natura (visivi, uditivi, tattili o dolorosi);
- assenza di comprensione o produzione verbale;
- incontinenza urinaria e fecale;
- variabile conservazione dei nervi cranici e dei riflessi spinali.

Il paziente in stato vegetativo è quindi vigile ma non consapevole, il che dimostra che le due componenti di coscienza possono essere completamente separate. Le funzioni tronco-encefaliche di un paziente in stato vegetativo sono conservate, sebbene siano presenti lesioni corticali e talamiche. Il metabolismo del cervello è diminuito del 40-50% rispetto ai valori normali. Il paziente in stato vegetativo è in grado di eseguire una serie di movimenti, quali digrignare i denti, sbattere e muovere gli occhi, deglutire, masticare, sbadigliare, piangere, sorridere, ma questi sono sempre movimenti riflessi e non collegati al contesto. Il comportamento motorio è ridotto a pochi movimenti stereotipati o riflessi ed è inadeguato se comparato all'intensità dello stimolo. Tipicamente i pazienti in stato vegetativo non sono in grado di seguire con gli occhi un oggetto in movimento o la propria immagine allo specchio.

Se il paziente è ancora in stato vegetativo un mese dopo la lesione cerebrale si dice che si trova in uno *stato vegetativo persistente*, mentre lo stato vegetativo è detto *permanente* quando si presume che sia irreversibile. L'uso dei termini persistente o permanente è attualmente sconsigliato, in quanto può dar luogo a diagnosi, prognosi o decisioni terapeutiche errate; è preferibile invece specificare la durata e la causa dello stato vegetativo.

Pur essendo la probabilità di recupero sempre minore con il passare del tempo dall'evento acuto, in base alle conoscenze attuali non è possibile stabilire un limite temporale che denoti l'irreversibilità dello stato vegetativo in quanto sono stati riportati alcuni casi di recupero parziale del contatto con il mondo esterno anche dopo moltissimo tempo.

### 1.1.3 Stato di minima coscienza

Lo stato di minima coscienza è una condizione clinica che può presentarsi dopo un coma o rappresentare l'evoluzione di un precedente stato vegetativo, ed è caratterizzata da una severa compromissione della coscienza, in cui si possono tuttavia individuare comportamenti finalizzati, volontari e riproducibili, a volte mantenuti sufficientemente a lungo e non classificabili come comportamenti riflessi. Anche se tali pazienti non sono in grado di comunicare in modo funzionale, possono a volte rispondere adeguatamente ai comandi verbali e con verbalizzazione comprensibile, e sono in grado di inseguire con lo sguardo oggetti, persone o immagini allo specchio; possono inoltre manifestare comportamenti di tipo emotivo.

Per considerare riproducibili le risposte osservate occorre tener conto sia della loro consistenza che della loro complessità; infatti, possono essere necessarie valutazioni estese e ripetute per determinare se una risposta osservata (movimento delle dita, chiusura degli occhi) si presenta in seguito a una richiesta o un avvenimento specifico, o è una semplice coincidenza. Viceversa per le risposte complesse (ad esempio la verbalizzazione comprensibile) possono essere sufficienti poche osservazioni per determinarne la consistenza.

L'attività metabolica dei pazienti in stato di minima coscienza è ridotta del 20-40% rispetto a un soggetto sano, le funzioni autonome sono conservate e le connessioni talamo-corticali e cortico-corticali sono in parte ristabilite o conservate.

Lo stato di minima coscienza, così come quello vegetativo, può essere transitorio, cronico o permanente.

Una volta che i pazienti sono in grado di comunicare in modo funzionale si dice che sono *emersi* dallo stato di minima coscienza; essi possono quindi utilizzare più oggetti in modo appropriato e il loro sistema di comunicazione

risulta adeguato e consistente. Poiché però quest'ultima condizione clinica, come lo stato di minima coscienza, è di recente introduzione, sono ancora necessarie ulteriori ricerche e convalide dei criteri diagnostici.

#### 1.1.4 Locked-in Syndrome

La locked-in syndrome è una condizione clinica in cui il paziente è vigile e cosciente ma presenta gravi deficit delle funzioni motorie tali da rendere difficoltoso il riscontro degli evidenti segni di coscienza. Il paziente con la forma classica di locked-in syndrome è generalmente tetraplegico e anartrico (impossibilitato a parlare in presenza di capacità linguistiche integre) ma conserva la possibilità di movimenti oculari sul piano verticale, che spesso sono l'unico possibile strumento di comunicazione. Oltre alla forma classica esistono la forma incompleta, in cui il paziente non è del tutto tetraplegico e anartrico, e la forma totale, in cui invece sono impossibili anche i movimenti oculari; in quest'ultimo caso la diagnosi differenziale con una condizione di coma o di stato vegetativo può essere particolarmente difficoltosa.

A conclusione di questa panoramica, va detto che ad oggi è difficile stabilire a livello individuale la prognosi per la sopravvivenza e il recupero dal coma, dallo stato vegetativo, dallo stato di minima coscienza e dalla locked-in syndrome, sebbene alcuni fattori come la giovane età del paziente, un'etiologia traumatica o la breve durata dello stato aumentano le possibilità di recupero.

#### 1.1.5 Diagnosi dei disordini di coscienza

Nella riabilitazione neurologica la distinzione tra stato vegetativo e di minima coscienza è di grande importanza, sia per le implicazioni in termini di prognosi e scelte terapeutiche, sia a livello medico-legale ed etico. Il metodo principale per la rilevazione di segni di coscienza è l'osservazione del comportamento. Tale valutazione è essenziale durante tutto il periodo di ricovero ospedaliero, al fine di ottenere informazioni sui progressi cognitivi e definire le cure appropriate, ed è perciò utilizzata per valutare la consapevolezza del paziente di sé e dell'ambiente.



Tuttavia ad oggi non è ancora disponibile alcuna tecnica per misurare in modo diretto la consapevolezza; infatti è possibile identificare la sua presenza, ma è molto più difficile dimostrarne l'assenza. Inoltre i movimenti volontari possono essere a volte erroneamente interpretati come riflessi, e le risposte motorie potrebbero essere limitate a causa di paralisi degli arti; in questo modo può succedere che un paziente in stato di minima coscienza venga diagnosticato in stato vegetativo, così come che un paziente con locked-in syndrome sia diagnosticato in stato vegetativo. Inoltre, in questi pazienti anche il livello di vigilanza è fluttuante, e pertanto può succedere che alcuni di essi entrino in uno stato di dormiveglia o addirittura si addormentino durante la valutazione.

Tutti questi fattori possono portare ad errori diagnostici. Alcuni studi hanno infatti dimostrato che il 20-40% dei pazienti diagnosticati in stato vegetativo hanno invece mostrato segni di coscienza quando sono stati valutati utilizzando strumenti più affidabili e standardizzati. Pertanto, al fine di evitare una diagnosi errata, è necessario che tali valutazioni siano fatte da personale preparato e con esperienza, usando strumenti standardizzati come scale cliniche e test individuali (per valutare se un comportamento specifico è eseguito in risposta ad un comando, e se il modo in cui viene effettuato cambia nel tempo, spontaneamente o per effetto del trattamento), al fine di rendere il più oggettivo possibile il giudizio clinico.

I più utilizzati strumenti clinici per la valutazione dello stato di coscienza dei pazienti con lesioni cerebrali sono la Glasgow Coma Scale (GCS), la Glasgow Liège Scale (GLS), la Coma Recovery Scale-Revised (CRS-R), la Full Outline of Unresponsiveness (FOUR), la Wessex Head Injury Matrix (WHIM), la scala Coma-Near Coma (CNC), il Western Neuro-Sensory Stimulation Profile (WNSSP) e la Sensory Modality Assessment and Rehabilitation Technique (SMART). Alcune scale sono di aiuto per la diagnosi nelle prime ore di ricovero dei pazienti nel reparto di terapia intensiva (ad esempio, GCS e FOUR), mentre altre come la CRS-R sono solitamente utilizzate durante tutto il ricovero.

Di seguito viene introdotta la CRS-R, la scala clinica che è stata utilizzata nel lavoro di tesi; il lettore interessato ad approfondire gli altri strumenti di valutazione può consultare, ad esempio, [12].

### 1.1.5.1 Coma Recovery Scale-Revised (CRS-R)

La Coma Recovery Scale-Revised (CRS-R, [10]) è uno strumento clinico sviluppato recentemente in modo specifico per supportare efficacemente la diagnosi degli stati vegetativo e di minima coscienza, oltre che per distinguere i pazienti in stato di minima coscienza dai soggetti che hanno recuperato la capacità di comunicare in modo funzionale; infatti, è l'unica scala ad incorporare esplicitamente i criteri diagnostici di stato vegetativo e di minima coscienza.

La CRS-R è costituita da un totale di 23 voci divise in sei sottoscale (uditiva, visiva, motoria, motricità orale e verbalizzazione, comunicazione e vigilanza) e ordinate in base al loro grado di complessità. Relativamente ad ogni sottoscala viene assegnato il punteggio più alto in corrispondenza di una risposta finalizzata e pienamente consapevole, mentre il punteggio minimo, pari a zero, è assegnato in caso di assenza di risposta allo stimolo.

In molti centri di ricerca la CRS-R è lo strumento principale per la valutazione comportamentale dei pazienti gravemente cerebrolesi, pertanto è stata tradotta e validata in diverse lingue.

### 1.1.5.2 Strumenti complementari di diagnosi

I metodi di *brain-imaging* e di *elettrofisiologia* permettono di indagare in modo oggettivo le funzioni cerebrali residue in pazienti affetti da disordini di coscienza e consentono pertanto di misurare l'entità del danno a fini diagnostici, prognostici e terapeutici.

Le tecniche di *Functional Neuroimaging* più comunemente utilizzate sono le seguenti:

- la *risonanza magnetica funzionale* (fMRI), che misura indirettamente, attraverso l'analisi delle diverse proprietà magnetiche dell'emoglobina a seconda che sia legata o no con l'ossigeno, le modificazioni del metabolismo e del flusso sanguigno cerebrale nelle regioni in cui si verifica maggiore attività neurale; infatti, quando sono attive, le cellule nervose consumano l'ossigeno trasportato dall'emoglobina degli eritrociti che attraversano i capillari sanguigni locali.

- la *tomografia ad emissione di positroni* (PET), in grado di misurare i cambiamenti nel metabolismo del cervello utilizzando un tracciante radioattivo (glucosio marcato) iniettato nel circolo del sangue, che si accumula nelle aree attive del cervello.
- l'elettroencefalogramma (EEG), che registra, a livello dello scalpo, l'attività elettrica generata direttamente dall'attività neuronale.

Le tecniche PET e fMRI, che misurano dunque i cambiamenti nel flusso ematico locale, hanno rispettivamente una buona e ottima risoluzione spaziale ma una scarsa risoluzione temporale; al contrario, l'EEG ha una eccellente risoluzione temporale (dell'ordine dei *ms*) ma una modesta risoluzione spaziale (dell'ordine dei *cm<sup>2</sup>*).

Una panoramica sull'ElettroEncefaloGrafia è fornita nella prossima sezione.

## 1.2 ElettroEncefaloGrafia

L'attività elettrica del cervello negli animali non anestetizzati venne descritta in modo qualitativo da Richard Caton già nella seconda metà del diciannovesimo secolo, ma venne analizzata per la prima volta in modo sistematico nel 1924 dallo psichiatra tedesco Hans Berger, il quale registrò per la prima volta le diverse forme d'onda del segnale elettrico cerebrale applicando degli elettrodi sullo scalpo umano. Berger introdusse il termine *elettroencefalogramma* (EEG) per indicare la registrazione dei potenziali cerebrali mediante elettrodi posti sulla cute e osservò nei suoi studi come tali forme d'onda dipendessero dallo stato mentale del soggetto, gettando così le basi dell'attuale *elettroencefalografia*.

A seguito dei miglioramenti nella strumentazione di acquisizione del segnale, l'EEG si è affermato come metodo di esplorazione non invasiva delle funzioni nervose dell'uomo, ed è stato per molti anni l'unico mezzo diagnostico a disposizione per lo studio delle patologie cerebrali. Ancora oggi rimane uno strumento fondamentale per ottenere informazioni dinamiche in tempo reale dell'attività elettrica generata durante l'attività cerebrale, fornendo dunque un importante contributo per l'inquadramento diagnostico, la va-

lutazione della gravità del quadro clinico e la formulazione di un giudizio prognostico.

L'EEG è una misura del flusso di corrente extracellulare che viene generato dalla somma delle attività elettriche sincrone di un numero molto elevato di neuroni fra loro interconnessi. Il contributo apportato da ciascun neurone corticale è estremamente basso ed è pertanto necessaria l'attivazione contemporanea di un'ampia popolazione neuronale per produrre un segnale che sia effettivamente rilevabile dallo scalpo; l'ampiezza del segnale EEG dipende quindi dal numero e dalla localizzazione dei neuroni, e dal grado di sincronia dell'attività neuronale generata.

### 1.2.1 Metodi di registrazione

Negli EEG convenzionali la registrazione è ottenuta applicando in posizioni standard sullo scalpo degli elettrodi di metallo (Ag/AgCl). La superficie della cute deve essere adeguatamente preparata prima di applicare gli elettrodi, sgrassando la zona di registrazione con alcool e quindi applicando un gel conduttore. In alternativa viene fatta indossare al soggetto una cuffia già predisposta che contiene tutti gli elettrodi.

Il sistema più utilizzato per posizionare gli elettrodi segue le direttive specificate dal Sistema Internazionale 10-20, proposto da Jasper nel 1958 (Figura 1.3); questo sistema utilizza alcuni punti di riferimento anatomici per standardizzare il posizionamento degli elettrodi sullo scalpo e assicura pertanto una denominazione coerente degli elettrodi tra i vari studi. 10% oppure 20% si riferiscono alla percentuale della distanza tra i due punti di repere cranici inion (prominenza alla base dell'osso occipitale) e nasion (attaccatura superiore del naso) e indicano le distanze proporzionali standard per identificare i punti in cui applicare gli elettrodi. Nella maggior parte dei casi vengono applicati 19 elettrodi lungo cinque linee: longitudinale esterna e longitudinale interna di destra, centrale, longitudinale esterna e longitudinale interna di sinistra.

Nel montaggio bipolare ciascun canale misura la differenza tra l'attività elettrica di due elettrodi adiacenti posti sullo scalpo, mentre in quello referenziale è misurata la differenza tra ciascun elettrodo attivo collocato sullo

scalpo e uno di riferimento posto solitamente sul lobo dell'orecchio. Nel montaggio a referenza media invece ogni canale registra la differenza tra l'attività elettrica di un elettrodo e un segnale di riferimento dato dalla media di quella registrata da tutti gli elettrodi.

Le differenze di potenziale tra elettrodi vengono filtrate, amplificate e quindi memorizzate in formato digitale dopo una conversione analogico-digitale del segnale. In questo modo i tracciati, archiviati permanentemente su supporti magnetici o ottici, sono sempre disponibili per eventuali elaborazioni successive.

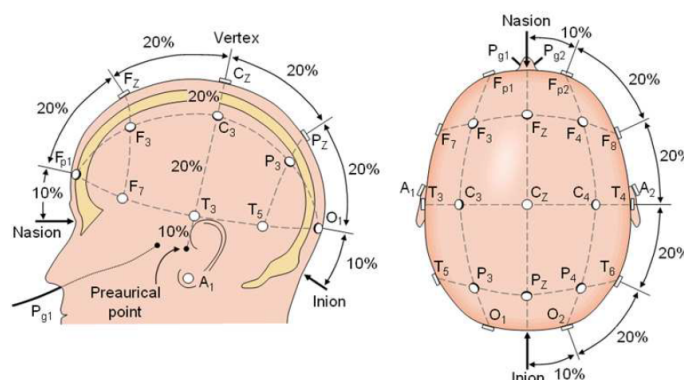


Figura 1.3: Sistema Internazionale 10-20. A ogni elettrodo è associata una sigla che ne identifica in modo univoco la posizione sullo scalpo. Le lettere identificano la regione della corteccia esplorata (Fp: frontopolare; F: frontale; C: centrale; P: parietale; T: temporale; O: occipitale), mentre i numeri (o la lettera z) indicano l'emisfero (numeri dispari: sinistra; numeri pari: destra; z: linea mediana).

Il segnale elettroencefalografico ha natura oscillatoria. In condizioni fisiologiche, nell'uomo l'ampiezza dei potenziali registrati sullo scalpo è compresa tra 10 e 100  $\mu\text{V}$  con frequenze comprese tra 0.5 e 80 Hz a seconda del grado di attività della corteccia cerebrale. In particolare si possono riconoscere cinque attività ritmiche caratteristiche (Figura 1.4), spesso collegate a particolari stati comportamentali:

- **ritmo delta:** comprende le frequenze inferiori a 4 Hz. È predominante nei neonati e viene normalmente riscontrato nei pazienti adulti durante gli stadi profondi del sonno o in condizioni patologiche quali il coma o

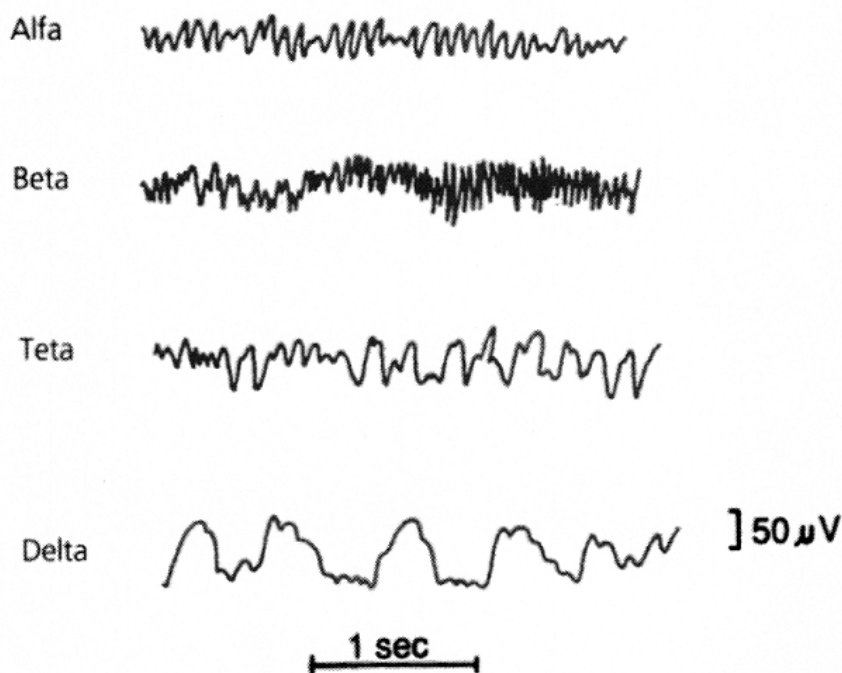


Figura 1.4: Diversi tipi di onde elettroencefalografiche normali.

lesioni cerebrali diffuse. Quando è presente nell'EEG di soggetti svegli, questa attività è considerata anormale.

- **ritmo theta:** è caratterizzato da oscillazioni di frequenza compresa tra 4 e 8 Hz. È predominante nelle regioni temporale e parietale nei bambini, ma può essere registrato in alcuni adulti durante periodi di disappunto e frustrazione. Si osserva nei due terzi dei giovani adulti durante il sonno.
- **ritmo alfa:** ha una frequenza compresa tra 8 e 13 Hz, un'ampiezza normalmente inferiore a  $50 \mu\text{V}$  ed è presente negli EEG della maggior parte delle persone normali sveglie, ad occhi chiusi e in uno stato di riposo mentale. È predominante nella regione occipitale ma talvolta può anche essere registrato nelle regioni frontale e parietale dello scalpo; tende invece a scomparire all'apertura degli occhi (Figura 1.5) o durante l'attività mentale.
- **ritmo beta:** è caratterizzato normalmente da oscillazioni di frequen-

za compresa tra 13 e 30 Hz ma che talvolta, particolarmente durante un'attività mentale intensa, possono arrivare fino a 50 Hz. Solitamente ha ampiezza inferiore a  $30 \mu\text{V}$ , sebbene l'ampiezza di tale ritmo varia sensibilmente da soggetto a soggetto. È prevalente nelle regioni frontale e parietale dello scalpo, è associato a livelli di coscienza come l'attenzione o la concentrazione, ed è il ritmo dominante negli individui in stato di allerta, ansiosi o che hanno gli occhi aperti.

- **ritmo gamma:** banda  $> 30$  Hz. L'attività gamma è un importante meccanismo alla base dei processi cognitivi e di attività mentali complesse (per esempio attenzione selettiva, riconoscimento dei volti, apprendimento per associazione, ...).



Figura 1.5: All'apertura degli occhi il ritmo alfa viene sostituito da onde asincrone di basso voltaggio.

### 1.2.2 La struttura del sonno

Si distinguono due differenti pattern di sonno: il *sonno lento* (sonno sincrono, o sonno non-REM) e il *sonno rapido* (sonno desincronizzato, o sonno REM, cioè caratterizzato dall'occorrenza di movimenti oculari rapidi). Sulla base di criteri internazionali il sonno viene ulteriormente differenziato in vari stadi e in cicli.

Nel sonno sincrono si identificano quattro stadi, ben differenziabili in base alle caratteristiche del segnale EEG, che costituiscono il progressivo approfondimento del sonno:

- **stadio I - addormentamento:** costituisce il passaggio dalla veglia al sonno; il ritmo alfa tende a diffondere anteriormente, poi si discontinua e il segnale diviene dominato diffusamente da componenti theta irregolari di bassa ampiezza; in questa fase è presente un incremento

di attività in banda beta, in particolare in sede fronto-centrale. Compaiono quindi figure di onde al vertice, con massima ampiezza in sede centrale bilaterale e in sede di vertice.

- **stadio II - sonno leggero:** il segnale si arricchisce di componenti lente in banda theta e delta, con un incremento anche di ampiezza rispetto allo stadio precedente. Compaiono le figure di fuso (spindles), treni di onde con frequenza generalmente comprese tra i 12-16 Hz di breve durata, e i complessi K.
- **stadi III e IV - sonno profondo:** il segnale EEG è caratterizzato da componenti lente in banda delta che costituiscono dal 20 al 50% del segnale nello stadio III e superano il 50% nello stadio IV. Vi è una progressiva riduzione di incidenza delle figure di fuso (assenti o quasi nello stadio IV).

Nel *sonno REM* il segnale EEG cambia in modo piuttosto repentino e si desincronizza, divenendo dominato da componenti irregolari, in banda theta, di bassa e media ampiezza, simili a quelle che caratterizzano lo stadio di addormentamento. Per un corretto riconoscimento di tale fase sono essenziali i parametri poligrafici: l'Elettro-oculogramma registra movimenti oculari rapidi, mentre l'Elettromiogramma indica la scomparsa dell'attività tonica muscolare. In genere, le persone che si risvegliano in fase REM riferiscono che stavano sognando, a differenza di quanto succede al risveglio da una fase non-REM.

In alcune patologie, in particolare in molte forme di epilessie, vi è un incremento dell'incidenza delle anomalie EEG (o addirittura la loro comparsa) in addormentamento e in sonno sincrono leggero; inoltre il passaggio dalla veglia al sonno e il risveglio sono momenti di estrema importanza nella definizione di sindromi epilettiche età-dipendenti.



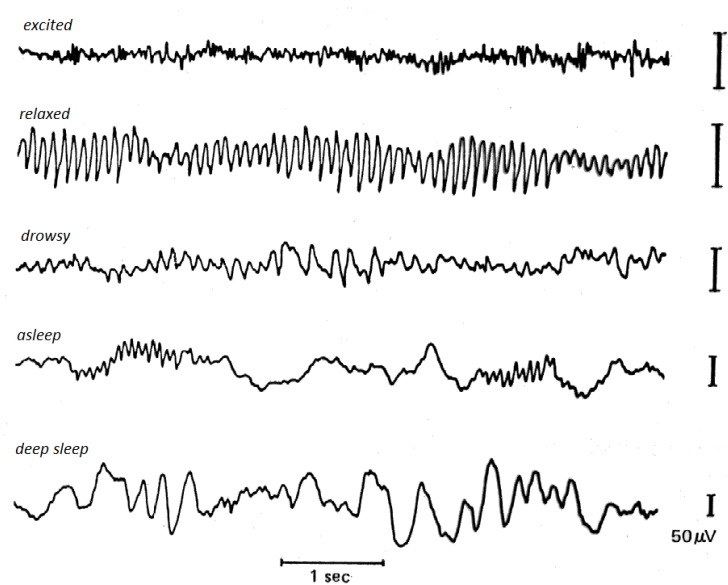


Figura 1.6: Variazioni del segnale EEG in veglia e in sonno.

# Capitolo 2

## Metodi

In questo capitolo vengono presentate brevemente le tecniche utilizzate nello studio del problema clinico in esame, focalizzando in particolar modo l'attenzione sui principali risultati e sulle proprietà di nostro interesse.

Nella Sezione 2.1 viene presentata la *trasformata discreta di Fourier* (DFT), strumento di grande utilità per trattare le successioni periodiche e fra i più importanti nell'analisi dei segnali digitali. Per mezzo della DFT è infatti possibile calcolare lo spettro di frequenza del segnale, da cui è possibile determinare il contributo relativo che le varie frequenze danno all'onda nella finestra di tempo analizzata. Tuttavia l'analisi di Fourier presenta un limite: passando dal dominio del tempo a quello delle frequenze si perde l'informazione temporale e pertanto gli spettri non ci danno alcuna informazione su come evolvono nel tempo le componenti di frequenza.

La Sezione 2.2 è dedicata alle *wavelets*, funzioni matematiche di grande interesse nella teoria dei segnali grazie alle loro proprietà di localizzazione sia in tempo che in frequenza. La trasformata wavelet è infatti caratterizzata dalla presenza di due indici, di cui uno percorre le frequenze mentre l'altro indica la localizzazione temporale.

Nella Sezione 2.3 si presenta l'*Analisi delle Componenti Principali* (PCA), tecnica statistica per l'analisi di dati multivariati che permette di descrivere la struttura di varianza-covarianza di un insieme di variabili aleatorie in termini di nuove variabili tra loro incorrelate ognuna delle quali è una combinazione lineare di quelle iniziali. L'obiettivo di questa tecnica è quello di

effettuare una riduzione dimensionale del problema, determinando a partire dalle  $p$  componenti originarie necessarie per spiegare l'intera variabilità del fenomeno un insieme di  $k \ll p$  vettori linearmente indipendenti, da usare come nuova base dello spazio, che spieghi la maggior parte della variabilità presente nei dati e su cui proiettare le osservazioni.

Infine, la Sezione 2.4 è dedicata alla *Cluster Analysis*. L'idea alla base di questa tecnica è che unità statistiche che appartengono allo stesso gruppo sono più "simili" tra loro rispetto ad unità che appartengono a gruppi differenti. Durante l'analisi non vengono fatte assunzioni né sul numero effettivo dei gruppi né sulla loro struttura; inoltre la vicinanza tra le unità è descritta in termini della misura di dissimilarità adottata, la cui scelta pertanto risulta cruciale.

Il lettore interessato ad approfondire gli argomenti qui trattati può consultare, ad esempio, [23],[26], [1] e [3] per la trasformata discreta di Fourier, [22], [1] e [4] per le wavelets, [15] e [16] per l'analisi delle componenti principali, [8], [15] e [13] per la parte di cluster analysis.

## 2.1 La trasformata discreta di Fourier

Sia  $L^2([0, 2\pi])$  lo spazio delle funzioni a valori reali di quadrato sommabile nell'intervallo  $[0, 2\pi]$ . I vettori:

$$\varphi_0(t) = \frac{1}{\sqrt{2\pi}}, \quad \varphi_{2n}(t) = \frac{\cos nt}{\sqrt{\pi}}, \quad \varphi_{2n-1}(t) = \frac{\sin nt}{\sqrt{\pi}}, \quad n = 1, 2, \dots$$

formano una base ortonormale di  $L^2([0, 2\pi])$ .

Pertanto,  $f \in L^2([0, 2\pi])$  si può riscrivere come:

$$f(t) = \frac{a_0}{\sqrt{2\pi}} + \sum_{n=1}^{\infty} \frac{a_n}{\sqrt{\pi}} \cos nt + \sum_{n=1}^{\infty} \frac{b_n}{\sqrt{\pi}} \sin nt = \sum_{n=0}^{\infty} c_n \varphi_n(t), \quad (2.1)$$

dove l'uguaglianza è intesa in  $L^2([0, 2\pi])$ , con

$$a_0 = \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} f(t) dt, \quad a_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(t) \cos nt dt,$$

$$b_n = \frac{1}{\sqrt{\pi}} \int_0^{2\pi} f(t) \sin nt dt, \quad c_n = \langle f, \varphi_n \rangle = \int_0^{2\pi} f(t) \varphi_n(t) dt.$$

Questa rappresentazione è detta sviluppo in *serie di Fourier* della  $f$ , e le quantità  $a_0, a_n, b_n$  sono i *coefficienti di Fourier* della  $f$ .

Ricordando che  $\lambda_n(t) := e^{int} = \cos nt + i \sin nt$ , l'Equazione 2.1 può essere riscritta nel seguente modo:

$$f(t) = \sum_{n=-\infty}^{\infty} \gamma_n e^{int}, \text{ con } \gamma_n = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-int} dt = \frac{1}{2\pi} \langle f, \lambda_n \rangle.$$

Supponiamo ora che  $N$  sia un numero intero positivo pari (in realtà la maggior convenienza dal punto di vista dei calcoli numerici si ha quando  $N$  è una potenza di 2). Definiamo *troncata di ordine  $N$*  della serie di Fourier la funzione

$$f_N^*(t) = \sum_{k=-\frac{N}{2}}^{\frac{N}{2}-1} \gamma_k e^{ikt} = \sum_{k=0}^{N-1} \tilde{\gamma}_k e^{i(k-\frac{N}{2})t},$$

dove

$$\tilde{\gamma}_k = \frac{1}{2\pi} \int_0^{2\pi} f(t) e^{-i(k-\frac{N}{2})t} dt = \frac{1}{2\pi} \langle f, \tilde{\lambda}_k \rangle, \quad k = 0, \dots, N-1 \quad (2.2)$$

e  $\tilde{\lambda}_k = e^{i(k-\frac{N}{2})t}$ . Detto  $S_N = \text{span} \{ \tilde{\lambda}_k, 0 \leq k \leq N-1 \}$ , se  $f \in L^2([0, 2\pi])$  la sua troncata di ordine  $N$  soddisfa la seguente proprietà ottimale di approssimazione nel senso dei minimi quadrati:

$$\| f - f_N^* \|_{L^2([0, 2\pi])} = \min_{g \in S_N} \| f - g \|_{L^2([0, 2\pi])}.$$

Posto  $h = 2\pi/N$  e  $t_j = jh$  per  $j = 0, \dots, N-1$  introduciamo il seguente *prodotto scalare discreto*:

$$\langle f, g \rangle_N = h \sum_{j=0}^{N-1} f(t_j) \overline{g(t_j)}. \quad (2.3)$$

Sostituendo nella 2.2  $\langle f, \tilde{\lambda}_k \rangle$  con  $\langle f, \tilde{\lambda}_k \rangle_N$ , otteniamo i *coefficienti discreti di Fourier* della funzione  $f$ :

$$\tilde{\gamma}_k = \frac{1}{N} \sum_{j=0}^{N-1} f(t_j) e^{-ikjh} e^{ij\pi} = \frac{1}{N} \sum_{j=0}^{N-1} f(t_j) W_N^{(k-\frac{N}{2})j}, \quad k = 0, \dots, N-1 \quad (2.4)$$

dove  $W_N = e^{-i\frac{2\pi}{N}}$  è la *radice principale di ordine  $N$*  dell'unità. Il polinomio trigonometrico

$$\Pi_N^F f(t) = \sum_{k=0}^{N-1} \tilde{\gamma}_k e^{i(k-\frac{N}{2})t} \quad (2.5)$$

è detto *serie discreta di Fourier di ordine  $N$*  di  $f$ .

Grazie alla seguente relazione:

$$\langle \lambda_j, \lambda_l \rangle_N = h \sum_{k=0}^{N-1} e^{-ik(l-j)h} = 2\pi\delta_{jl}, \quad 0 \leq l, j \leq N-1, \quad (2.6)$$

dove  $\delta_{jl}$  è il simbolo di Kronecker, il polinomio trigonometrico  $\Pi_N^F f$  risulta essere l'interpolato di Fourier di  $f$  nei nodi  $t_j$ , ovvero  $\Pi_N^F f(t_j) = f(t_j)$ ,  $j = 0, \dots, N-1$ . Infatti, usando le (2.4) e (2.6) nella (2.5), si trova

$$\Pi_N^F f(t_j) = \sum_{k=0}^{N-1} \tilde{\gamma}_k e^{ikjh} e^{-ijh\frac{N}{2}} = \sum_{l=0}^{N-1} f(t_l) \left[ \frac{1}{N} \sum_{k=0}^{N-1} e^{-ik(l-j)h} \right] = f(t_j) \quad (2.7)$$

da cui come conseguenza discende che

$$f(t_j) = \sum_{k=0}^{N-1} \tilde{\gamma}_k e^{ik(j-\frac{N}{2})h} = \sum_{k=0}^{N-1} \tilde{\gamma}_k W_N^{-(j-\frac{N}{2})k}, \quad j = 0, \dots, N-1. \quad (2.8)$$

La trasformazione  $\{f(t_j)\} \rightarrow \{\tilde{\gamma}_k\}$  descritta dalla (2.4) è detta *trasformata discreta di Fourier* (o DFT, dall'inglese *Discrete Fourier Transform*), mentre la trasformazione (2.8) da  $\{\tilde{\gamma}_k\}$  a  $\{f(t_j)\}$  è detta *trasformata inversa* (IDFT).

Il costo computazionale della DFT e della IDFT è dell'ordine di  $N^2$ ; tale numero può essere proibitivo se  $N$  è "grande". L'uso della *trasformata rapida di Fourier* (FFT, da *Fast Fourier Transform*) richiede solo un numero di flops dell'ordine di  $N \log_2 N$ , a patto che  $N$  sia una potenza di 2; tale algoritmo, sviluppato da Cooley e Tukey nel 1965 e per i cui dettagli si rimanda ad esempio a [26], fattorizza  $N$  con un approccio *divide-et-impera* e sfrutta la simmetria tra numeri complessi coniugati.

## 2.2 Le wavelets

Le wavelets sono definite a partire dal concetto di multirisoluzione ortogonale.

**Definizione 2.2.1.** *Sia  $\{V_j\}_{j \in \mathbb{Z}}$  una successione di sottospazi chiusi  $V_j \subseteq L^2(\mathbb{R})$  e  $\varphi \in V_0$ . Una multirisoluzione ortogonale per  $L^2(\mathbb{R})$  è una coppia  $(\{V_j\}_j, \varphi)$  tale che:*

1.  $V_j \subset V_{j+1}$
2.  $\overline{\bigcup_j V_j} = L^2(\mathbb{R})$  e  $\bigcap_{j=-\infty}^{+\infty} V_j = \{0\}$
3.  $x \mapsto f(x) \in V_j \Leftrightarrow x \mapsto f(2x) \in V_{j+1}$
4.  $\{\varphi(l-k)\}_{k \in \mathbb{Z}}$  è una base ortonormale per  $V_0$  e  $\int_{\mathbb{R}} \varphi \neq 0$ , cioè  $\varphi$  non ha media nulla.

La proiezione di  $f \in L^2(\mathbb{R})$  sui sottospazi  $\{V_j\}_j$  fornisce un'approssimazione di  $f$  sempre migliore al crescere dell'indice  $j$ . Inoltre, grazie alla proprietà (3) della Definizione 2.2.1,  $\{2^{j/2}\varphi(2^j x - k)\}_k$  è una base ortonormale per  $V_j$ . La funzione  $\varphi$  è detta *funzione di scaling* o *wavelet padre*.

In molte applicazioni può essere però interessante conoscere l'informazione di dettaglio necessaria per passare da  $V_j$  a  $V_{j+1}$ , a partire da un sottospazio  $V_0$ , piuttosto che considerare tutti i sottospazi  $V_j$ . Per questo motivo si introduce la successione  $W_j = V_{j+1} \setminus V_j$  dei complementi di  $V_j$  in  $V_{j+1}$ . Una *wavelet madre* è una funzione  $\psi \in W_0$  tale che  $\{\psi(x-m)\}_{m \in \mathbb{Z}}$  sia una base ortonormale per  $W_0$ . Grazie alla proprietà (2) della Definizione 2.2.1 segue che:

$$L^2(\mathbb{R}) = V_{j_0} \oplus \bigoplus_{j=j_0}^{+\infty} W_j = \bigoplus_{j \in \mathbb{Z}} W_j.$$

Inoltre la famiglia  $\{\psi_{j,k}\}_{j,k \in \mathbb{Z}}$  definita come  $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$  è una base ortonormale di  $L^2(\mathbb{R})$ . Pertanto per ogni  $f \in L^2(\mathbb{R})$  si ha che:

$$\begin{aligned} f &= \sum_j \sum_k \langle f, \psi_{j,k} \rangle \psi_{j,k} = \\ &= \sum_k \langle f, \varphi_{j_0,k} \rangle \varphi_{j_0,k} + \sum_{j=j_0}^{+\infty} \sum_k \langle f, \psi_{j,k} \rangle \psi_{j,k} = \\ &= \sum_k s_{j_0,k} \varphi_{j_0,k} + \sum_{j=j_0}^{+\infty} \sum_k d_{j,k} \psi_{j,k}. \end{aligned} \quad (2.9)$$

dove  $\langle \cdot, \cdot \rangle$  è il prodotto scalare in  $L^2(\mathbb{R})$ ,  $s_{j_0,k} := \langle f, \varphi_{j_0,k} \rangle$  e  $d_{j,k} := \langle f, \psi_{j,k} \rangle$ . I coefficienti  $\{s_{j_0,k}\}_{k \in \mathbb{Z}}$  e  $\{d_{j,k}\}_{j \in \mathbb{Z} \cap \{j \geq j_0\}, k \in \mathbb{Z}}$  sono chiamati *trasformata wavelet discreta* di  $f$ .

Si può dimostrare che  $\varphi$  e  $\psi$  soddisfano le seguenti equazioni di dilatazione/raffinamento:

$$\varphi(x) = \sum_k \sqrt{2}h_k\varphi(2x - k) \text{ e } \psi(x) = \sum_k \sqrt{2}g_k\psi(2x - k) \quad (2.10)$$

per opportune successioni  $\{h_k\}_k$  e  $\{g_k\}_k$  chiamate rispettivamente *filtro di scaling* e *filtro wavelet*. Queste equazioni sono fondamentali per lo sviluppo e l'implementazione della *Fast Wavelet Transform* (FFT), algoritmo che permette di effettuare la trasformata e l'antitrasformata wavelet discreta in  $O(n)$  operazioni e per i cui dettagli si rimanda a [2].

### 2.2.1 Le wavelets ortogonali di Daubechies

La famiglia di wavelets ortogonali introdotta da Ingrid Daubechies nel 1988 è tra le più usate per effettuare la stima di funzioni regolari, in quanto consente di ottenere wavelets di regolarità arbitraria con il numero desiderato  $p$  di momenti nulli.

**Definizione 2.2.2.** Una funzione  $f \in L^2(\mathbb{R})$  ha i primi  $p$  momenti nulli se

$$\int_{\mathbb{R}} x^k f(x) dx = 0 \quad \forall k \in \{1, \dots, p-1\}$$

Vale infatti il seguente risultato.

**Teorema 2.2.1.** Sia  $\psi$  una wavelet indotta dalla multirisoluzione  $(\{V_j\}_j, \varphi)$ . Le seguenti affermazioni sono equivalenti:

1.  $\psi$  ha i primi  $p$  momenti nulli.
2. la trasformata di Fourier  $\hat{\psi}(\xi)$  e le sue  $p-1$  derivate si annullano in  $\xi = 0$ .
3.  $H(\xi) = \sum_k h_k e^{ik\xi}$  e le sue  $p-1$  derivate si annullano in  $\xi = \pi$ .

Grazie alla condizione (3) è possibile costruire  $H$ , e quindi il filtro  $h_k$  in modo da soddisfare la condizione per il numero di momenti nulli  $p$  desiderato. Si osservi però che tale condizione risulta verificata per varie scelte di  $H$ , e pertanto esistono più famiglie di wavelets di Daubechies.

La regolarità delle funzioni  $\varphi$  e  $\psi$  associate al filtro aumenta all'aumentare di  $p$ . Per  $p = 1$  si ritrova il filtro di scaling di Haar, di particolare rilevanza dal punto di vista storico e per le sue proprietà. Proposta nel 1909 dal matematico ungherese Alfred Haar, la base di Haar è infatti la prima base wavelet ad esser stata introdotta, ed è inoltre l'unica ad avere funzione di scaling reale, simmetrica a supporto compatto. Tuttavia la funzione di scaling e la wavelet di Haar non sono neppure continue, e pertanto tale base non risulta adatta a stimare funzioni regolari.

### 2.2.2 Stimatore wavelet non lineare di un dato funzionale

L'Equazione 2.9 ci dice che ogni funzione  $f \in L^2(\mathbb{R})$  può essere rappresentata per mezzo di una base wavelet. Sfruttando ciò, è possibile ottenere una stima della funzione  $f$  a partire dalle sue osservazioni discrete e affette da errore di misurazione. Consideriamo il modello statistico:

$$y_i = f(x_i) + \epsilon_i, \quad i = 1, \dots, n, \quad n = 2^J, \quad J \in \mathbb{N}, \quad (2.11)$$

dove  $x_i = \frac{i}{n}$  sono punti equispaziati e  $\epsilon_i$  sono errori indipendenti e identicamente distribuiti (i.i.d.) con distribuzione  $N(0, \sigma^2)$ . Una volta effettuata la trasformata wavelet l'errore si trasmetterà anche sui coefficienti, cioè avremo:

$$d_{j,k} = d_{j,k}^0 + \rho_{j,k}$$

dove  $d_{j,k}$  sono i coefficienti empirici,  $d_{j,k}^0$  i coefficienti veri della funzione  $f$  e  $\rho_{j,k} \sim N(0, \sigma_d^2)$  sono i coefficienti corrispondenti alla trasformata dell'errore. Il metodo più semplice per trovare una stima  $\hat{d}_{j,k}$  dei coefficienti wavelet veri  $d_{j,k}^0$ , a partire dai coefficienti empirici  $d_{j,k}$ , è chiamato hard-thresholding: si fissa un valore di soglia  $t$  e si considera che tutti i coefficienti al di sotto di questa soglia siano dovuti solo a rumore, ponendo pertanto

$$\hat{d}_{j,k}^H = d_{j,k} \mathbb{I}_{\{|d_{j,k}| > t\}}.$$

Infatti l'idea alla base degli stimatori non lineari è che soltanto i coefficienti di modulo maggiore sono quelli significativi per la stima.

Il metodo appena descritto ha però il difetto di non tenere conto del fatto che anche nei coefficienti considerati significativi per la stima è presente una



componente di rumore. Uno stimatore più raffinato è pertanto il cosiddetto *soft-thresholding*:

$$\hat{d}_{j,k}^S = \text{sign}(d_{j,k})(|d_{j,k}| - t)_+.$$

Questo stimatore effettua lo *shrinkage* sui coefficienti che superano il valore di soglia prefissato con l'intento di eliminare la componente dovuta a rumore. La corrispondente stima della funzione  $f$  è quindi

$$\hat{f}(x) = \sum_{k=0}^{2^{j_0}-1} s_{j_0,k} \varphi_{j_0,k}(x) + \sum_{j=j_0}^{J-1} \sum_{k=0}^{2^j-1} \hat{d}_{j,k}^S \psi_{j,k}(x) \quad (2.12)$$

dove  $j_0$  è il livello al di sotto del quale la procedura di thresholding non è applicata, e pertanto i coefficienti dei livelli inferiori al  $j_0$ -esimo non subiscono alcuno shrinkage.

Sono state proposte numerose strategie per la scelta della soglia  $t$ , la più popolare delle quali è l'*universal threshold*:

$$t = \sigma_d \sqrt{2 \log n}.$$

Nell'ipotesi che il coefficiente empirico abbia distribuzione normale con media il coefficiente esatto e varianza  $\sigma_d^2$ , tale scelta del valore di soglia è giustificata infatti dal seguente risultato.

**Proposizione 2.2.1.** *Siano  $X_i \sim N(0, 1)$  i.i.d. e  $A_n = \{\max_{i=1, \dots, n} |X_i| \leq \sqrt{2 \log n}\}$ . Allora:*

$$\mathbb{P}(A_n) \rightarrow 1 \text{ per } n \rightarrow +\infty.$$

Dal Teorema 2.2.1 segue infatti che se il numero  $n$  di osservazioni è sufficientemente elevato, allora l'*universal threshold*  $t$  è tale da "contenere" l'errore con alta probabilità.

La stima di  $\sigma_d$  è basata sul fatto che i coefficienti wavelet dell'ultimo livello, il  $(J - 1)$ -esimo, sono essenzialmente rumore. Donoho e Johnstone propongono in [5] di utilizzare uno stimatore robusto dato dalla mediana della deviazione assoluta dalla mediana (MAD):

$$\hat{\sigma}_d = \frac{\text{mediana}(|d_{J-1} - \text{mediana}(d_{J-1})|)}{0.6745} \quad (2.13)$$

### 2.2.3 Estensione al caso di funzioni multidimensionali

Sia  $\mathbf{f} : \mathbb{R} \ni s \mapsto (f_1(s), f_2(s), \dots, f_p(s)) \in \mathbb{R}^p$  la funzione da stimare. Le osservazioni di  $\mathbf{f}$  a nostra disposizione sono pertanto generate dal modello statistico

$$\mathbf{w}_i = \mathbf{f}(s_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n = 2^J, \quad (2.14)$$

dove  $\boldsymbol{\epsilon}_i$  sono errori normali multivariati i.i.d. di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $\sigma^2 \mathbb{I}_p$ .

L'idea di stimare separatamente le funzioni  $f_1, f_2, \dots, f_p$  applicando separatamente a ciascuna di esse la procedura descritta nella Sezione 2.2.2 non è soddisfacente, in quanto ci si aspetta che se una curva  $p$ -dimensionale presenta una caratteristica significativa in un punto dello spazio, allora tale caratteristica si dovrebbe riflettere contemporaneamente su tutte le  $p$  coordinate. Per questo motivo Pigoli e Sangalli hanno proposto in [22] una tecnica di stima che tenga conto della natura vettoriale della funzione. In particolare la procedura di thresholding proposta è tale che la base wavelet utilizzata per la stima di tutte le coordinate  $f_1, f_2, \dots, f_p$  di  $\mathbf{f}$  sia la stessa: una specifica funzione della base wavelet, con una determinata localizzazione in spazio e in frequenza, o è usata per ogni coordinata di  $\mathbf{f}$  se serve a catturare una caratteristica significativa, oppure in caso contrario non è utilizzata per alcuna coordinata di  $\mathbf{f}$ .

A partire dal modello descritto dalla (2.14), ipotizzando di utilizzare una trasformata ortogonale, si ha che

$$\mathbf{d}_{j,k} = \mathbf{d}_{j,k}^0 + \boldsymbol{\rho}_{j,k},$$

dove  $\mathbf{d}_{j,k} \in \mathbb{R}^p$  è il vettore dei coefficienti wavelet empirici,  $\mathbf{d}_{j,k}^0 \in \mathbb{R}^p$  è il vettore dei coefficienti veri della funzione  $p$ -dimensionale  $\mathbf{f}$ , mentre  $\boldsymbol{\rho}_{j,k} \in \mathbb{R}^p$  è la trasformata wavelet dell'errore e avrà distribuzione normale multivariata di media  $\mathbf{0}$  e matrice di varianza-covarianza  $\sigma_d^2 \mathbb{I}_p$ . Per decidere se stimare  $\mathbf{d}_{j,k}^0$  con il vettore nullo oppure no ci si focalizza sulla norma euclidea  $\|\mathbf{d}_{j,k}\|$  dei coefficienti empirici, per capire se è possibile fissare un valore di soglia in modo analogo a quanto visto per il caso monodimensionale. Poichè  $\|\boldsymbol{\rho}_{j,k}/\sigma_d\|^2 \sim \chi^2(p)$ , siamo ora dunque alla ricerca di una soglia che “contenga” con alta probabilità  $n$  realizzazioni da una  $\chi^2(p)$ .

Vale il seguente risultato.

**Proposizione 2.2.2.** *Siano  $Y_i \sim \chi^2(p)$  i.i.d. e sia  $A_n = \{\max_{i=1,\dots,n} Y_i \leq c_p \log n\}$ , dove*

$$c_p = \begin{cases} 2 & \text{se } p = 1 \\ 3 & \text{se } p \geq 2 \end{cases}$$

Allora

$$\mathbb{P}(A_n) \rightarrow 1 \text{ per } n \rightarrow +\infty.$$

Si osservi che per  $p = 1$ , cioè quando  $Y_i$  è il quadrato di una variabile aleatoria di tipo normale standard, ritroviamo il risultato illustrato per il caso monodimensionale, cioè l'universal threshold  $t = \hat{\sigma}_d \sqrt{2 \log n}$  su  $|d_{j,k}|$ . Per quanto riguarda il caso multidimensionale ( $p \geq 2$ ) la soglia sarà invece pari a  $\hat{\sigma}_d^2 (3 \log n)$  su  $\|\mathbf{d}_{j,k}\|^2$  o, equivalentemente,  $t_p = \hat{\sigma}_d \sqrt{3 \log n}$  su  $\|\mathbf{d}_{j,k}\|$ , dove  $\hat{\sigma}_d$  è stimato raggruppando i coefficienti del livello più fine di tutte e  $p$  le direzioni per mezzo dello stimatore MAD definito dalla (2.13).

Il più semplice stimatore per il caso multidimensionale è perciò il seguente *hard-thresholding*:

$$\hat{\mathbf{d}}_{j,k} = \begin{cases} \mathbf{0} & \text{se } \|\mathbf{d}_{j,k}\| \leq t_p \\ \mathbf{d}_{j,k} & \text{se } \|\mathbf{d}_{j,k}\| > t_p \end{cases}$$

Questo stimatore presenta però le stesse problematiche evidenziate nel caso monodimensionale, cioè non tiene conto del fatto che anche i coefficienti che superano la soglia contengono una componente di errore. Uno stimatore *soft-thresholding* che effettui uno shrinkage sul vettore dei coefficienti agendo in modo analogo a quanto visto per il caso monodimensionale è pertanto più appropriato, sebbene l'estensione al caso  $p$ -dimensionale non è banale.

Partendo dall'ipotesi che gli errori nelle  $p$  dimensioni siano indipendenti e ugualmente distribuiti, essi contribuiscono nello stesso modo alla norma del vettore dei coefficienti empirici. Pertanto Pigoli e Sangalli propongono di stimare  $\mathbf{d}_{j,k}^0$  conservando la direzione del vettore dei coefficienti empirici  $\mathbf{d}_{j,k}$  e diminuendone la norma di  $t_p$  (Figura 2.1). Si ottiene perciò il seguente stimatore *soft-thresholding*:

$$\hat{\mathbf{d}}_{j,k} = \left(1 - \frac{t_p}{\|\mathbf{d}_{j,k}\|}\right)_+ \mathbf{d}_{j,k} = \begin{cases} \mathbf{0} & \text{se } \|\mathbf{d}_{j,k}\| \leq t_p \\ \left(1 - \frac{t_p}{\|\mathbf{d}_{j,k}\|}\right) \mathbf{d}_{j,k} & \text{se } \|\mathbf{d}_{j,k}\| > t_p \end{cases} \quad (2.15)$$

In particolare, lo stimatore (2.15) soddisfa la proprietà di usare la stessa base wavelets per le  $p$  componenti.

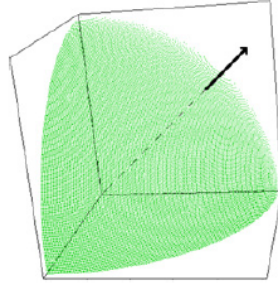


Figura 2.1: (Da [22]). Esempio in tre dimensioni della procedura di soft-thresholding descritta dalla (2.15): lo shrinkage agisce solo sulla norma del vettore  $\mathbf{d}_{j,k}$ , preservandone la direzione.

### 2.2.3.1 Errori correlati nelle $p$ dimensioni

Il metodo (2.15) per la stima dei coefficienti wavelet nel caso multidimensionale assume che le  $p$  componenti di errore nelle  $p$  dimensioni siano incorrelate, cioè  $\text{Var}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbb{I}_p$ . Tuttavia in molte applicazioni, come ad esempio nello studio dei segnali biomedici, risulta più opportuno tenere conto della correlazione fra le componenti di errore nelle diverse direzioni; in questi casi infatti le coordinate del segnale rappresentano differenti proiezioni della stessa dinamica fisica affetta, in fase di registrazione, da errori di misurazione o altre fonti di rumore.

Il modello da considerare ora pertanto è il seguente:

$$\mathbf{w}_i = \mathbf{f}(s_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n = 2^J,$$

dove  $\boldsymbol{\epsilon}_i$  sono errori normali multivariati i.i.d. di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $S$  generica. Ipotizzando di utilizzare una trasformazione wavelet ortogonale, anche ogni vettore  $\mathbf{d}_{j,k}$  dei coefficienti wavelet avrà

distribuzione normale multivariata di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $S$ . Quindi  $\mathbf{d}'_{j,k} S^{-1} \mathbf{d}_{j,k} \sim \chi^2(p)$ , e pertanto, in accordo alla Proposizione 2.2.2, la soglia  $3 \log n$  è appropriata per questa quantità.

Anche in questo caso la matrice  $S$  può essere stimata a partire dai coefficienti del livello  $J-1$ , ad esempio con la matrice di covarianza campionaria dei coefficienti  $\mathbf{d}_{J-1,k}$  ([22]).

## 2.3 Analisi delle Componenti Principali (PCA)

Sia  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$  un vettore aleatorio con matrice di covarianza  $\Sigma$  e autovalori  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Consideriamo le combinazioni lineari:

$$\begin{aligned} Y_1 &= \mathbf{a}'_1 \mathbf{X} = a_{11}X_1 + a_{12}X_2 + \dots + a_{1p}X_p \\ Y_2 &= \mathbf{a}'_2 \mathbf{X} = a_{21}X_1 + a_{22}X_2 + \dots + a_{2p}X_p \\ &\vdots \\ Y_p &= \mathbf{a}'_p \mathbf{X} = a_{p1}X_1 + a_{p2}X_2 + \dots + a_{pp}X_p \end{aligned}$$

Le prima componente principale è la trasformazione lineare  $\mathbf{a}'_1 \mathbf{X}$  che massimizza  $\text{Var}(\mathbf{a}'_1 \mathbf{X})$  sotto il vincolo  $\mathbf{a}'_1 \mathbf{a}_1 = 1$ , cioè è tale che soddisfi:

$$\begin{aligned} \max_{\mathbf{a}_1} \text{Var}(\mathbf{a}'_1 \mathbf{X}) &= \mathbf{a}'_1 \Sigma \mathbf{a}_1 \\ \text{s.t. } \mathbf{a}'_1 \mathbf{a}_1 &= 1 \end{aligned} \tag{2.16}$$

Osserviamo che il vincolo di norma unitaria per il vettore  $\mathbf{a}_1$  è essenziale per la risoluzione del Problema (2.16); appare infatti evidente che in assenza di questo vincolo il massimo non può essere raggiunto per un valore finito di  $\mathbf{a}_1$ .

La seconda componente principale è la trasformazione lineare  $\mathbf{a}'_2 \mathbf{X}$  che massimizza  $\text{Var}(\mathbf{a}'_2 \mathbf{X})$  sotto i vincoli  $\text{Cov}(\mathbf{a}'_1 \mathbf{X}, \mathbf{a}'_2 \mathbf{X}) = \mathbf{a}'_1 \Sigma \mathbf{a}_2 = 0$  e  $\mathbf{a}'_2 \mathbf{a}_2 = 1$ , e così in modo analogo fino alla definizione di tutte le  $p$  componenti principali. Precisamente, l' $i$ -esima componente principale è la trasformazione lineare  $\mathbf{a}'_i \mathbf{X}$  che massimizza  $\text{Var}(\mathbf{a}'_i \mathbf{X})$  sotto i vincoli  $\mathbf{a}'_i \mathbf{a}_i = 1$  e  $\text{Cov}(\mathbf{a}'_i \mathbf{X}, \mathbf{a}'_k \mathbf{X}) = \mathbf{a}'_i \Sigma \mathbf{a}_k = 0$  per  $k < i$ .

Le variabili  $\mathbf{a}'_i \mathbf{X}$ ,  $i = 1, \dots, p$ , sono dette *componenti principali* (o *scores*), mentre i coefficienti dei vettori  $\mathbf{a}_i$  sono detti *loadings*. Nel caso in cui la matrice  $\Sigma$  è incognita, si procede allo stesso modo utilizzando la matrice di covarianza campionaria  $\mathbf{S}$  calcolata a partire da  $\mathbf{x}_1, \dots, \mathbf{x}_p$  realizzazioni del vettore aleatorio  $\mathbf{X}$ .

Valgono i seguenti risultati.

**Proposizione 2.3.1.** *Sia  $\Sigma$  la matrice di covarianza associata al vettore aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ . Siano  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  le coppie autovalore-autovettore della matrice  $\Sigma$  con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Allora l' $i$ -esima componente principale è data da:*

$$Y_i = \mathbf{e}'_i \mathbf{X} = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p, \quad i = 1, \dots, p. \quad (2.17)$$

Inoltre

$$\begin{aligned} \text{Var}(Y_i) &= \mathbf{e}'_i \Sigma \mathbf{e}_i = \lambda_i \quad i = 1, \dots, p \\ \text{Cov}(Y_i, Y_k) &= \mathbf{e}'_i \Sigma \mathbf{e}_k = 0 \quad i \neq k \end{aligned}$$

**Proposizione 2.3.2.** *Sia  $\Sigma$  la matrice di covarianza associata al vettore aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ . Siano  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  le coppie autovalore-autovettore della matrice  $\Sigma$  con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Siano  $Y_1 = \mathbf{e}'_1 \mathbf{X}, \dots, Y_p = \mathbf{e}'_p \mathbf{X}$  le componenti principali. Allora*

$$\sigma_{11} + \sigma_{22} + \dots + \sigma_{pp} = \sum_{i=1}^p \text{Var}(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p = \sum_{i=1}^p \text{Var}(Y_i) \quad (2.18)$$

**Proposizione 2.3.3.** *Sia  $\Sigma$  la matrice di covarianza associata al vettore aleatorio  $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ . Siano  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  le coppie autovalore-autovettore della matrice  $\Sigma$  con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ . Siano  $Y_1 = \mathbf{e}'_1 \mathbf{X}, \dots, Y_p = \mathbf{e}'_p \mathbf{X}$  le componenti principali. Allora*

$$\rho_{Y_i, X_k} = \frac{e_{ik} \sqrt{\lambda_i}}{\sqrt{\sigma_{kk}}}, \quad i, k = 1, \dots, p$$

sono i coefficienti di correlazione tra le componenti  $Y_i$  e le variabili  $X_k$ .

Pertanto dalla Proposizione (2.3.1) segue che le componenti principali sono fra loro incorrelate e hanno varianza pari agli autovalori della matrice  $\Sigma$ .

Dalla Proposizione (2.3.2) segue che la quota di variabilità totale spiegata dall'  $i$ -esima componente principale è pari a

$$\frac{\lambda_i}{\sum_{i=1}^p \lambda_i}, \quad i = 1, 2, \dots, p.$$

Tipicamente il numero  $k$  di componenti principali da utilizzare per non avere eccessiva perdita di informazione è scelto in modo tale che

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} \times 100 = 80\% - 90\%.$$

Per la Proposizione (2.3.3) si ha inoltre che il loading  $e_{ik}$ , il cui valore è indice dell'importanza della  $k$ -esima variabile nell' $i$ -esima componente principale, è proporzionale al coefficiente di correlazione tra  $Y_i$  e  $X_k$ . Il valore dei loadings è solitamente utilizzato per interpretare il significato fisico da attribuire alle componenti principali.

### 2.3.1 Componenti Principali per variabili standardizzate

Le componenti principali possono anche essere ottenute a partire dalle variabili standardizzate:

$$\mathbf{Z} = \begin{bmatrix} Z_1 = \frac{X_1 - \mu_1}{\sqrt{\sigma_{11}}} \\ Z_2 = \frac{X_2 - \mu_2}{\sqrt{\sigma_{22}}} \\ \vdots \\ Z_p = \frac{X_p - \mu_p}{\sqrt{\sigma_{pp}}} \end{bmatrix} = (\mathbf{V}^{1/2})^{-1}(\mathbf{X} - \boldsymbol{\mu}),$$

dove  $\mathbf{V}^{1/2} = \text{diag}(\sqrt{\sigma_{11}}, \dots, \sqrt{\sigma_{pp}})$ . Segue dunque che  $\mathbb{E}(\mathbf{Z}) = \mathbf{0}$  e  $\text{Cov}(\mathbf{Z}) = (\mathbf{V}^{1/2})^{-1} \boldsymbol{\Sigma} (\mathbf{V}^{1/2})^{-1} = \boldsymbol{\varrho}$ . Le componenti principali di  $\mathbf{Z}$  possono quindi essere ottenute a partire dagli autovettori della matrice di correlazione  $\boldsymbol{\varrho}$  di  $\mathbf{X}$ . Tuttavia è importante sottolineare che le coppie autovalore-autovettore derivate dalla matrice  $\boldsymbol{\Sigma}$  in generale non sono le stesse di quelle derivate dalla matrice  $\boldsymbol{\varrho}$ .

Indicando ancora con  $Y_i$  l' $i$ -esima componente principale e con  $(\lambda_i, \mathbf{e}_i)$  l' $i$ -esima coppia autovalore-autovettore ottenute da  $\boldsymbol{\varrho}$ , vale la seguente proposizione.

**Proposizione 2.3.4.** *L' $i$ -esima componente principale delle variabili standardizzate  $\mathbf{Z}' = [Z_1, Z_2, \dots, Z_p]$  con  $\text{Cov}(\mathbf{Z}) = \boldsymbol{\rho}$  è*

$$Y_i = \mathbf{e}_i' \mathbf{Z} = \mathbf{e}_i' (\mathbf{V}^{1/2})^{-1} (\mathbf{X} - \boldsymbol{\mu}), \quad i = 1, \dots, p$$

Inoltre

$$\sum_{i=1}^p \text{Var}(Y_i) = \sum_{i=1}^p \text{Var}(Z_i) = p \quad (2.19)$$

e

$$\rho_{Y_i, Z_k} = e_{ik} \sqrt{\lambda_i} \quad i, k = 1, \dots, p$$

dove  $(\lambda_1, \mathbf{e}_1), (\lambda_2, \mathbf{e}_2), \dots, (\lambda_p, \mathbf{e}_p)$  le coppie autovalore-autovettore della matrice  $\boldsymbol{\rho}$  con  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ .

Dalla (2.19) segue pertanto che la quota di variabilità spiegata dall' $i$ -esima componente principale di  $\mathbf{Z}$  è pari a  $\frac{\lambda_i}{p}$ ,  $i = 1, \dots, p$ . Tipicamente in questo caso le  $k$  componenti principali da considerare sono quelle in corrispondenza delle quali l'autovalore è maggiore di 1.

## 2.4 Cluster Analysis

Sia  $\mathbb{X}$  la matrice dei dati di dimensione  $n \times p$  in cui i vettori riga rappresentano le  $n$  unità statistiche osservate sulle  $p$  variabili.

**Definizione 2.4.1.** *Date le seguenti proprietà:*

1.  $d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^p$
2.  $d(\mathbf{x}, \mathbf{y}) = 0$  se e solo se  $\mathbf{x} = \mathbf{y}$
3.  $d(\mathbf{x}, \mathbf{y}) \leq d(\mathbf{x}, \mathbf{z}) + d(\mathbf{z}, \mathbf{y}) \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^p$
4.  $d(\mathbf{x}, \mathbf{y}) \leq \max(d(\mathbf{x}, \mathbf{z}), d(\mathbf{z}, \mathbf{y})) \quad \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^p$

una funzione  $d : \mathbb{R}^p \times \mathbb{R}^p \rightarrow [0, 1]$  si definisce:

- a) dissimilarità se soddisfa le proprietà 1,2
- b) metrica o distanza se soddisfa le proprietà 1,2,3
- c) ultrametrica se soddisfa le proprietà 1,2,4



Nel caso in cui le  $p$  variabili siano di tipo quantitativo, esempi notevoli di metriche sono i seguenti:

- la *distanza euclidea*:  $d_E(\mathbf{x}, \mathbf{y}) = \left[ \sum_{i=1}^p |x_i - y_i|^2 \right]^{1/2}$ , con  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$
- la *distanza di Minkowski*:  $d_{MI}(\mathbf{x}, \mathbf{y}) = \left[ \sum_{i=1}^p |x_i - y_i|^m \right]^{1/m}$ ,  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$   
Per  $m = 1$   $d_{MI}(\mathbf{x}, \mathbf{y})$  è detta *distanza di Manhattan*, mentre per  $m = 2$  si ritrova la distanza euclidea.
- la *distanza di Mahalanobis*:  $d_{MA}(\mathbf{x}, \mathbf{y}) = \sqrt{(\mathbf{x} - \mathbf{y})' \mathbf{S}^{-1} (\mathbf{x} - \mathbf{y})}$ , con  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^p$  e  $\mathbf{S}$  matrice di covarianza campionaria
- la *distanza di Canberra*:

$$d_C(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^p \frac{|x_i - y_i|}{(x_i + y_i)}$$

con  $\mathbf{x}, \mathbf{y} \in \mathbb{R}_+^p$

Scelta la distanza  $d(\cdot, \cdot)$  da utilizzare è possibile quindi calcolare, a partire dalla matrice  $\mathbb{X}$  dei dati, la *matrice di dissimilarità*  $\mathbf{D} = [d_{i,j}]$  di dimensione  $n \times n$ , in cui  $d_{i,j} = d(\mathbf{x}_i, \mathbf{x}_j)$  e  $\mathbf{x}_i, \mathbf{x}_j$  sono due vettori riga della matrice  $\mathbb{X}$  che rappresentano le unità statistiche  $i$  e  $j$ .

Sebbene in linea teorica dovremmo esaminare tutti i raggruppamenti possibili in modo da poter scegliere la partizione migliore fra tutte quelle costruibili, cioè quella che rende conto al meglio delle distanze tra i vari individui, tale ricerca risulta però irrealizzabile nella pratica. Pertanto in letteratura sono stati proposti algoritmi di clusterizzazione che permettono di trovare raggruppamenti “ragionevoli” senza esaminare tutte le possibili configurazioni, tra cui i *metodi gerarchici*.

### 2.4.1 Algoritmi gerarchici

Gli algoritmi di tipo gerarchico limitano la ricerca del raggruppamento ottimo ad una successione di partizioni ottenute mediante una procedura iterativa di divisione oppure accoppiamento di cluster. Nei *metodi gerarchici divisivi*, qui non trattati e per i quali si rimanda ad esempio a [8], a partire dal

raggruppamento iniziale formato da tutte le unità statistiche si procede per partizioni successive fino a giungere a gruppi formati da una sola unità. Nei *metodi gerarchici agglomerativi* invece inizialmente ogni unità è un cluster e si procede per partizioni annidate successive fondendo ad ogni iterazione in un unico nuovo cluster i due raggruppamenti più “simili” della partizione precedente.

Per entrambe le classi di metodi queste successioni possono essere rappresentate graficamente tramite un *dendrogramma*, grafico dalla struttura ad albero in cui i rami ad ogni livello rappresentano una partizione della successione.

#### 2.4.1.1 Metodi gerarchici agglomerativi

L'algoritmo alla base dei metodi gerarchici agglomerativi è descritto dai seguenti passi:

**Passo 0:** Le  $n$  unità statistiche sono suddivise in  $n$  gruppi costituiti ciascuno da una sola unità. La matrice di dissimilarità è  $D = [d_{i,j}]$ .

**Passo 1:** Si fondono insieme i due gruppi che appartengono alla partizione ottenuta al passo precedente e che risultano più simili in termini di *distanza intercluster*. Si aggiorna quindi la matrice di dissimilarità rimuovendo le righe e le colonne corrispondenti ai cluster che sono stati uniti e rimpiazzandole con una riga e una colonna che tengono conto della distanza tra il nuovo gruppo e i rimanenti cluster.

**Passo 2:** Si ripete il Passo 1 fino a quando si ottiene la partizione finale, cioè quella costituita da un unico cluster comprendente tutte le  $n$  unità statistiche.

Tra le distanze intercluster più note in letteratura ci sono i cosiddetti *metodi di linkage*, definiti nel seguente modo: siano  $U$  e  $V$  due gruppi generici senza unità in comune, costituiti rispettivamente da  $n_U$  e  $n_V$  oggetti, e sia  $x_{U_i}$  l'unità  $i$ -esima del cluster  $U$ :

- nel metodo **single linkage** (nearest neighbor) la distanza tra i cluster  $U$  e  $V$  è pari a

$$d_{sl}(U, V) = \min(d(x_{U_i}, x_{V_j})), \quad i = 1, \dots, n_U, j = 1, \dots, n_V$$

e dunque coincide con la più piccola distanza tra gli elementi dei due gruppi. Questo metodo tende però a non riconoscere come cluster distinti gruppi ben definiti ma collegati da filamenti di dati (*chaining effect*).

- nel metodo **complete linkage** (farthest neighbor) la distanza tra i cluster  $U$  e  $V$  è pari a

$$d_{cl}(U, V) = \max(d(x_{Ui}, x_{Vj})), \quad i = 1, \dots, n_U, j = 1, \dots, n_V$$

e dunque coincide con la massima distanza tra gli elementi dei due gruppi. Questo metodo tende però a formare strutture ellissoidali anche quando non sono presenti.

- nel metodo **average linkage** la distanza tra i cluster  $U$  e  $V$  è pari a

$$d_{al}(U, V) = \frac{1}{n_U n_V} \sum_{i=1}^{n_U} \sum_{j=1}^{n_V} d(x_{Ui}, x_{Vj}), \quad i = 1, \dots, n_U, j = 1, \dots, n_V$$

e dunque coincide con la distanza media tra gli elementi dei due gruppi. Come il complete linkage tende però a formare cluster ellissoidali.

- il metodo **Ward's linkage** si basa sulla scomposizione della devianza totale nella somma delle devianze *within groups* e *between groups*:  $DEV_T = DEV_W + DEV_B$ . Quando si fondono due cluster la devianza within aumenta, mentre quella between diminuisce. Ad ogni passo pertanto il metodo aggrega fra loro i due gruppi per cui vi è il minor incremento della varianza within.

Per misurare bontà del raggruppamento, oltre che per scegliere la distanza  $d(\cdot, \cdot)$  e la distanza intercluster che riflettono nel modo migliore le somiglianze tra le unità, si introducono la *matrice cofenetica*  $\tilde{D} = [\tilde{d}_{i,j}]$  e il *coefficiente di correlazione cofenetica* (detto anche *coefficiente cofenetico*)  $c$ .

La matrice cofenetica  $\tilde{D} = [\tilde{d}_{i,j}]$  è una matrice di dissimilarità in cui ogni elemento  $\tilde{d}_{i,j}$  è pari al livello gerarchico più basso in cui le unità statistiche  $i$  e  $j$  si fondono.

Il coefficiente cofenetico  $c$  misura invece la correlazione tra i valori di dissimilarità e i valori delle distanze cofenetiche ricavabili dal dendrogramma:

$$c = \frac{\sum_{i < j} (d_{i,j} - \bar{d})(\tilde{d}_{i,j} - \bar{\tilde{d}})}{\left[ \sum_{i < j} (d_{i,j} - \bar{d})^2 \sum_{i < j} (\tilde{d}_{i,j} - \bar{\tilde{d}})^2 \right]^{\frac{1}{2}}}, \text{ con } \bar{d} = \frac{\sum_{i < j} d_{i,j}}{\frac{N(N-1)}{2}} \text{ e } \bar{\tilde{d}} = \frac{\sum_{i < j} \tilde{d}_{i,j}}{\frac{N(N-1)}{2}}$$

Secondo Rohlf e Fisher un processo di clusterizzazione è da ritenersi buono quando ha coefficiente cofenetico  $c > 0.8$ .

# Capitolo 3

## Analisi

I dati analizzati riguardano EEG in sonno e in veglia di 19 pazienti appartenenti alle due categorie di disordine di coscienza in esame (9 in stato vegetativo e 10 in stato di minima coscienza), registrati per mezzo di 19 elettrodi applicati sullo scalpo in posizioni opportune (8 per emisfero e 3 centrali, secondo il Sistema Internazionale 10-20) e riferiti ai lobi auricolari (A1+A2). Per ogni paziente si dispone inoltre dell'informazione relativa al punteggio della *Coma Recovery Scale-Revised* (CRS-R). Tali informazioni sono riassunte nella Tabella A.1 in Appendice A.

Tutti i segnali in considerazione hanno durata pari a 64 secondi, frequenza di campionamento 256 Hz, e sono stati prefiltrati con un filtro Notch al fine di rimuovere eventuali disturbi generati dalla rete elettrica. In Europa infatti l'energia elettrica viene distribuita sotto forma di corrente alternata sinusoidale con portante a 50 Hz, la quale può influire sugli strumenti per registrare i dati; pertanto è opportuno rimuovere questa frequenza dal segnale prima di effettuare qualsiasi tipo di analisi.

### 3.1 Smoothing dei dati

Dal punto di vista matematico l'EEG è una funzione vettoriale

$$\mathbf{f} : \mathbb{R} \ni s \mapsto (f_1(s), f_2(s), \dots, f_p(s)) \in \mathbb{R}^p, \text{ con } p = 19.$$

Possiamo quindi interpretare le osservazioni a nostra disposizione come realizzazioni di una variabile aleatoria multivariata nell'intervallo  $(s_{min}, s_{max}) = (0s, 64s)$  in punti successivi equispaziati  $s_i$ ,  $i = 1, \dots, n = 2^J$ ,  $J = 14$ , generate con frequenza di campionamento pari a 256 Hz dal modello statistico

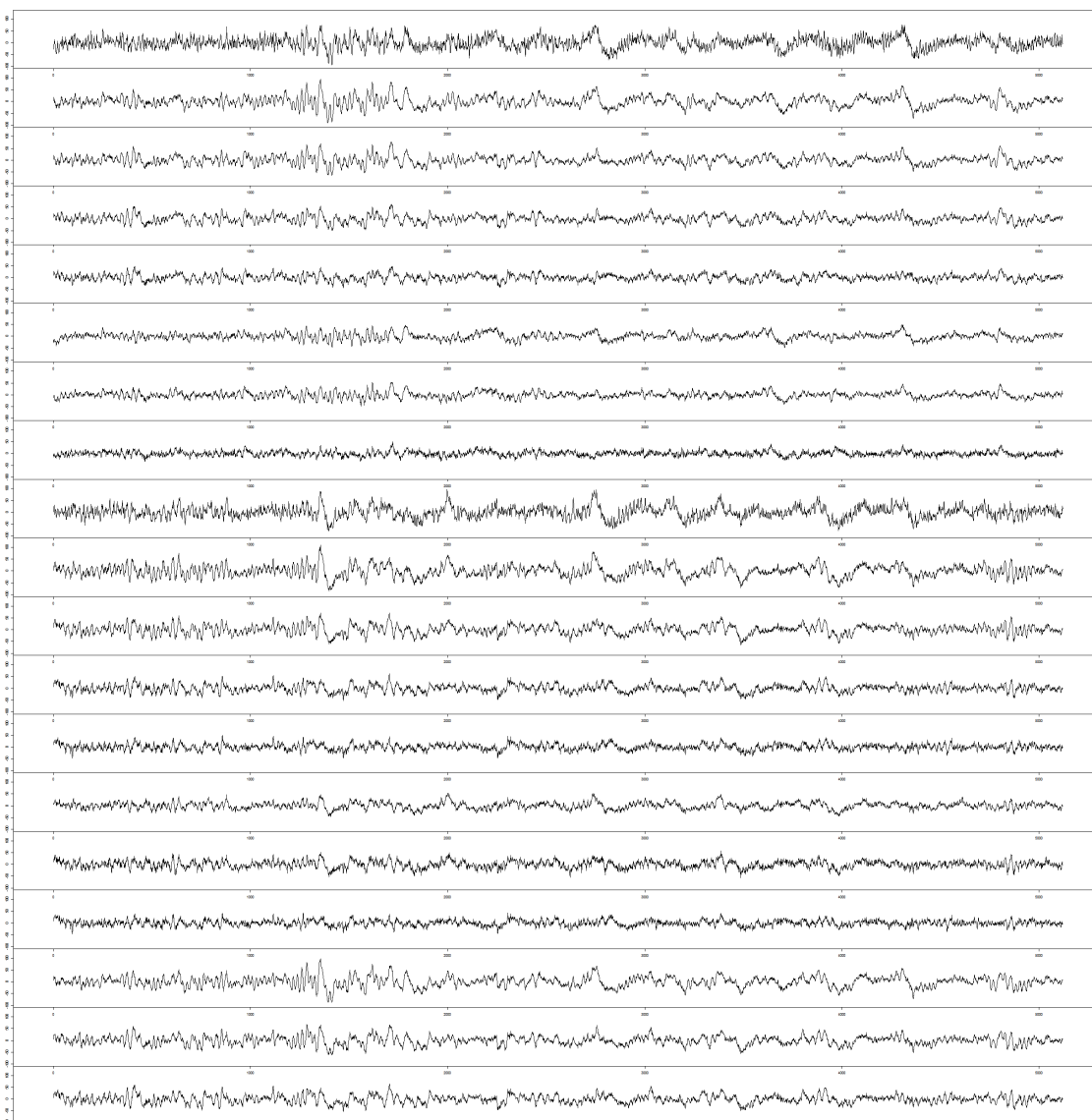
$$\mathbf{w}_i = \mathbf{f}(s_i) + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n = 2^J, \quad J = 14,$$

dove  $\boldsymbol{\epsilon}_i$  sono errori normali multivariati i.i.d. di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $S$  generica e incognita. La scelta di un modello di generazione delle osservazioni con errori correlati è giustificata dal fatto che, durante l'acquisizione del segnale EEG, ciascun elettrodo registra una differente proiezione della stessa dinamica cerebrale affetta da errori di misurazione o altre fonti di rumore dovute, ad esempio, ad attività muscolare.

È pertanto necessario introdurre una base funzionale che permetta di ricostruire opportunamente il dato a partire dalle sue istanze di campionamento, effettuando un'operazione di *smoothing* tale da filtrare la parte di variabilità non rilevante ai fini dell'analisi (il rumore) e che consenta inoltre di sopperire alla mancanza di regolarità tipica delle osservazioni discrete.

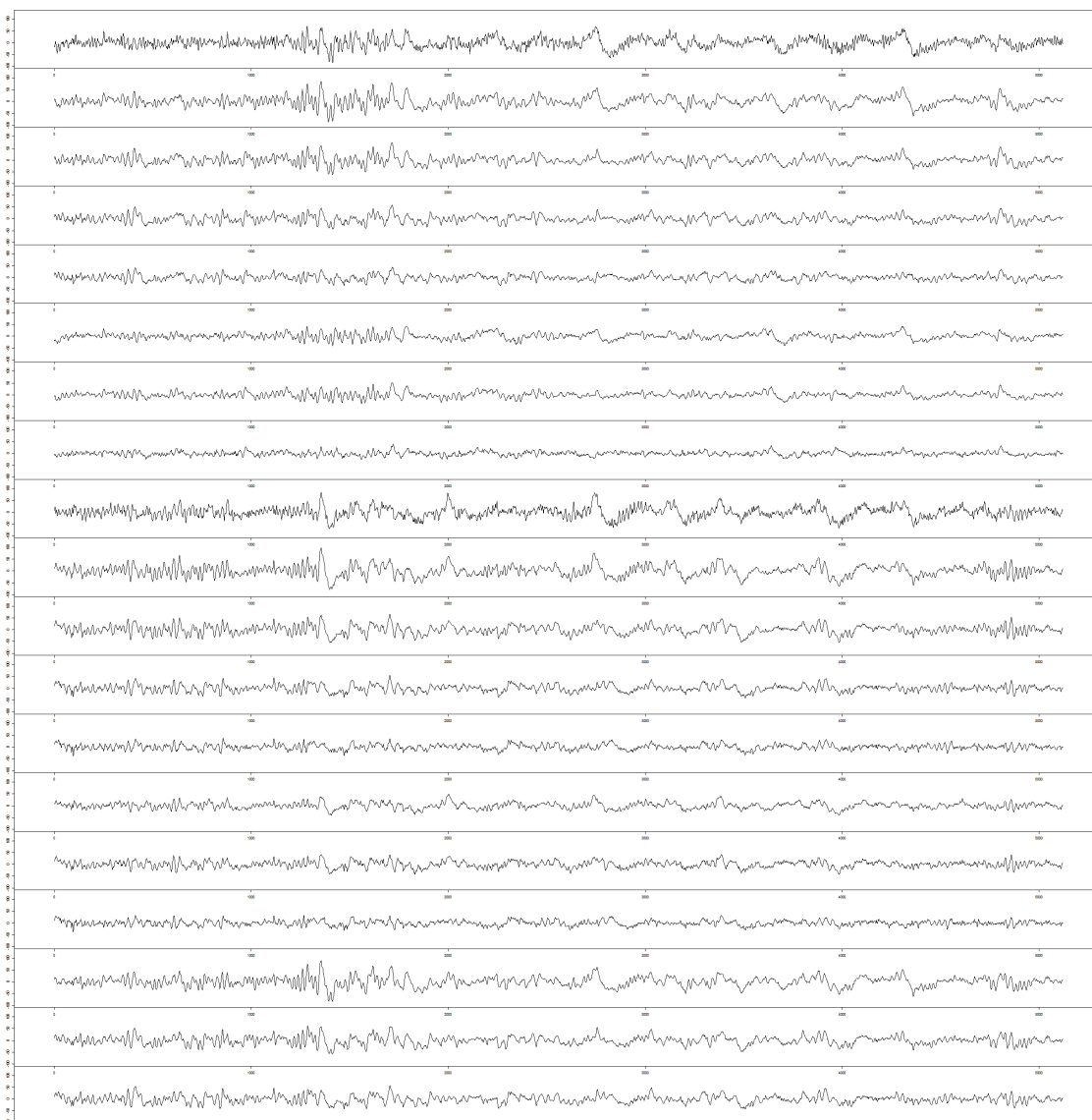
In questo lavoro di tesi lo smoothing dei dati funzionali è stato effettuato utilizzando le wavelets ortogonali di Daubechies di ordine 10 (cioè con 10 momenti nulli), molto usate in letteratura per le loro proprietà di regolarità. La scelta delle wavelets come base funzionale è giustificata dalle loro proprietà di localizzazione sia in tempo che in frequenza, oltre che dal fatto che forniscono una rappresentazione sparsa del segnale.

Lo stimatore utilizzato per ricostruire i dati ed eliminare possibili artefatti dovuti ad attività muscolare è quello proposto da Pigoli e Sangalli in [22] e discusso in Sezione 2.2.3.1, in quanto tiene conto della natura vettoriale dei dati e della correlazione tra le componenti di errore nelle  $p$  direzioni tipica dei segnali biomedici (e in particolare dunque anche degli EEG), e inoltre utilizza la stessa base wavelets per stimare tutte le componenti. In particolare il valore di soglia utilizzato è stato scelto in accordo alla Proposizione 2.2.2, lo shrinkage è stato effettuato sui vettori dei coefficienti di ogni livello dell'espansione wavelets e la matrice  $S$  è stata stimata con la matrice di covarianza campionaria dei coefficienti  $\mathbf{d}_{J-1,k}$  dell'ultimo livello.



(a)

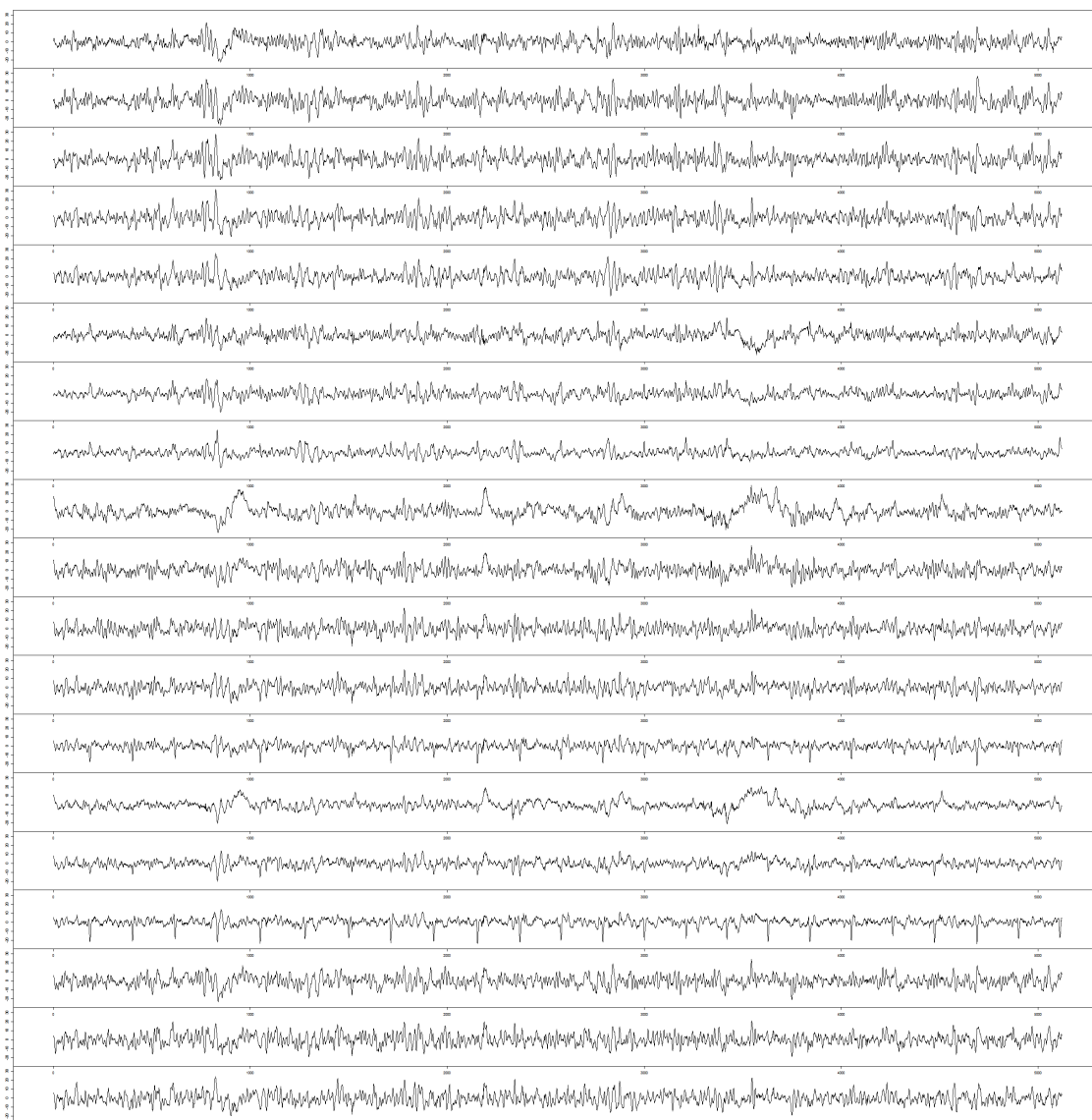
Figura 3.1: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, ottenuto interpolando linearmente le istanze di campionamento.



(b)

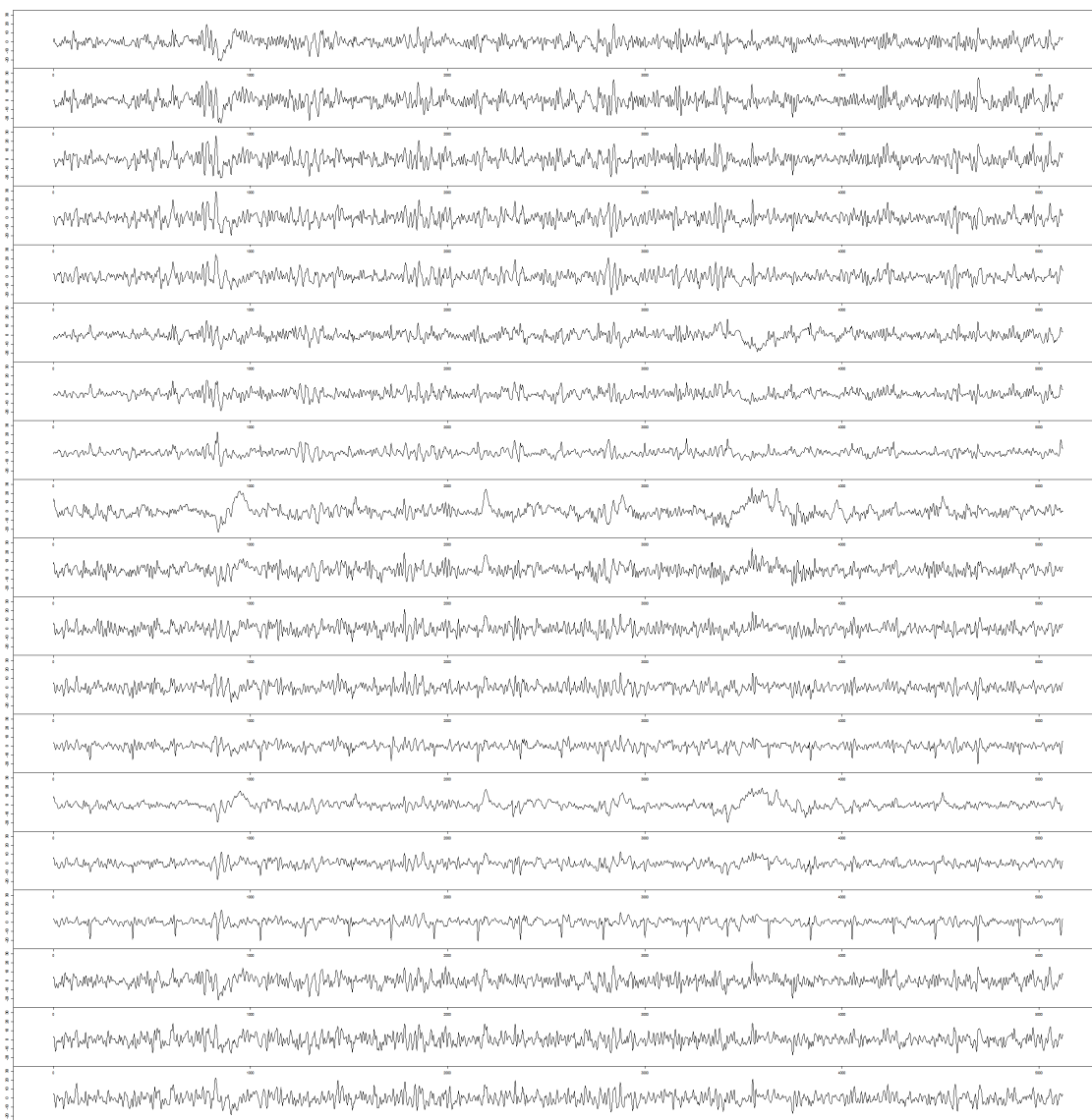
Figura 3.1: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets.





(a)

Figura 3.2: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, ottenuto interpolando linearmente le istanze di campionamento.



(b)

Figura 3.2: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets.

Le Figure 3.1.(a)-(b) e 3.2.(a)-(b) mostrano, su un frammento di segnale della durata di 20 secondi, la procedura di smoothing via wavelets effettuata sugli EEG del Paziente 10, rispettivamente in sonno e in veglia.

## 3.2 Standardizzazione dei dati e PCA

L'elettroencefalogramma registra in corrispondenza di 19 canali l'evoluzione temporale dell'attività elettrica cerebrale per mezzo di elettrodi posizionati sullo scalpo in posizioni opportune in accordo con il Sistema Internazionale 10-20.

Con l'obiettivo di effettuare una riduzione dimensionale del dato, e quindi di sintetizzare l'informazione contenuta nei 19 canali, siamo pertanto interessati ad indagare la struttura di varianza-covarianza dei canali osservati nel tempo. Ciò può essere fatto effettuando su tutti gli EEG stimati via wavelets a disposizione, sia quelli relativi al sonno che quelli relativi alla veglia, un' *Analisi delle Componenti Principali* (PCA) in cui le variabili aleatorie in esame sono i 19 canali mentre le repliche sono i 16384 istanti di campionamento.

I segnali elettroencefalografici sono però caratterizzati da un'ampia variabilità sia tra individui diversi, imputabile a fattori come ad esempio l'età o il sesso, che intra-individuale, a livello dei singoli canali di registrazione dell'EEG oppure in relazione allo stato psicologico e al livello di vigilanza del soggetto. Scegliamo dunque di effettuare la PCA sugli EEG ricostruiti via wavelets e standardizzati, in quanto in questa fase dell'analisi siamo interessati alle dipendenze e ai legami tra i canali e non all'intensità dei loro segnali.

Le Figure 3.3 e 3.4 mostrano gli autovalori delle matrici di correlazione dei canali relative agli EEG registrati rispettivamente in sonno e in veglia, e suggeriscono l'utilizzo di tre componenti principali per quasi tutti i pazienti. Pertanto, per privilegiare l'uniformità dell'analisi tra i vari soggetti, scegliamo di considerare come significative per tutti i pazienti, sia per il sonno che per la veglia, tre componenti principali.

Analizzando i loadings delle basi individuate dalla PCA per i vari soggetti, si nota che sia per la veglia che per il sonno è tendenzialmente possibile

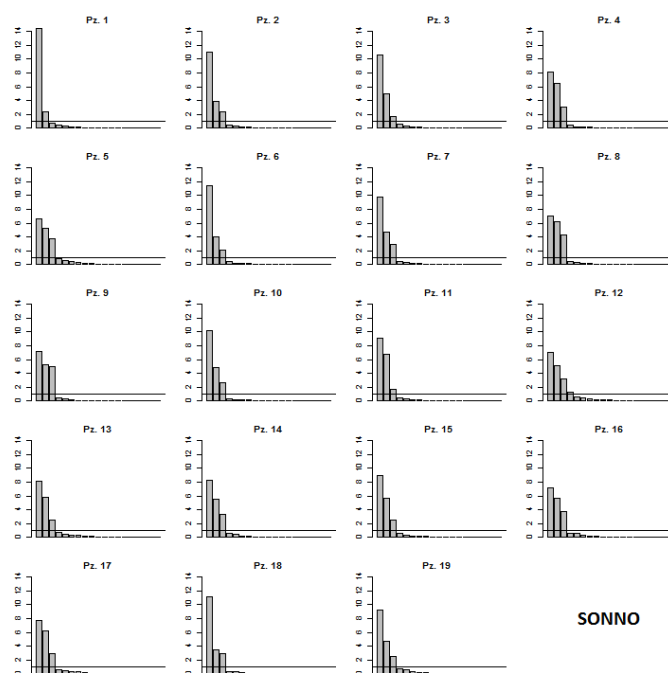


Figura 3.3: Autovalori delle matrici di correlazione dei canali relative agli EEG registrati in sonno.

riordinare le prime tre componenti principali (solamente in un caso infatti, EEG del Paziente 15 registrato in veglia, la terza componente di cui terremo conto è in realtà la quarta quanto a variabilità spiegata) in accordo alla somiglianza tra i loadings; tale riordinamento è stato effettuato *visivamente*.

Le componenti riordinate, che chiameremo d'ora in avanti *Componente A*, *Componente B* e *Componente C*, hanno un chiaro significato fisico:

- la **Componente A** rappresenta una media dell'attività cerebrale registrata dai 19 canali;
- la **Componente B** contrasta l'attività cerebrale registrata dai canali dell'emisfero destro con quella acquisita dai canali dell'emisfero sinistro;
- la **Componente C** contrasta l'attività cerebrale registrata nella parte anteriore dei due emisferi con quella acquisita dai canali posizionati nella parte posteriore dello scalpo.

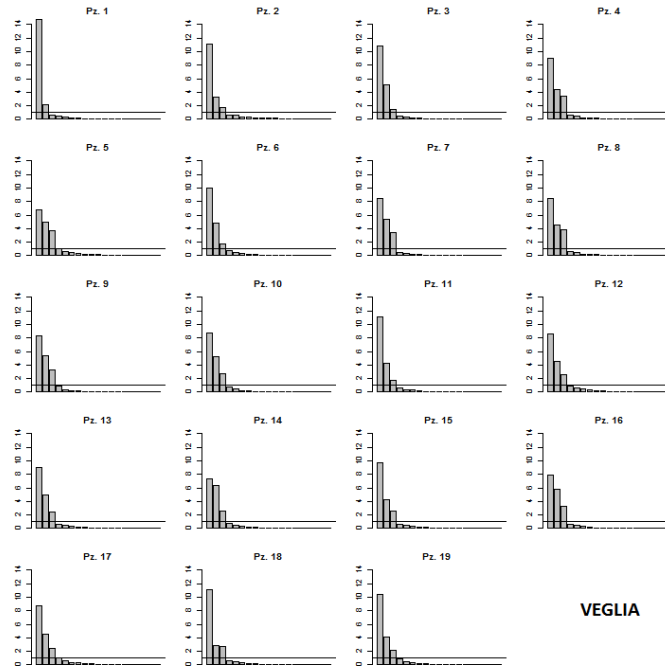


Figura 3.4: Autovalori delle matrici di correlazione dei canali relative agli EEG registrati in veglia.

I loadings  $e_{ik}$  delle Componenti A, B e C, ognuno dei quali è indice dell'importanza della  $k$ -esima variabile del vecchio sistema di riferimento (cioè del  $k$ -esimo canale) nell' $i$ -esima componente principale ( $i = A, B, C$ ), sono riportati nelle Figure A.3.(a)-(c) e A.4.(a)-(c) in Appendice A, rispettivamente per il sonno e per la veglia. In particolare, per quanto riguarda le componenti B e C, i canali sono stati riordinati e colorati in base alla posizione sullo scalpo del rispettivo elettrodo, in modo da metterne in risalto il significato associato. Si nota come in alcuni casi accada che due componenti, in particolare quando spiegano quote di variabilità simili, tendano a mescolarsi e a “confondere” il proprio significato. Tuttavia come risulta chiaro dalle Figure A.5.(a)-(d) e A.6.(a)-(d) in Appendice A, nelle quali gli elettrodi di acquisizione del segnale elettroencefalografico sono colorati in base al valore assunto dai loadings della rispettiva componente, il significato da attribuire agli spazi individuati su cui proiettare i dati è comunque lo stesso, a cambiare è solamente la base in cui vengono rappresentati. Infatti l’“atomo” di questi spazi è dato dal quarto di scalpo (anteriore sinistro, anteriore destro,

posteriore sinistro o posteriore destro).

Infine, le Figure 3.7.(a)-(b) e 3.8.(a)-(b) mettono a confronto, per il Paziente 10, il dato stimato via wavelets e standardizzato globalmente con il dato ricostruito a partire dalle Componenti A, B e C individuate dalla PCA, rispettivamente in sonno e in veglia. I grafici confermano che il dato ricostruibile a partire dalle tre componenti considerate costituisce una buona approssimazione dei segnali originari contenuti nei canali.

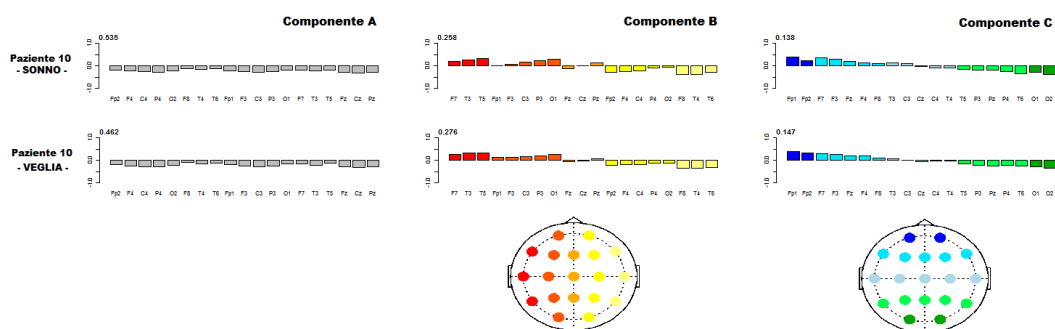


Figura 3.5: Loadings delle Componenti A, B e C relative agli EEG in sonno e in veglia del Paziente 10. I numeri in alto a sinistra fanno riferimento alle quote di variabilità spiegata dalle tre componenti. I canali delle componenti B e C sono stati riordinati per evidenziare i contrasti.

**Osservazione 1.** Il segnale elettroencefalografico non è stazionario; perciò, per poter utilizzare determinati modelli matematici che richiedono stazionarietà del segnale, solitamente si individuano spezzoni temporali all'interno dei quali le caratteristiche statistiche dei segnali si possano considerare costanti con buona approssimazione. La scelta della durata di tali finestre temporali deriva da un compromesso ragionevole tra il tentativo di avere un segnale stazionario e la necessità di selezionare un numero sufficiente di dati per poter caratterizzare adeguatamente la dinamica del segnale. A seconda della complessità del segnale, questa finestra può variare da pochi millisecondi a decine di secondi.

Pertanto, per avere conferma del fatto che l'analisi effettuata standardizzando il dato globalmente non si limitasse a spiegare il drift (valore medio della variabile in funzione della localizzazione) del segnale, è stata eseguita una ulteriore indagine nella quale ogni dato è stato suddiviso in 16 segmenti

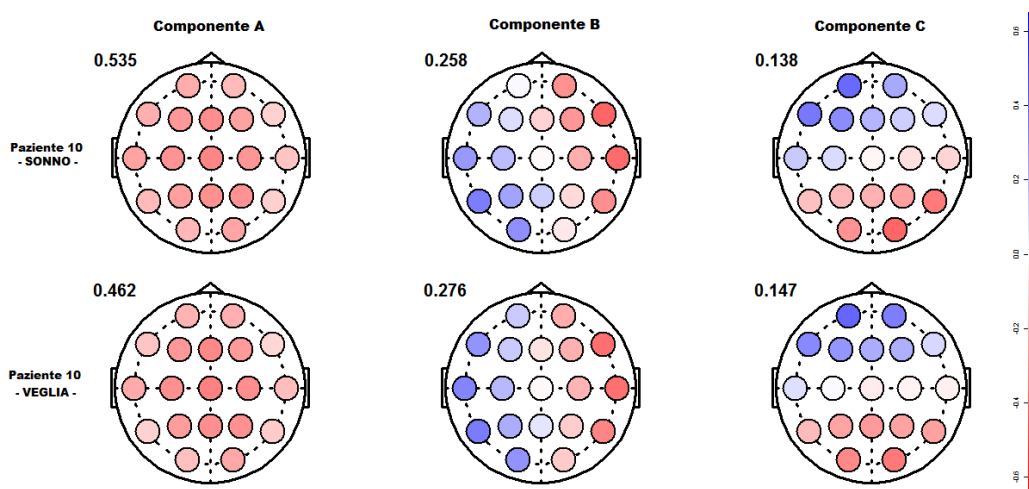


Figura 3.6: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG in sonno e in veglia del Paziente 10.

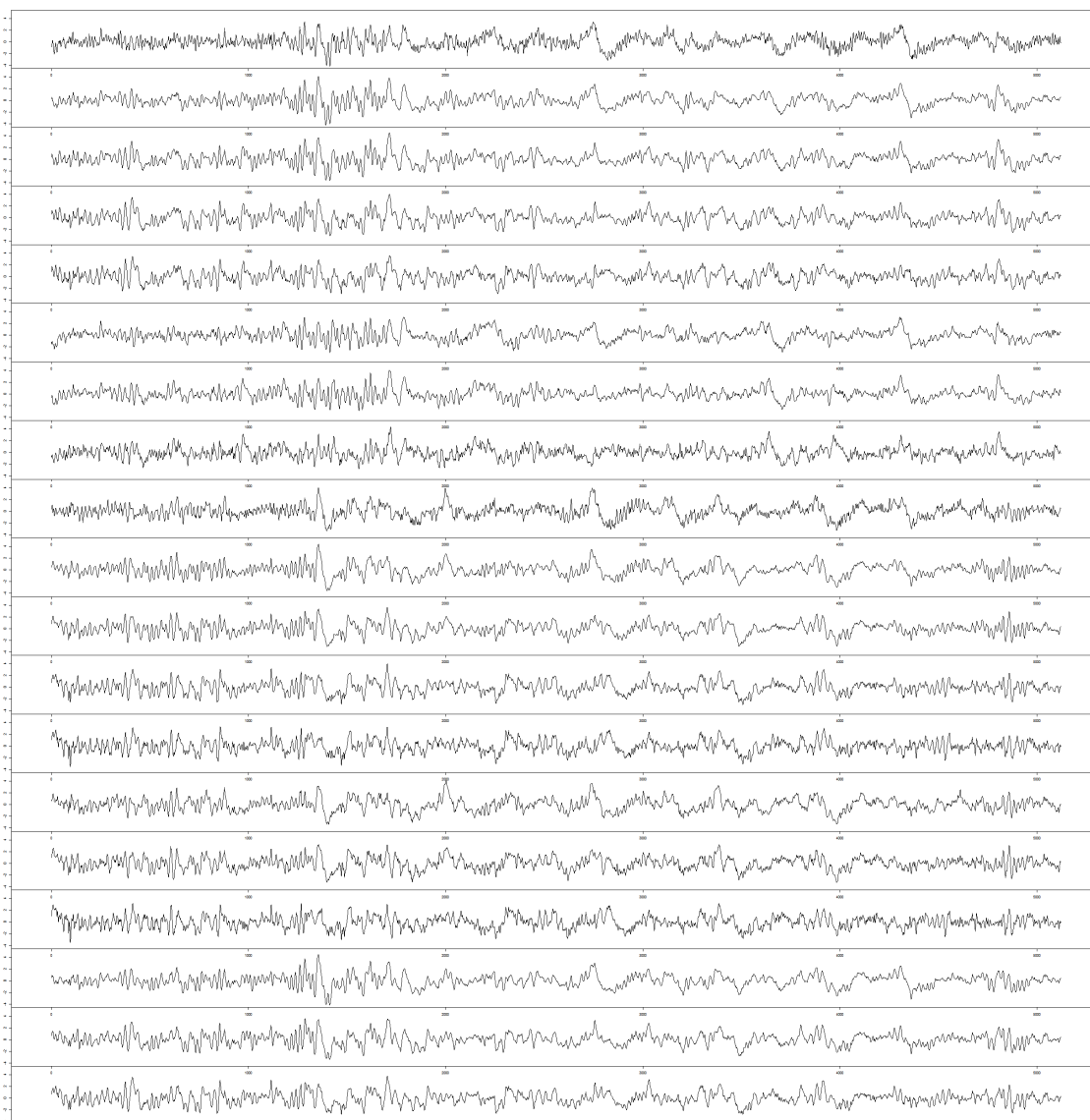
non sovrapposti (ciascuno della durata quindi di 4s, pari a 1024 istanti di campionamento), e su ogni blocco, dopo averlo standardizzato, si è seguito l'approccio precedentemente spiegato. I risultati che si ottengono sono comparabili con quelli della precedente analisi; infatti, sebbene i loadings vengono stimati su porzioni di dato, la struttura di PCA che si ottiene è la stessa.

A supporto di questa affermazione, in Appendice A si riportano nelle Figure A.8.(a)-(d) e A.9.(a)-(d) i loadings ottenuti nei singoli blocchi per il Paziente 10, considerando rispettivamente il dato registrato in sonno e in veglia.

### 3.3 Analisi spettrale

L'analisi spettrale tramite trasformata di Fourier (e in particolare via FFT) è la tecnica più comunemente utilizzata per quantificare il contenuto in frequenza dell'attività cerebrale misurata per mezzo dell'EEG.

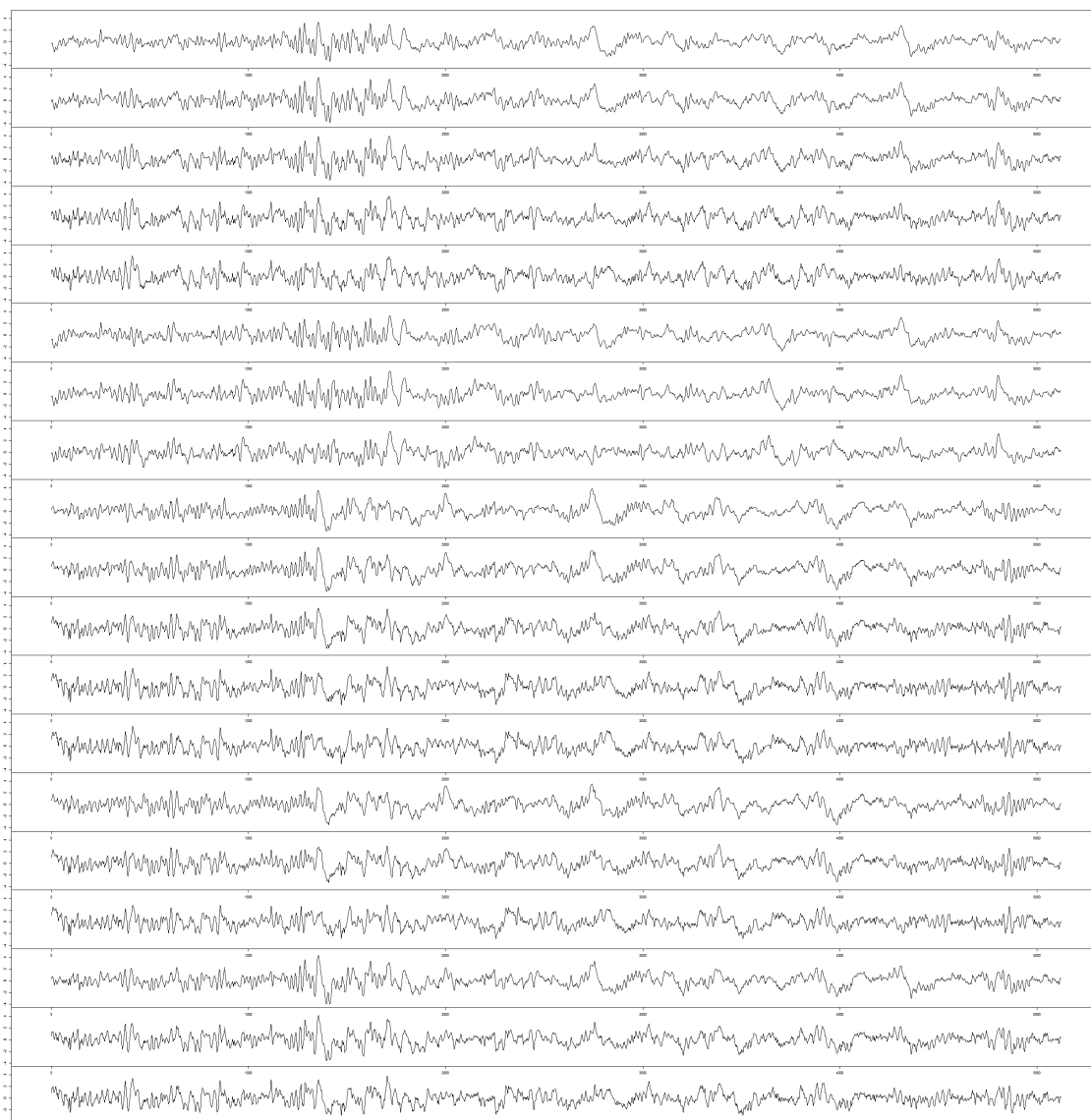
La massima frequenza spettrale ricostruibile per i dati in esame, campionati a 256 Hz, è pari a 128 Hz. Gli intervalli interquantile al 90% degli spettri di ampiezza in Figura A.7.(a)-(d), ricavati per ogni EEG mediando



(a)

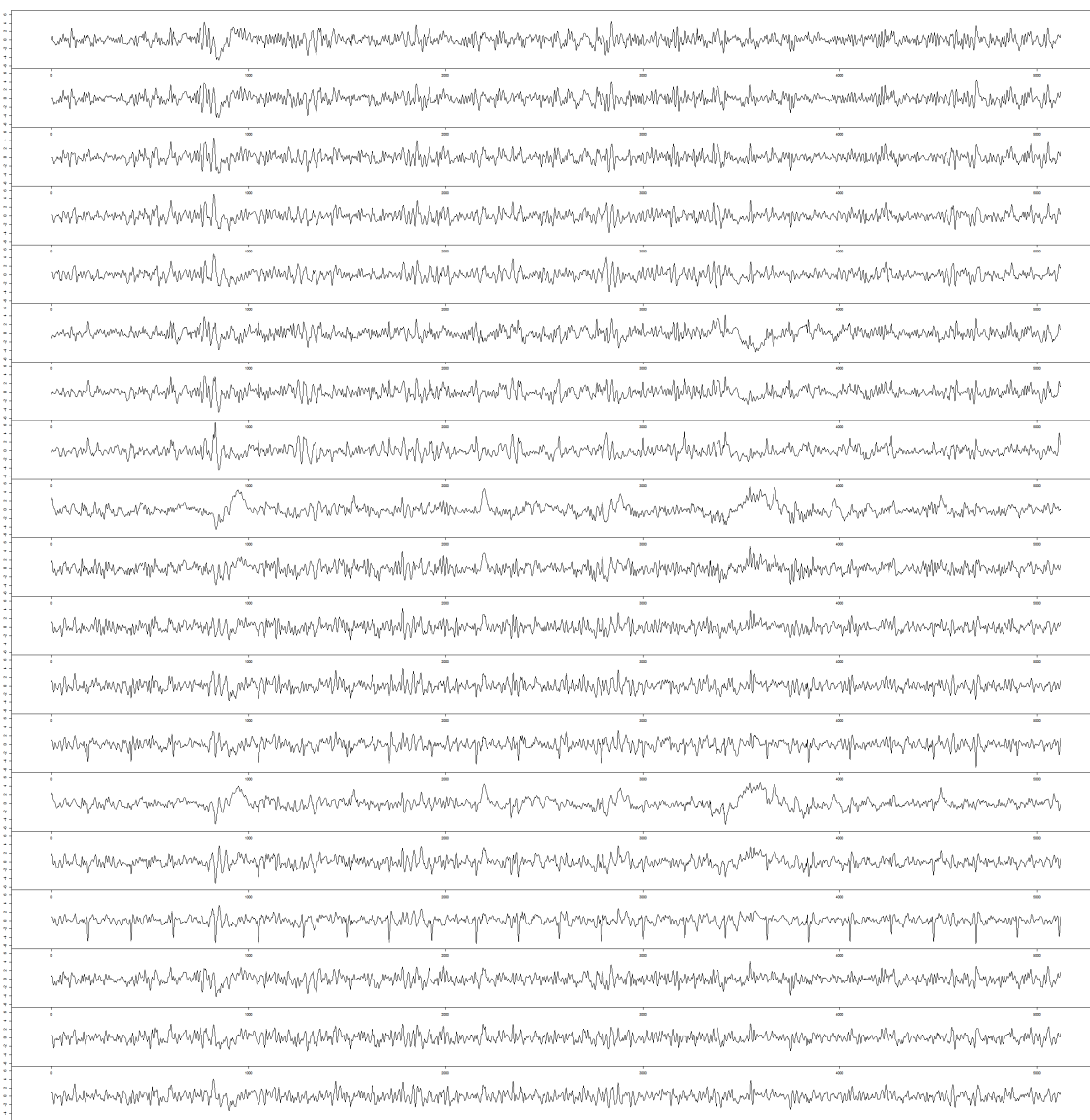
Figura 3.7: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets e la standardizzazione globale dei canali.





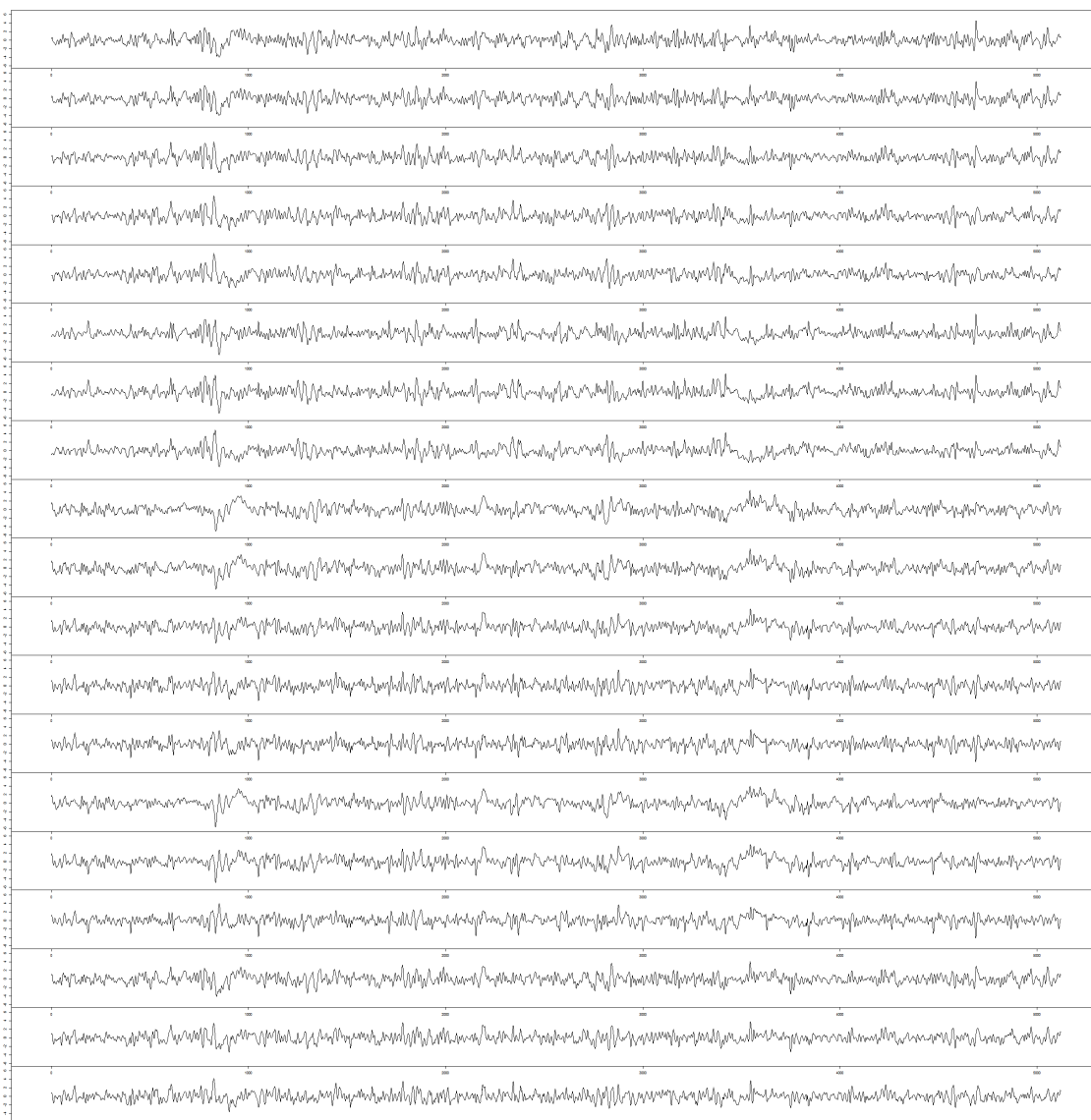
(b)

Figura 3.7: Ricostruzione del segnale raffigurato in Figura 3.7.(a) utilizzando le Componenti A, B e C.



(a)

Figura 3.8: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets e la standardizzazione globale dei canali.



(b)

Figura 3.8: Ricostruzione del segnale raffigurato in Figura 3.8.(a) utilizzando le Componenti A, B e C.

ad ogni frequenza e canale per canale i valori degli spettri ottenuti a partire da segmenti non sovrapposti di segnale lunghi 4s, e poi calcolando per ogni frequenza il valore mediano e i quantili di ordine 5% e 95% degli spettri ottenuti per i 19 canali, mostrano che il segnale ricostruito a partire dalle tre componenti principali considerate mantiene il contenuto in frequenza del segnale stimato tramite wavelets, e che anche per il residuo il contenuto in frequenza è dominante alle basse frequenze.

Oltre a ciò, possiamo ancora notare che i picchi di frequenza si registrano, sia per il sonno che per la veglia, all'interno delle bande di frequenza  $\delta$  o  $\theta$ .

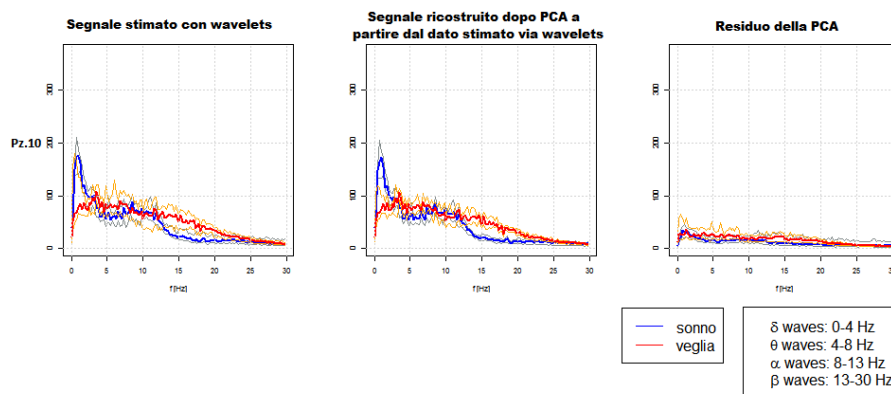


Figura 3.9: Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Paziente 10)

### 3.4 Cluster Analysis

L'analisi delle componenti principali ha permesso di ricavare per i soggetti in esame spazi tridimensionali simili su cui proiettare i dati, in quanto individuati da componenti a cui è possibile attribuire lo stesso significato fisico, ma con importanti differenze per quanto riguarda il valore assunto dai loadings ad esse associati. Inoltre, come si evince in particolare dalle Figure A.5.(a)-(d) e A.6.(a)-(d) in Appendice A, la sintesi dei dati effettuata per mezzo dei loadings ha permesso di mettere in risalto per alcuni pazienti la

presenza di zone spente di cervello, non direttamente osservabili esaminando i tracciati dei segnali registrati dai 19 canali. Ci chiediamo quindi se sia possibile usare queste informazioni ricavate sui loadings delle tre componenti per raggruppare correttamente i pazienti in base alla categoria di coscienza di appartenenza.

L'utilizzo della distanza euclidea come metrica per descrivere la vicinanza tra le unità in esame e dunque clusterizzare i soggetti in base al valore assunto dai loadings delle Componenti A, B e C non è però ottimale in questo caso, in quanto non tutti i pazienti manifestano come spenta la stessa area di cervello. I raggruppamenti che sono stati ottenuti ricorrendo agli algoritmi gerarchici agglomerativi, scegliendo come metrica la distanza euclidea e come distanze intercluster i metodi single, complete, average e Ward's linkage, hanno infatti confermato le criticità derivabili da una tale scelta.

Il prossimo passo dell'analisi, che richiederà l'implementazione di tecniche specifiche e che è segnalato come sviluppo futuro del lavoro di tesi, sarà pertanto finalizzato alla ricerca di una metrica che permetta di raggruppare tra loro pazienti caratterizzati dalla presenza di zone spente del cervello, non necessariamente le stesse.

## Conclusioni

In questo lavoro sono stati studiati gli elettroencefalogrammi in sonno e in veglia di pazienti in stato vegetativo e in stato di minima coscienza.

I segnali EEG possono essere interpretati dal punto di vista matematico-statistico come realizzazioni discrete, affette da errori di misurazione correlati nelle diverse coordinate, di funzioni aleatorie  $\mathbf{f} : \mathbb{R} \mapsto \mathbb{R}^{19}$ .

La prima fase dell'analisi ha pertanto riguardato la ricostruzione del dato a partire dalle sue istanze di campionamento, effettuata mediante un'operazione di *smoothing* tale da filtrare la parte di variabilità non rilevante ai fini dell'analisi e che ha permesso inoltre di sopperire alla mancanza di regolarità tipica delle osservazioni discrete. Tale operazione è stata effettuata ricorrendo all'utilizzo delle wavelets ortogonali di Daubechies di ordine 10 e in particolare allo stimatore proposto da Pigoli e Sangalli in [22] nel caso di errori normali multivariati i.i.d. di media  $\mathbf{0} \in \mathbb{R}^p$  e matrice di varianza-covarianza  $S$  generica e incognita.

Successivamente è stata studiata la struttura di correlazione dei 19 canali di acquisizione dell'EEG attraverso l'analisi delle componenti principali sui dati stimati via wavelets e standardizzati. Questa indagine ha permesso di trovare per tutti gli EEG in esame, dopo un riordinamento visivo effettuato in accordo alla somiglianza tra i loadings delle componenti ottenute, spazi tridimensionali simili su cui proiettare i dati dal chiaro significato fisico, in quanto la prima componente che li individua rappresenta una media dei segnali registrati dai 19 canali, mentre la seconda e la terza contrastano

rispettivamente l'attività cerebrale registrata negli emisferi destro-sinistro e nelle parti anteriore-posteriore dei due emisferi.

I grafici dei dati ricostruiti a partire dalle tre componenti individuate dalla PCA e gli intervalli interquantile degli spettri di ampiezza hanno confermato la bontà dell'approssimazione effettuata.

L'analisi dei loadings delle tre componenti individuate ha inoltre permesso di individuare per alcuni pazienti delle aree di cervello spente, non direttamente osservabili esaminando i tracciati dei segnali registrati dai 19 canali. Il prossimo passo dell'analisi pertanto sarà volto alla ricerca di una metrica che consenta di raggruppare i pazienti in base alla presenza o meno di tali zone.

Il lavoro svolto resta quindi aperto a possibili sviluppi futuri, quali ad esempio:

- l'implementazione di una funzione che automatizzi il procedimento di riallineamento delle Componenti A, B e C individuate mediante l'analisi delle componenti principali;
- l'implementazione di tecniche specifiche finalizzato alla ricerca di una metrica che permetta di raggruppare tra loro pazienti che manifestano la presenza di zone spente del cervello.

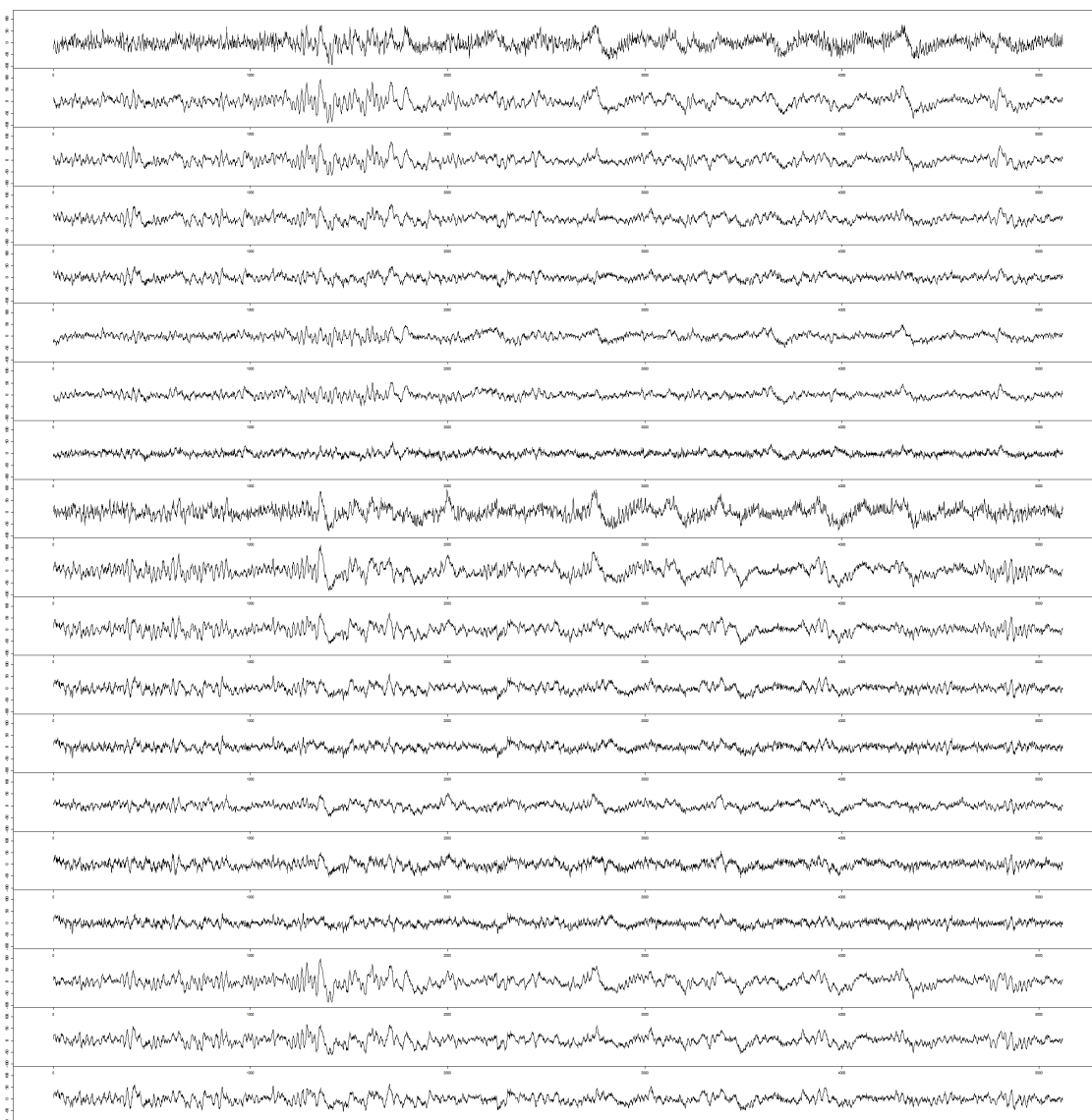
## Figure e Tabelle del Capitolo 3

In questa appendice sono riportate le figure del Capitolo 3, oltre che una tabella riassuntiva delle informazioni a disposizione sugli stati di coscienza dei pazienti in esame.

<b>ID</b> <b>Paziente</b>	<b>Stato di</b> <b>coscienza</b>	<b>CRS-R</b>	<b>ID.</b> <b>Paziente</b>	<b>Stato di</b> <b>coscienza</b>	<b>CRS-R</b>
Pz. 1	VS	5	Pz. 10	MCS	9
Pz. 2	VS	5	Pz. 11	MCS	9
Pz. 3	VS	6	Pz. 12	MCS	9
Pz. 4	VS	6	Pz. 13	MCS	9
Pz. 5	VS	6	Pz. 14	MCS	10
Pz. 6	VS	6	Pz. 15	MCS	10
Pz. 7	VS	7	Pz. 16	MCS	10
Pz. 8	VS	7	Pz. 17	MCS	11
Pz. 9	VS	8	Pz. 18	MCS	11
			Pz. 19	MCS	13

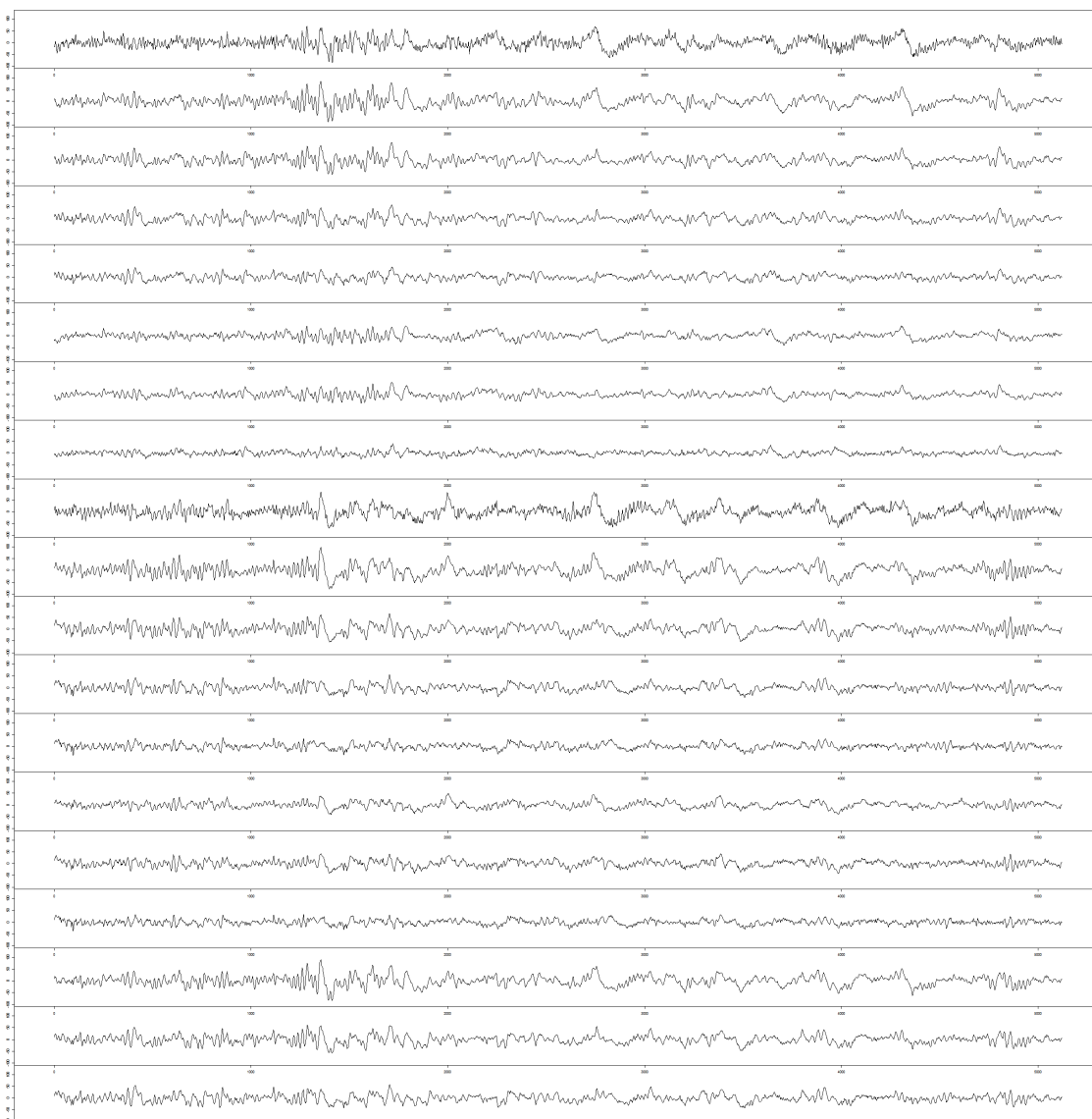
Tabella A.1: Identificativo, stato di coscienza e punteggio CRS-R dei pazienti esaminati.





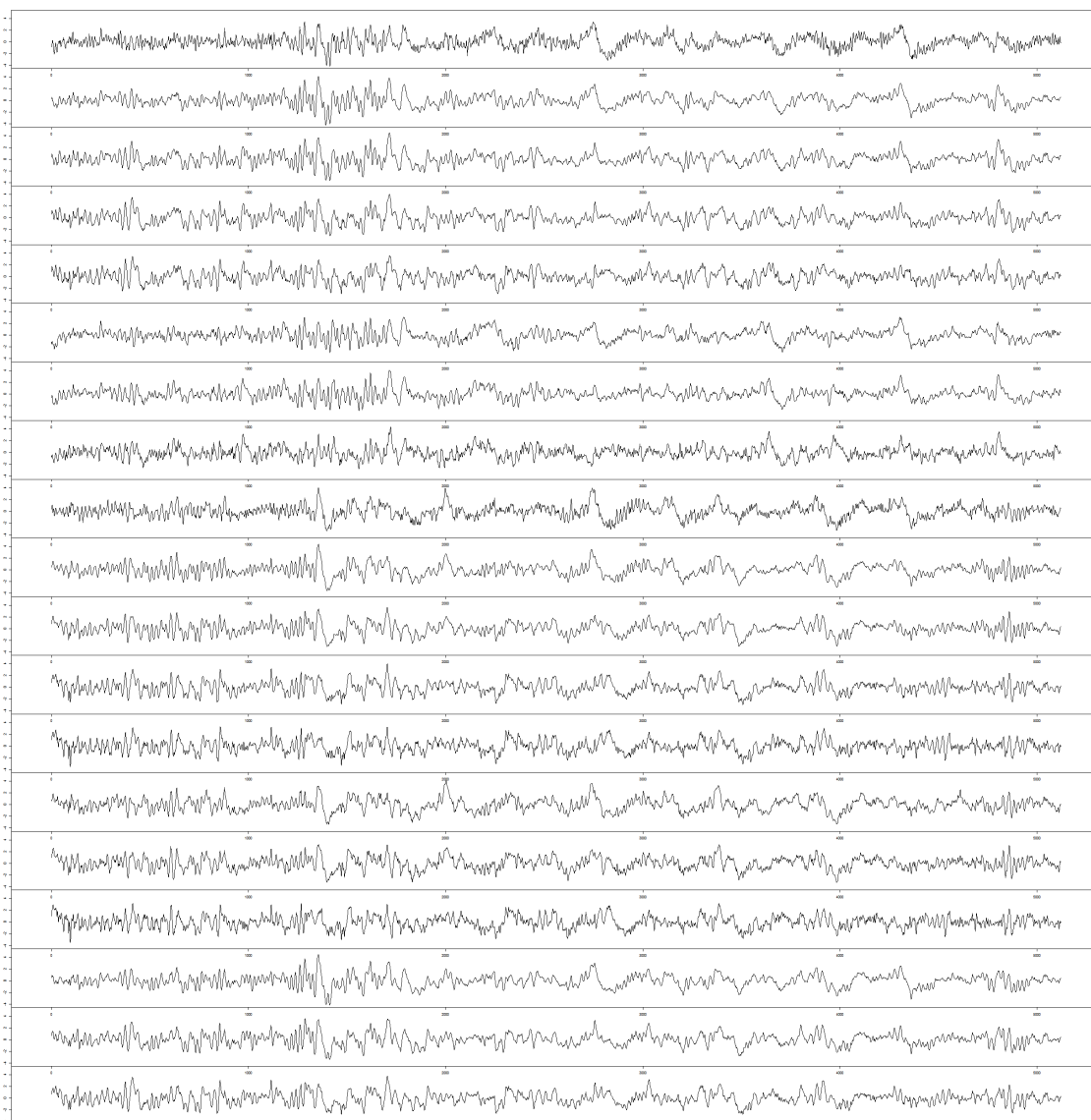
(a)

Figura A.1: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, ottenuto interpolando linearmente le istanze di campionamento.



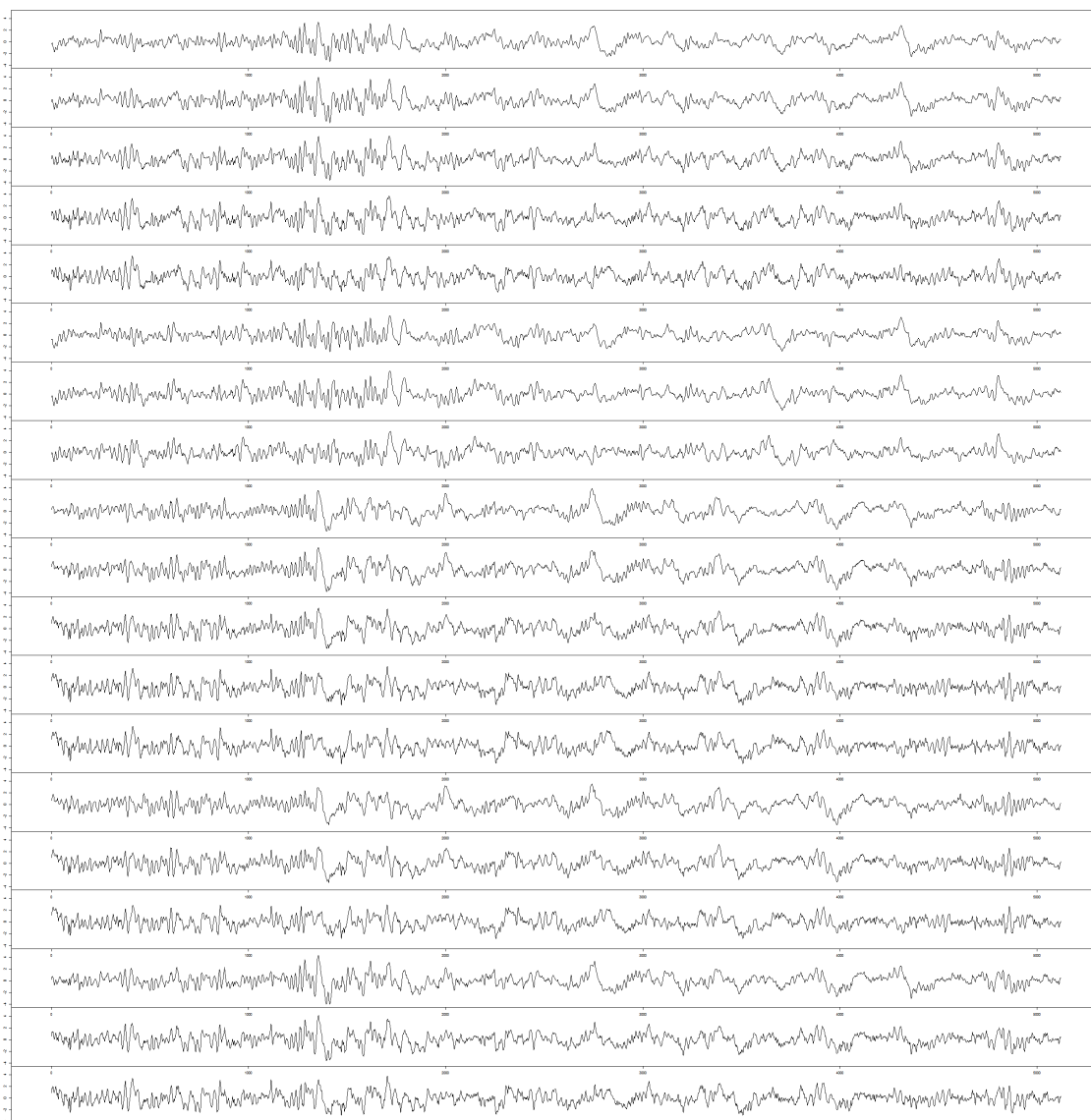
(b)

Figura A.1: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets.



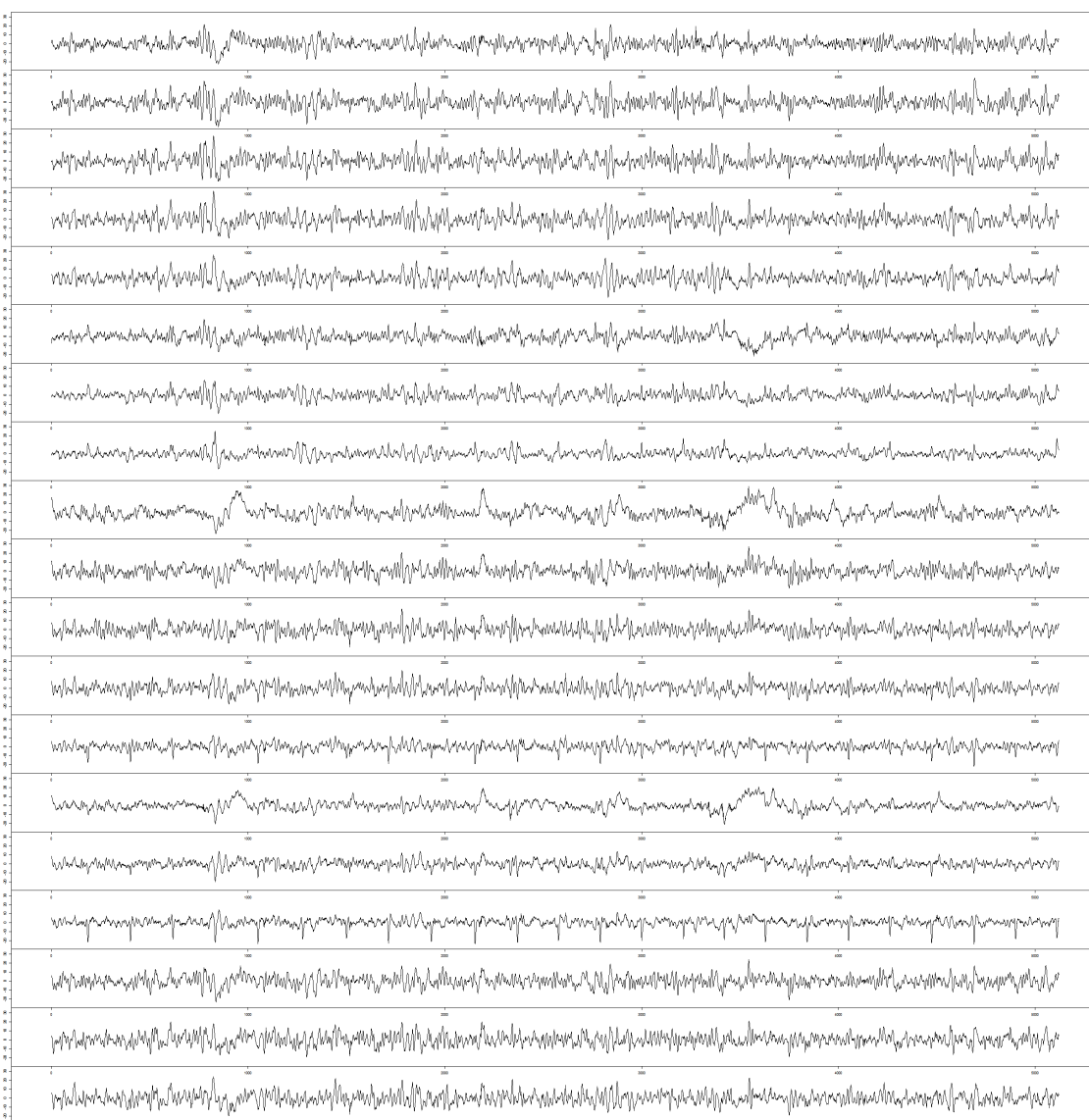
(c)

Figura A.1: Frammento EEG del Paziente 10, di durata 20s e registrato in sonno, dopo lo smoothing via wavelets e la standardizzazione globale dei canali.



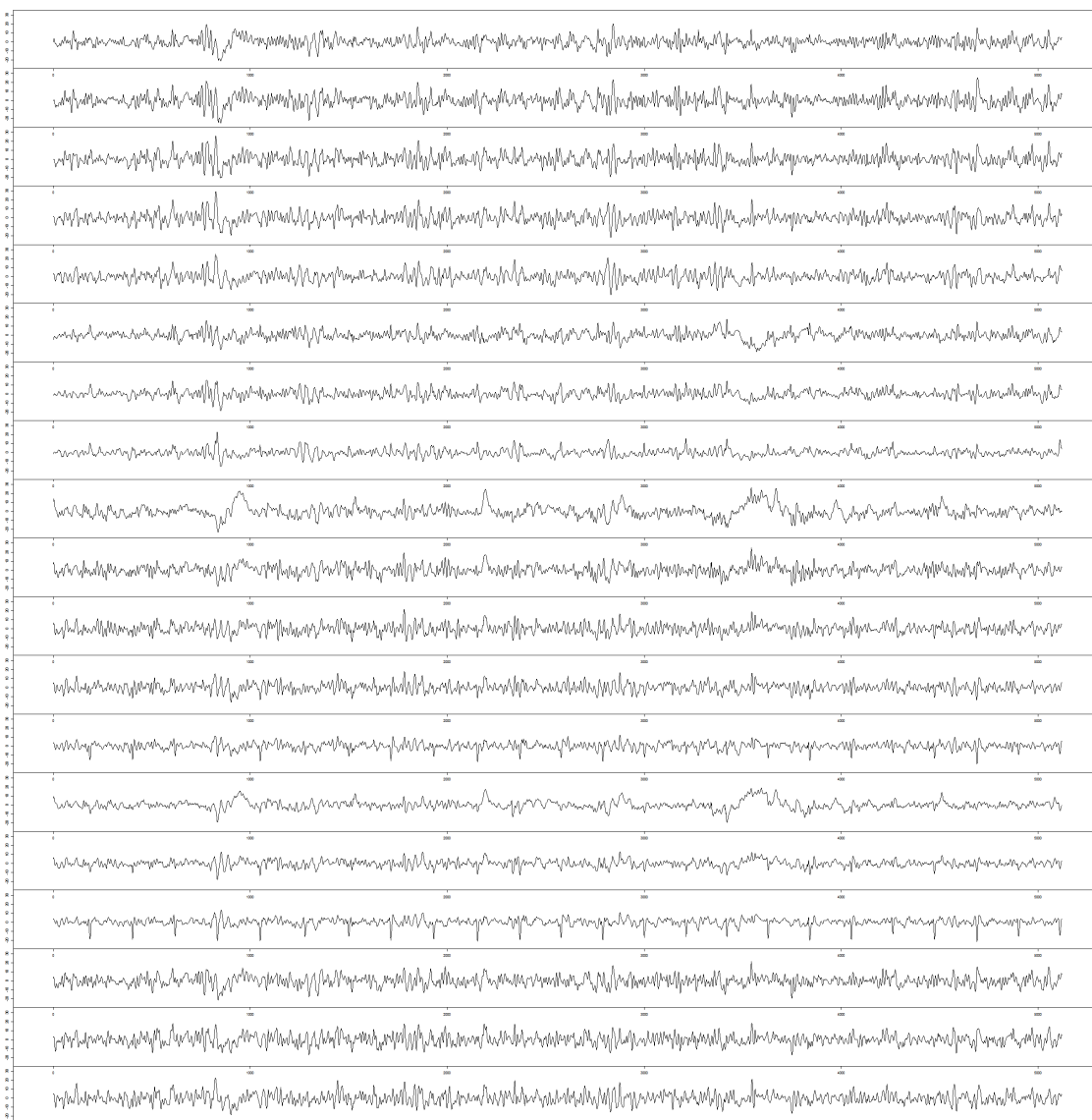
(d)

Figura A.1: Ricostruzione del segnale raffigurato in Figura A.1.(c) utilizzando le Componenti A, B e C.



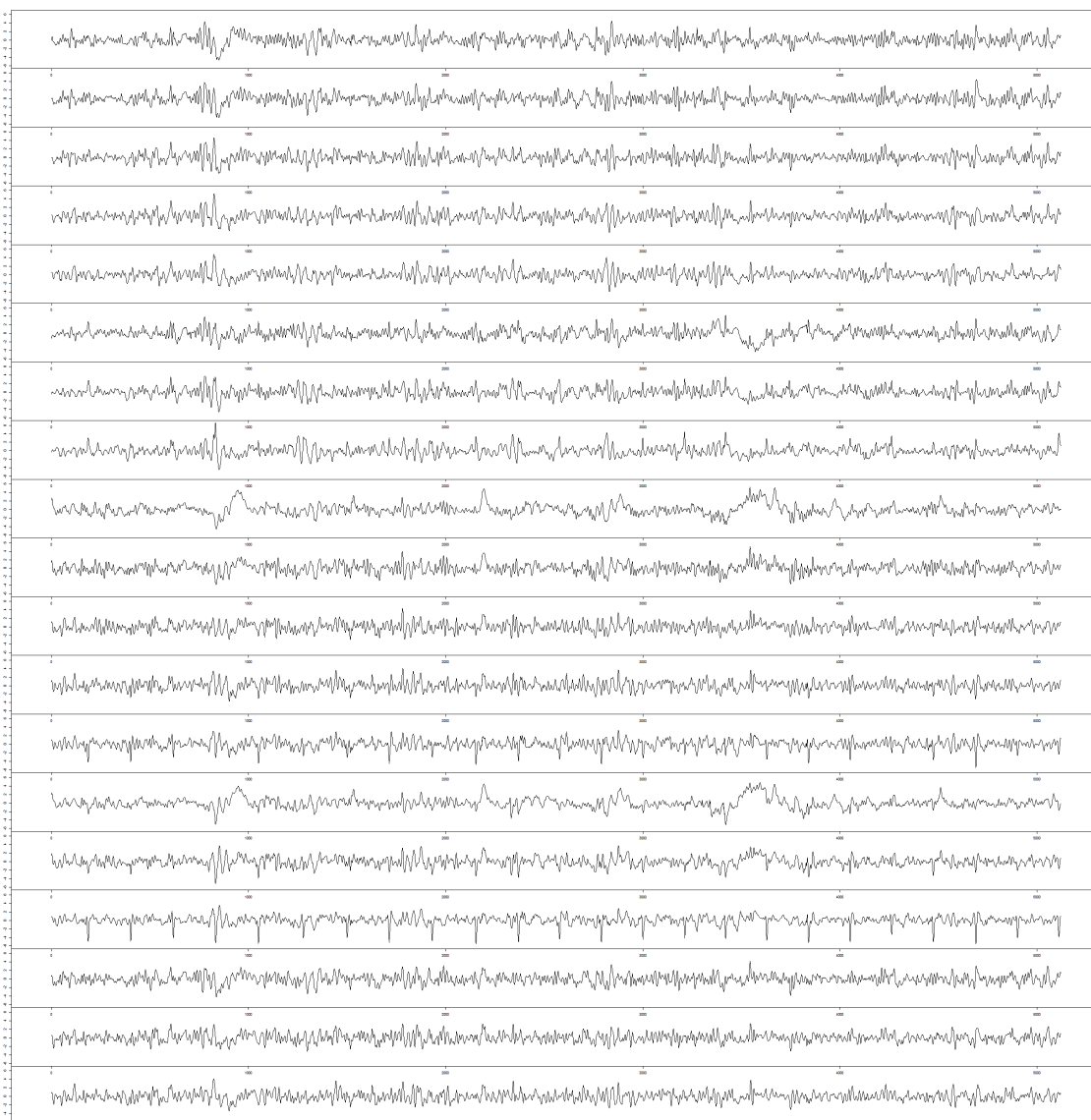
(a)

Figura A.2: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, ottenuto interpolando linearmente le istanze di campionamento.



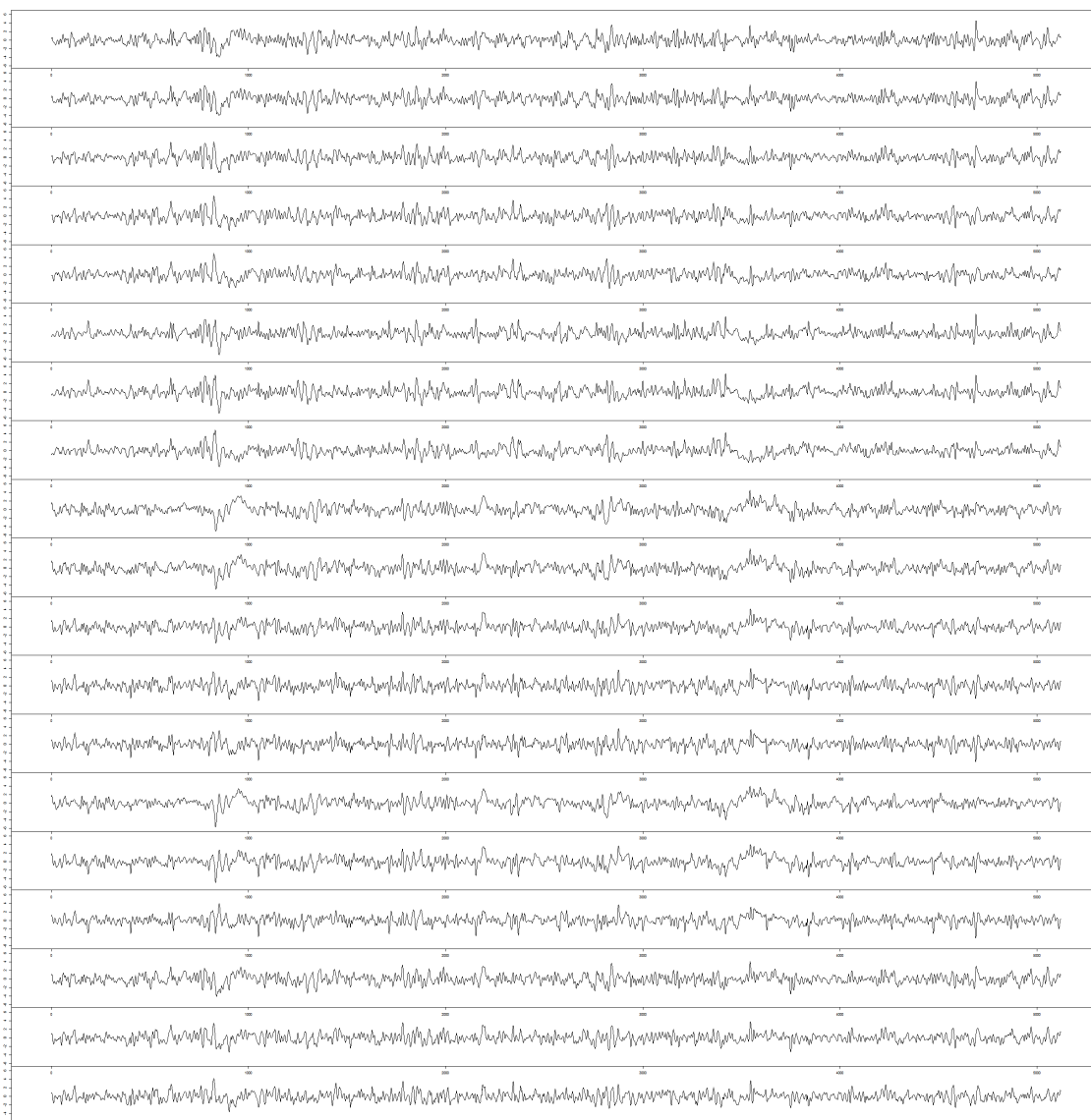
(b)

Figura A.2: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets.



(c)

Figura A.2: Frammento EEG del Paziente 10, di durata 20s e registrato in veglia, dopo lo smoothing via wavelets e la standardizzazione globale dei canali.



(d)

Figura A.2: Ricostruzione del segnale raffigurato in Figura A.2.(c) utilizzando le Componenti A, B e C.



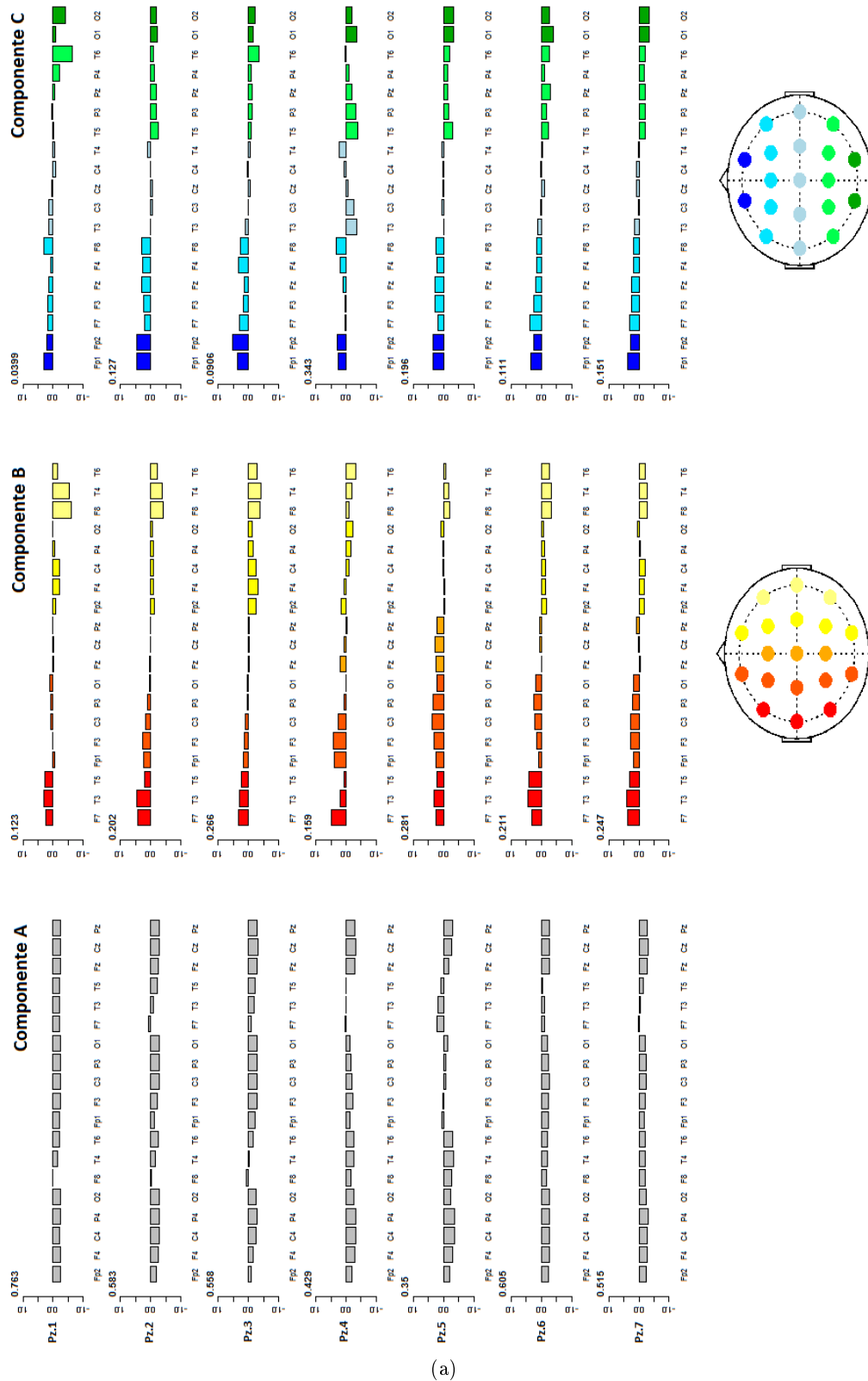


Figura A.3: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 1 a 7)

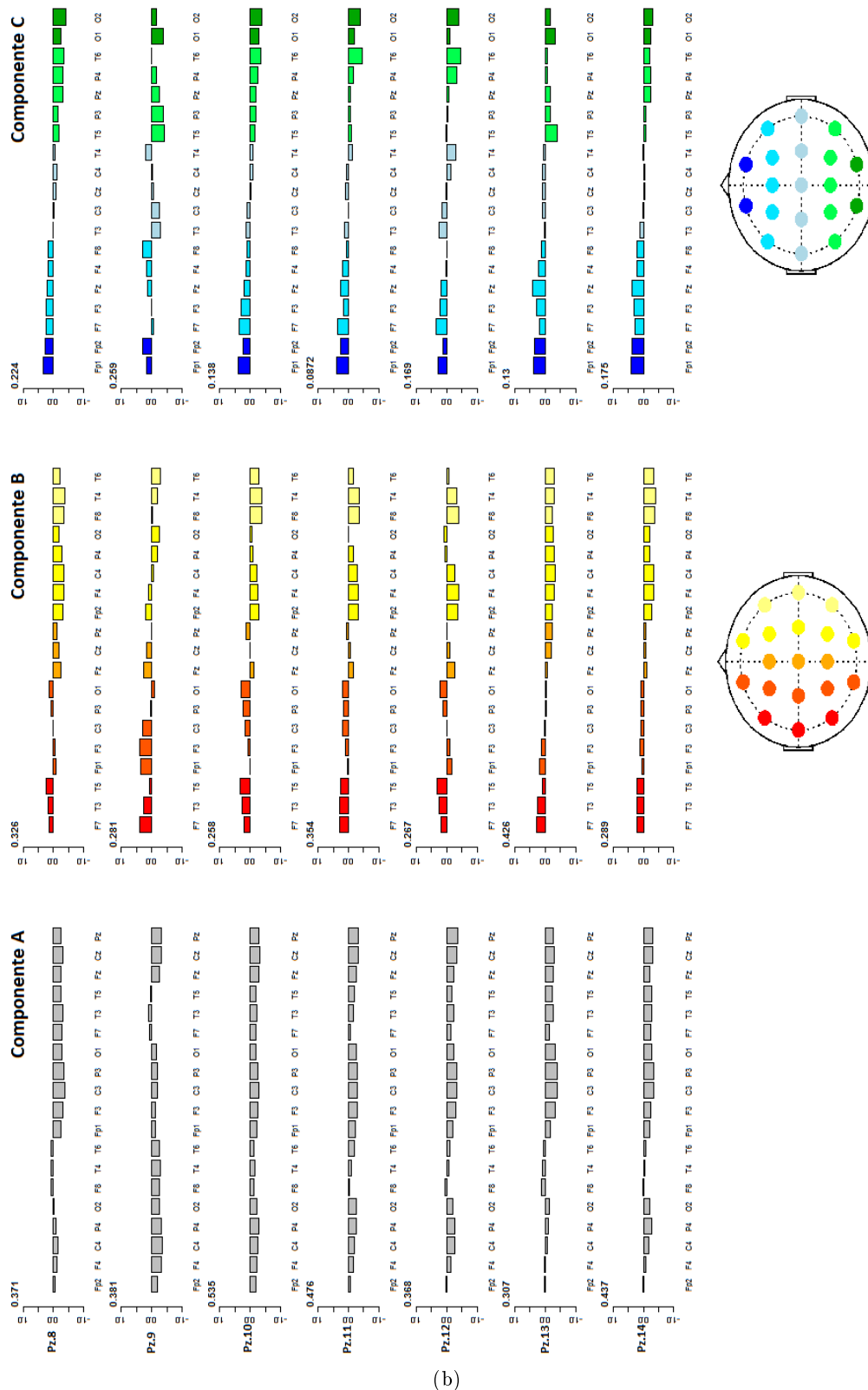


Figura A.3: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 8 a 14)

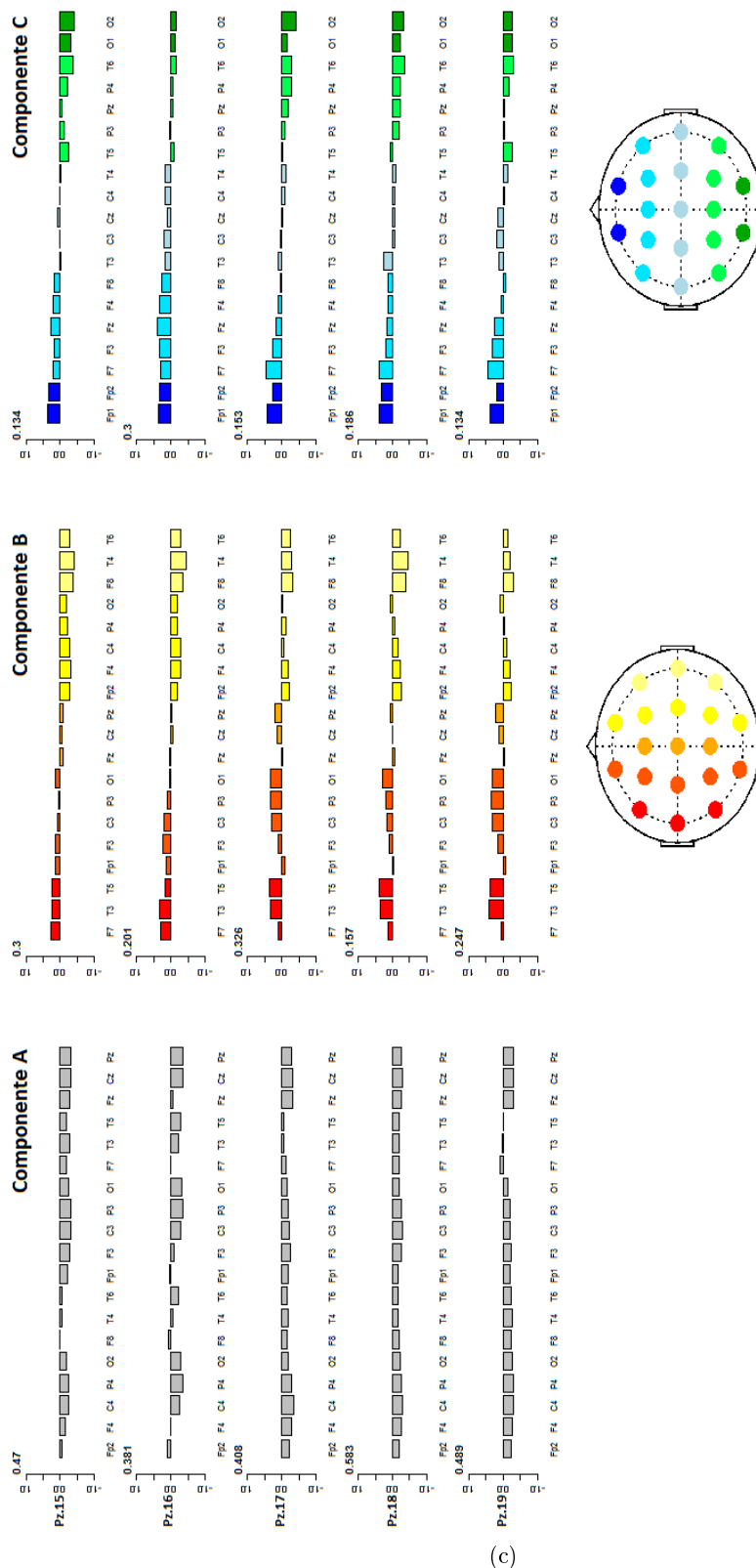


Figura A.3: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in sonno. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti (Pazienti da 15 a 19)

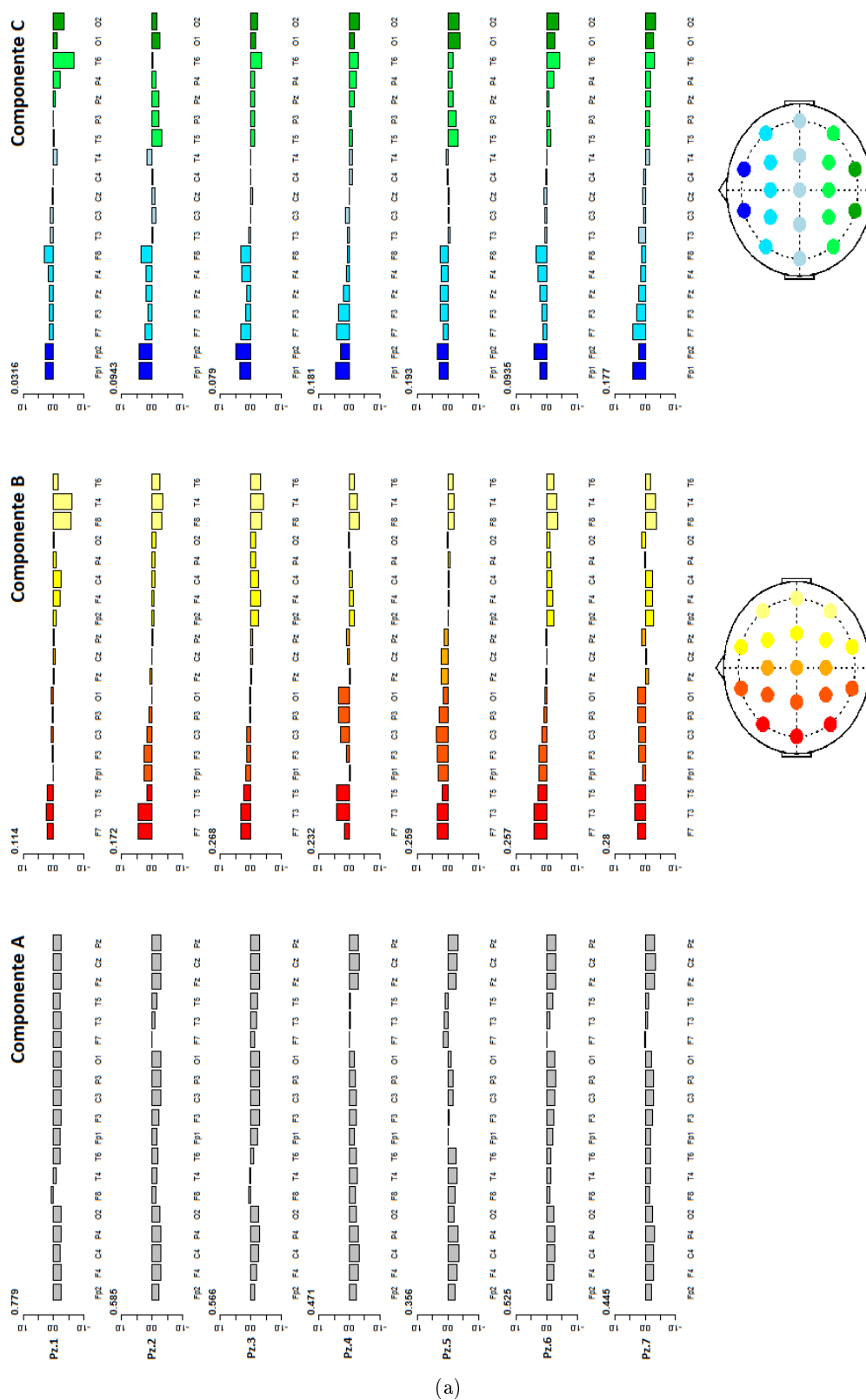


Figura A.4: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 1 a 7)

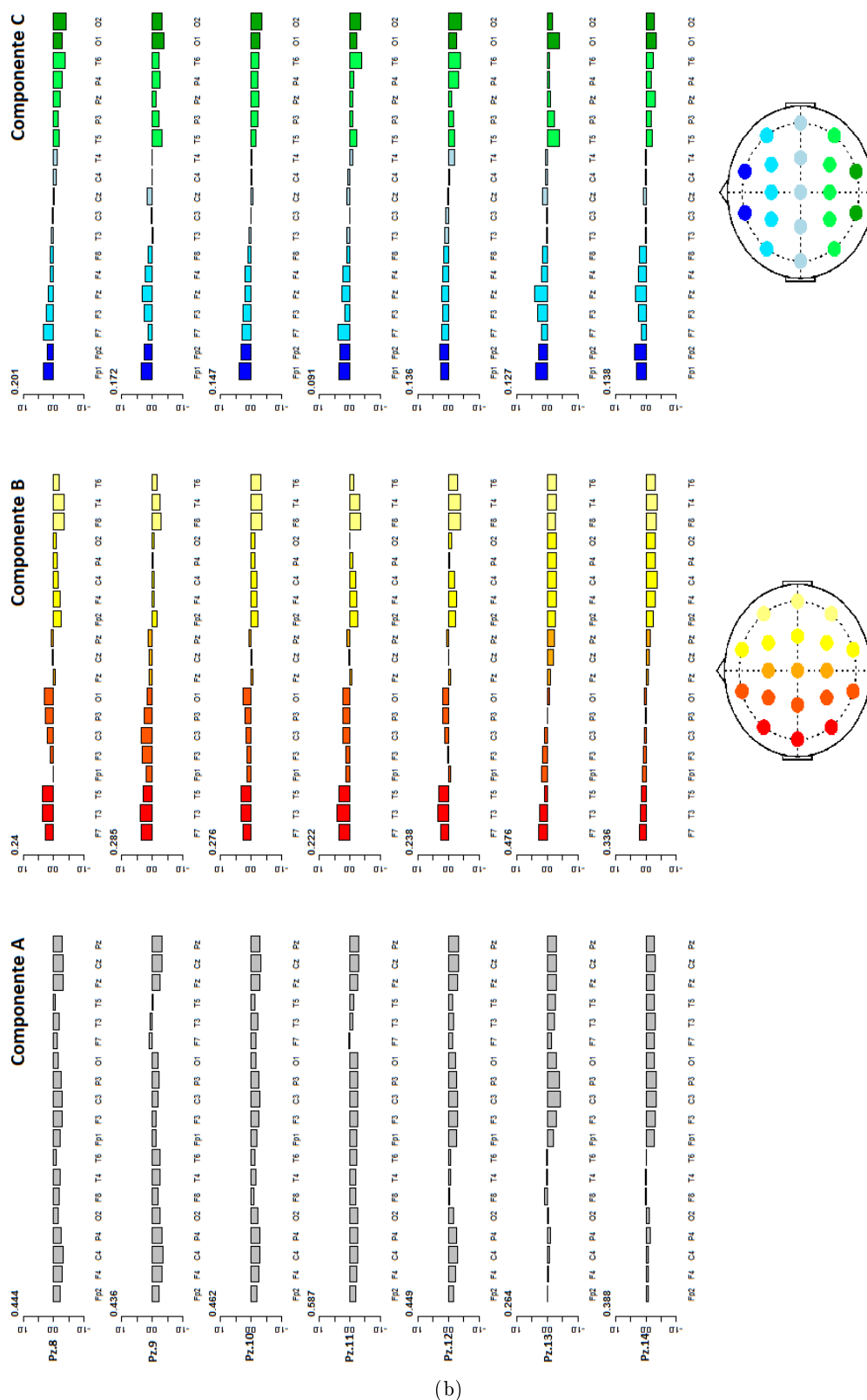


Figura A.4: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 8 a 14)

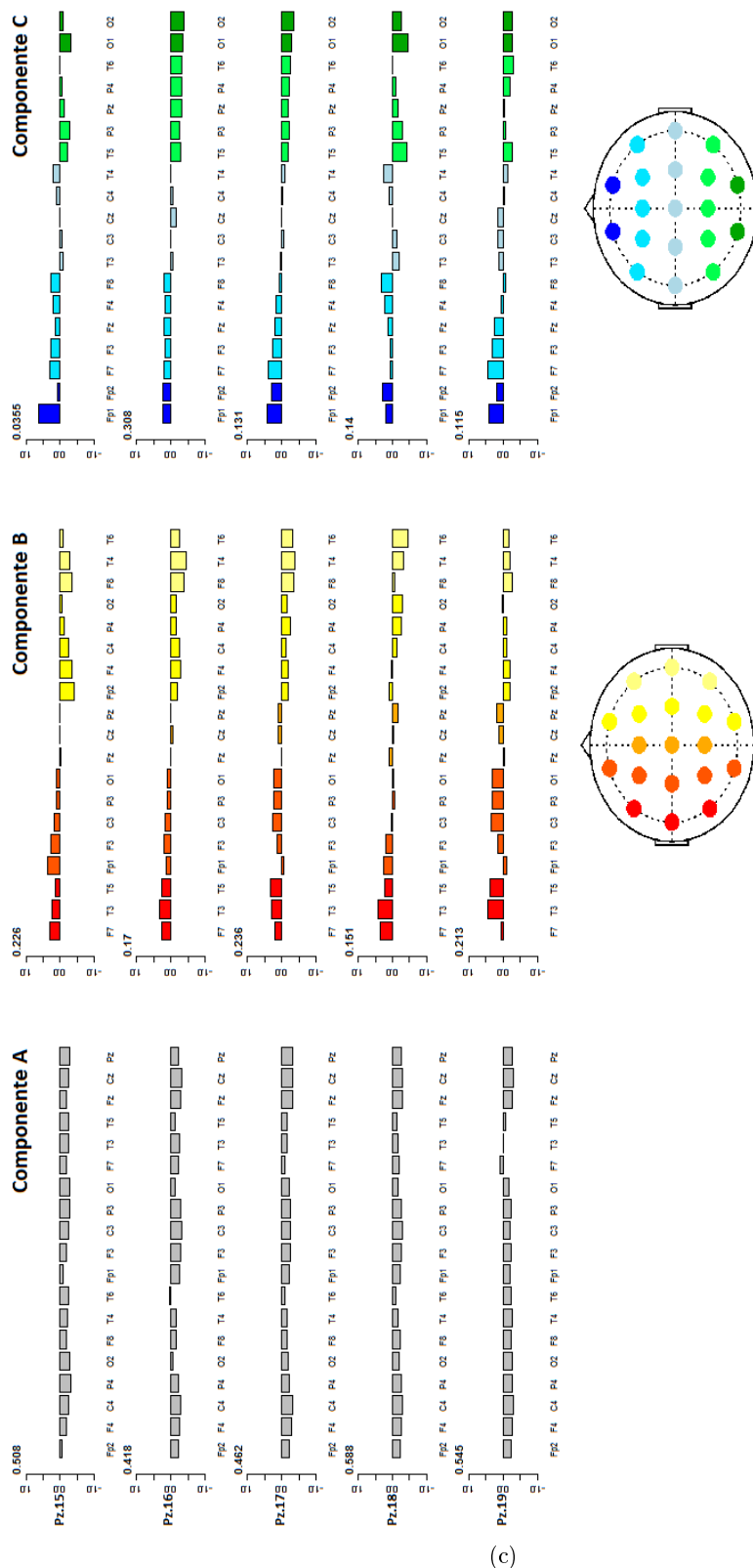


Figura A.4: Componenti A, B e C dopo il riallineamento visivo, relative agli EEG in veglia. I numeri in alto a sinistra fanno riferimento, per ogni paziente, alle quote di variabilità spiegata dalle tre componenti. I canali delle Componenti B e C sono stati riordinati per evidenziare i contrasti. (Pazienti da 15 a 19)

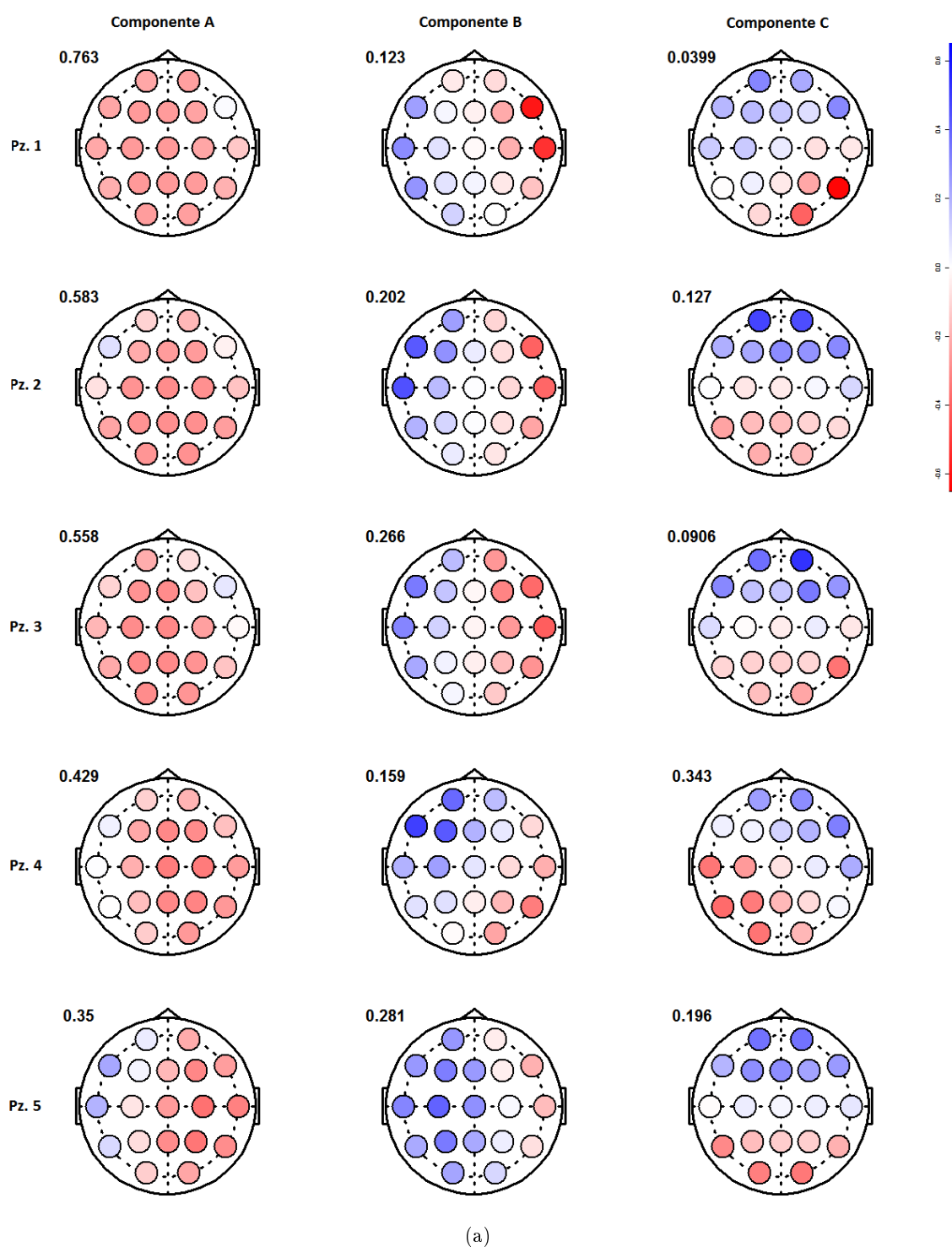
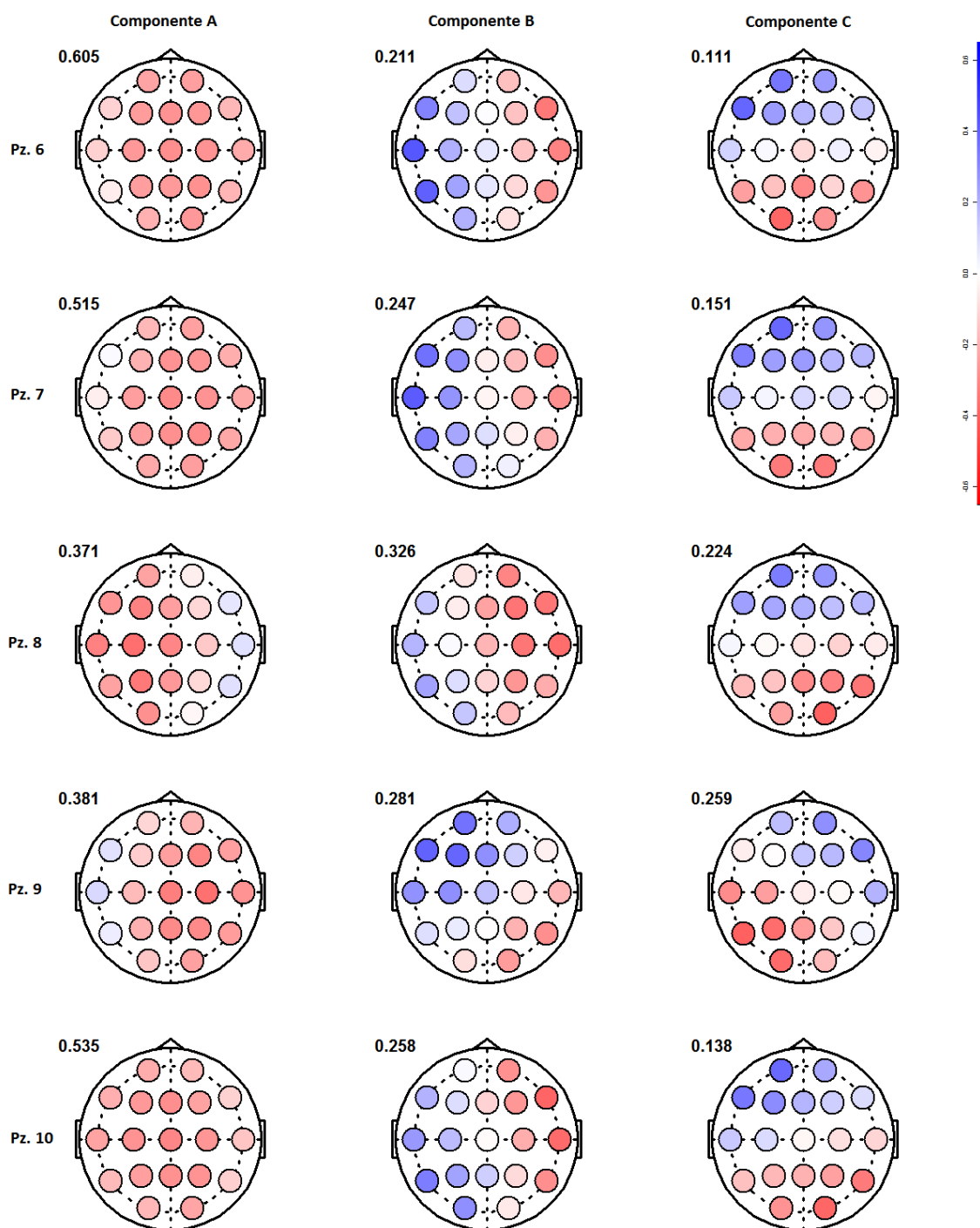


Figura A.5: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 1 a 5)



(b)

Figura A.5: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 6 a 10)



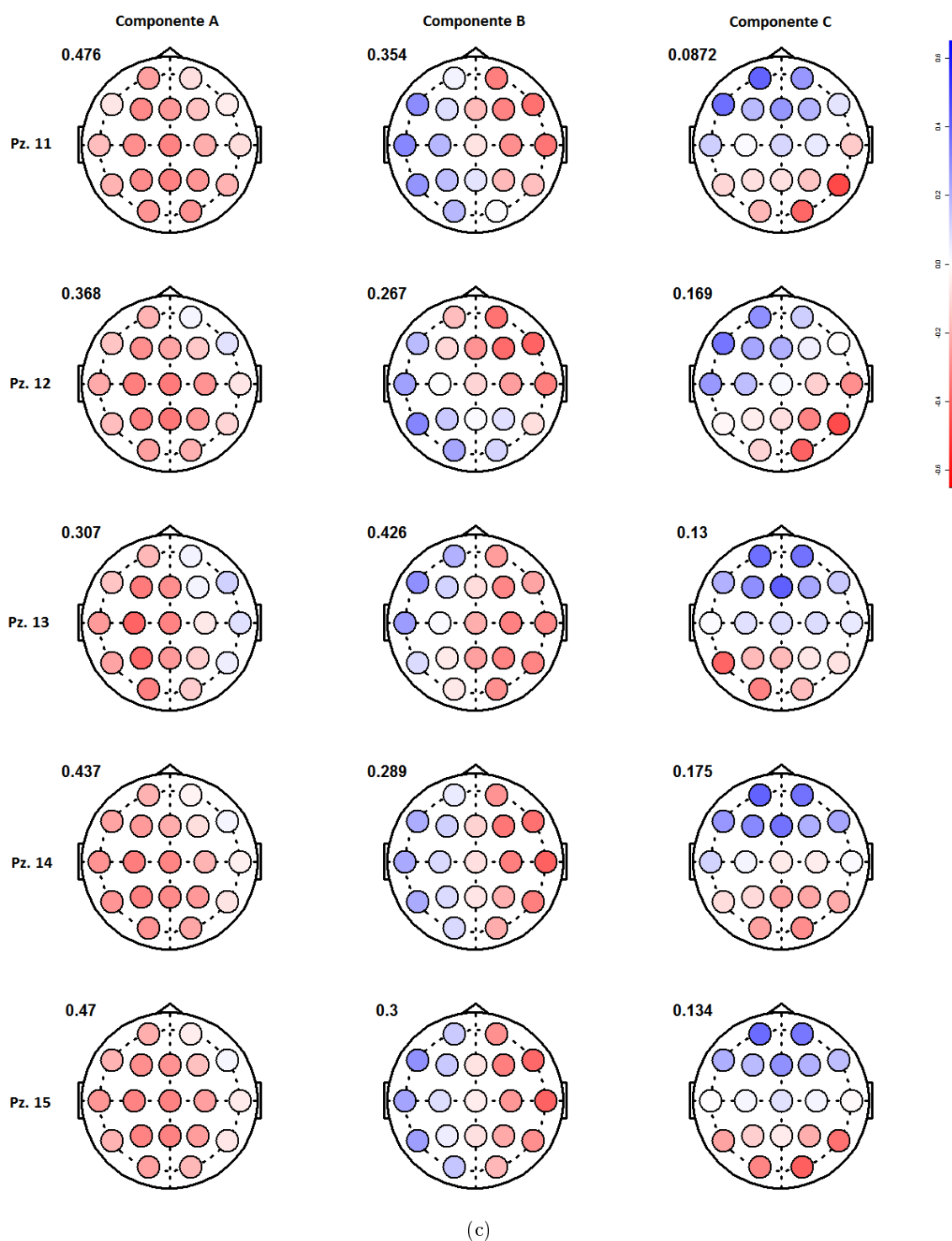


Figura A.5: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 11 a 15)

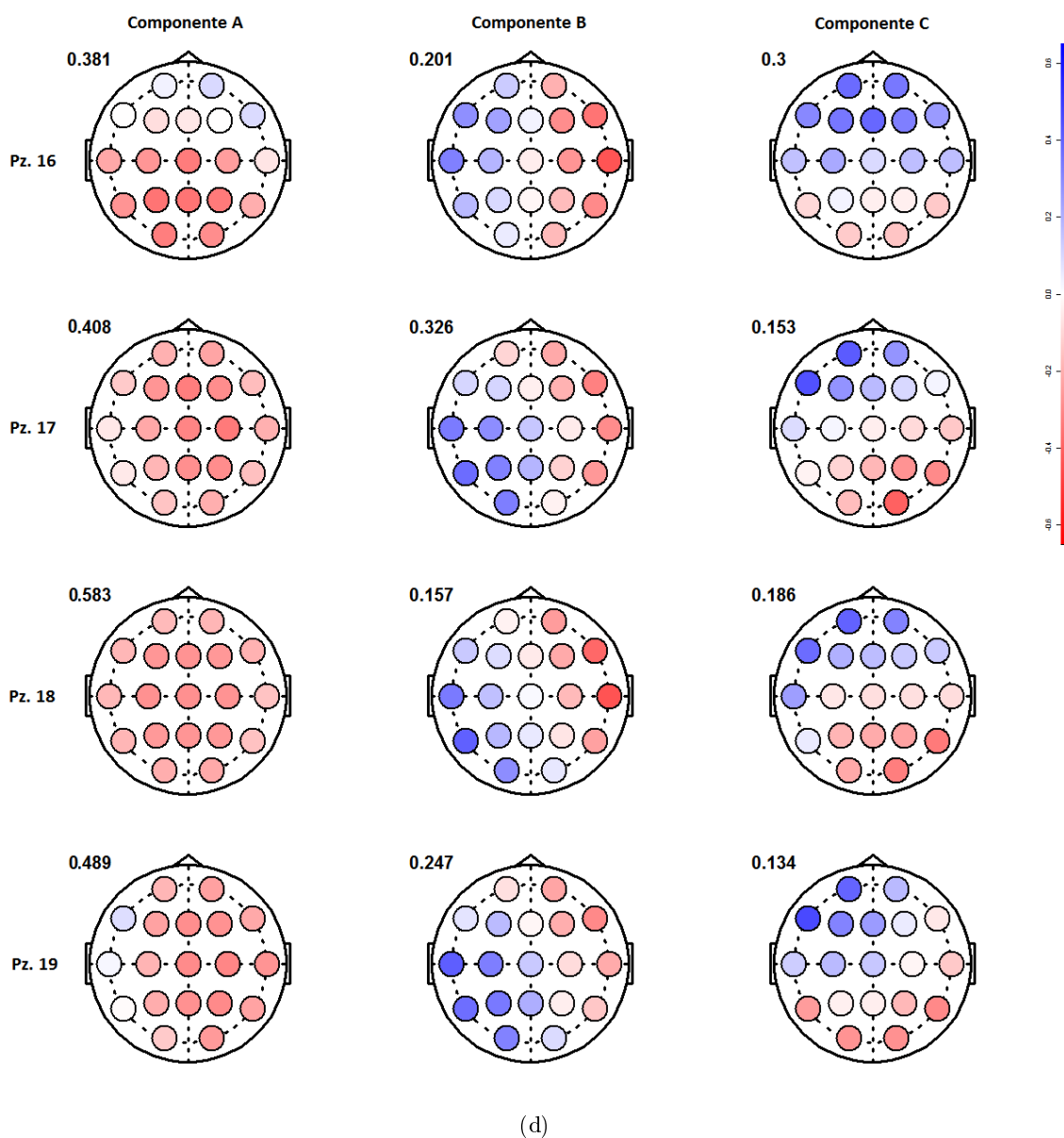


Figura A.5: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in sonno. (Pazienti da 16 a 19)

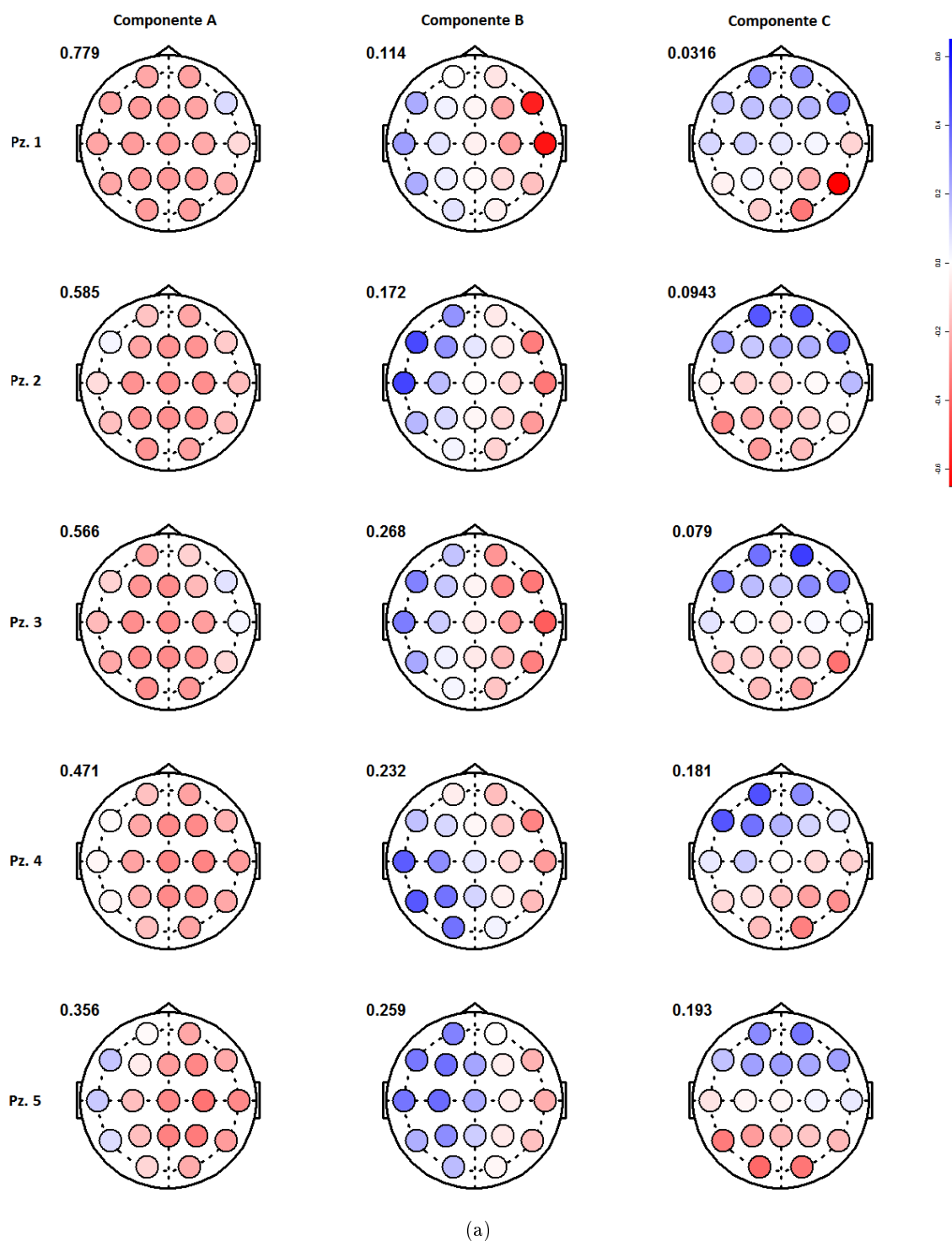


Figura A.6: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 1 a 5)

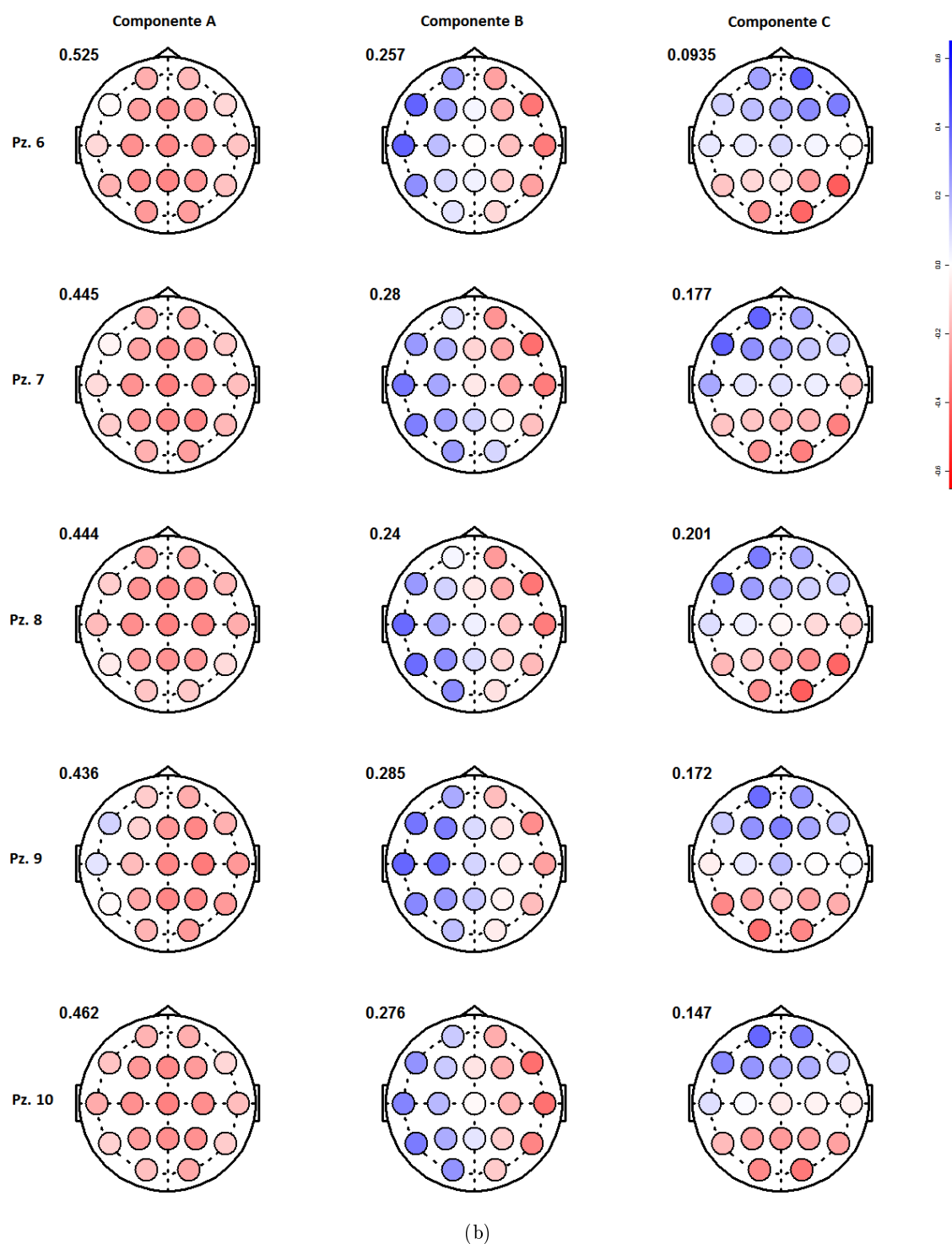


Figura A.6: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 6 a 10)

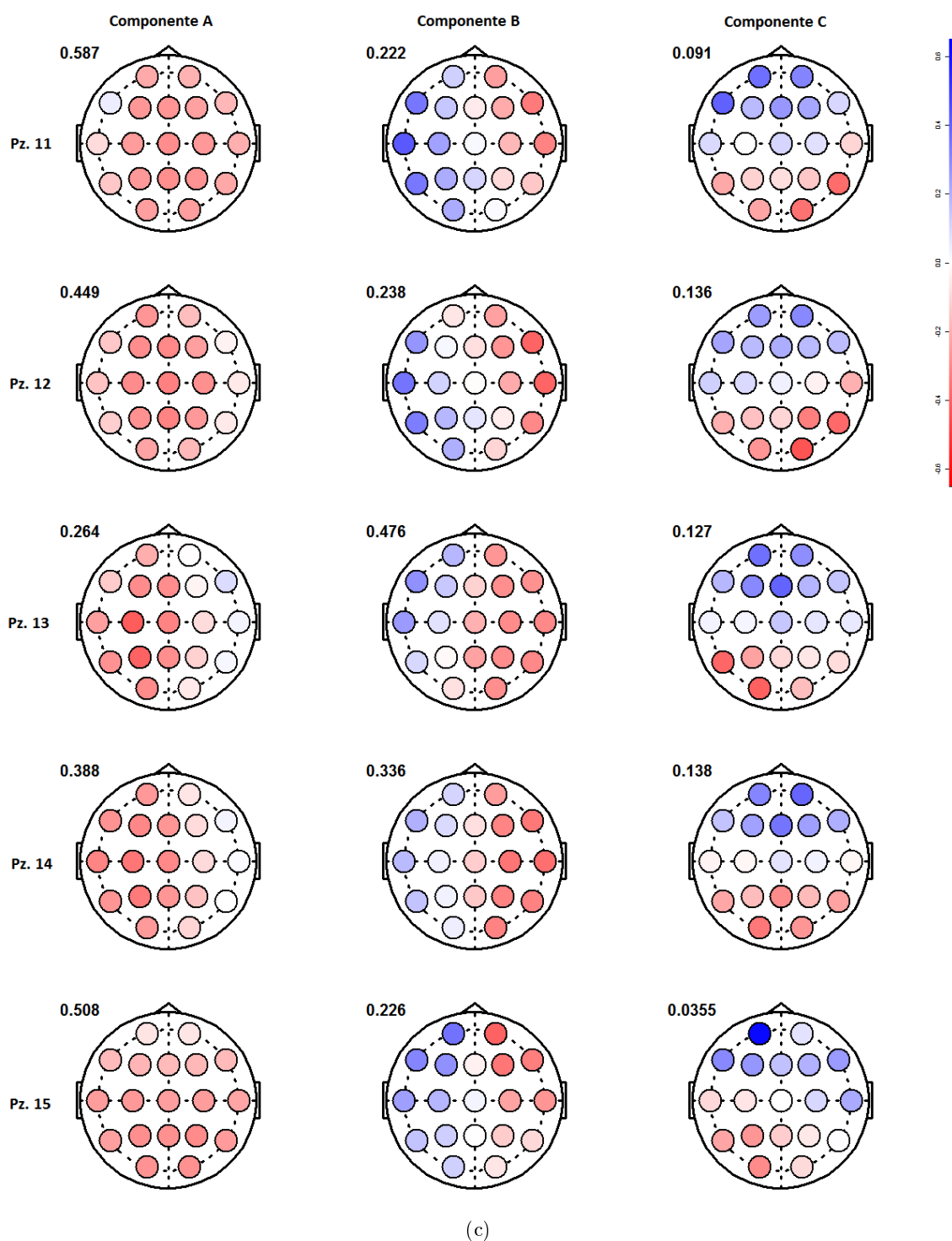


Figura A.6: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 11 a 15)

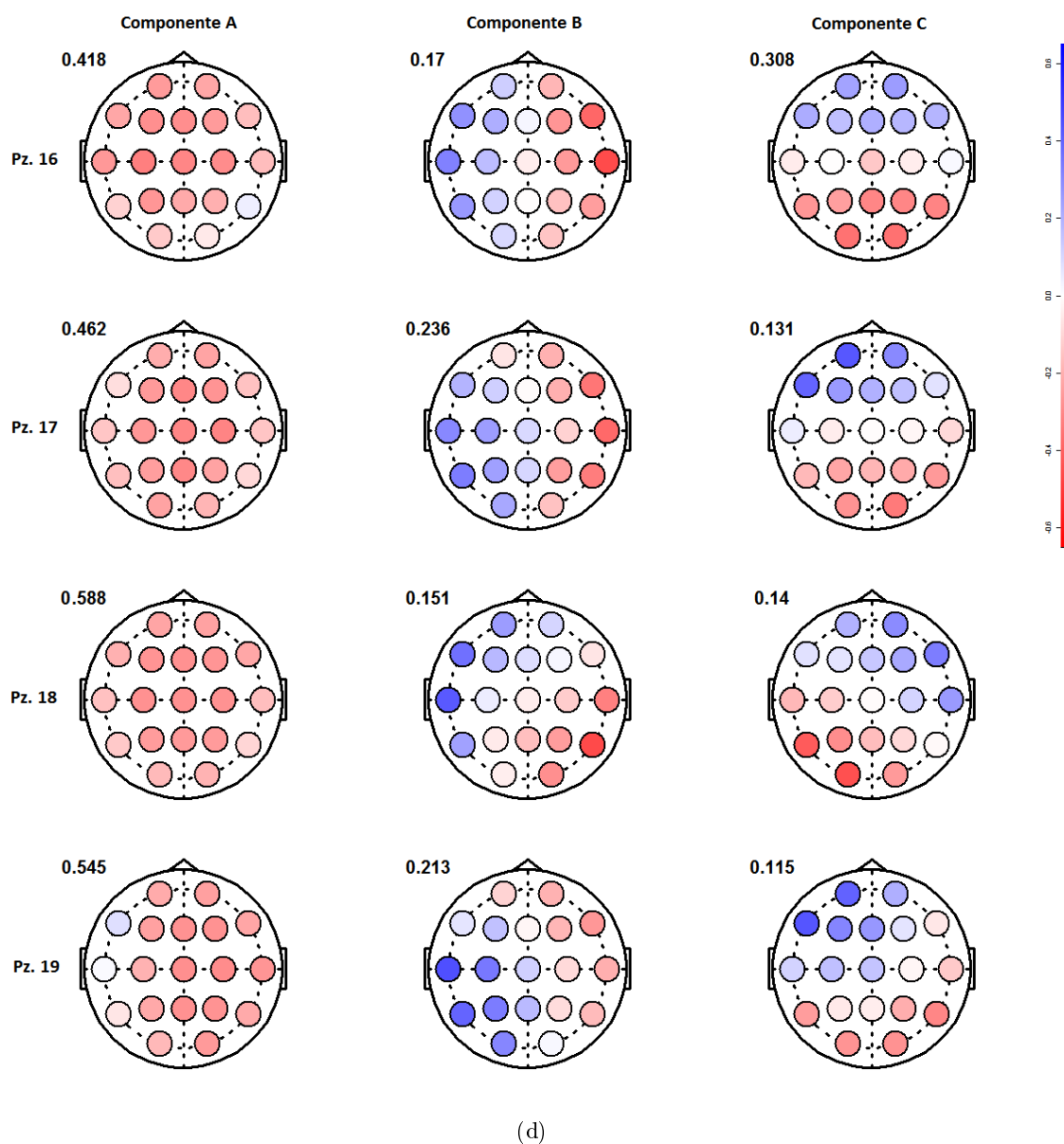
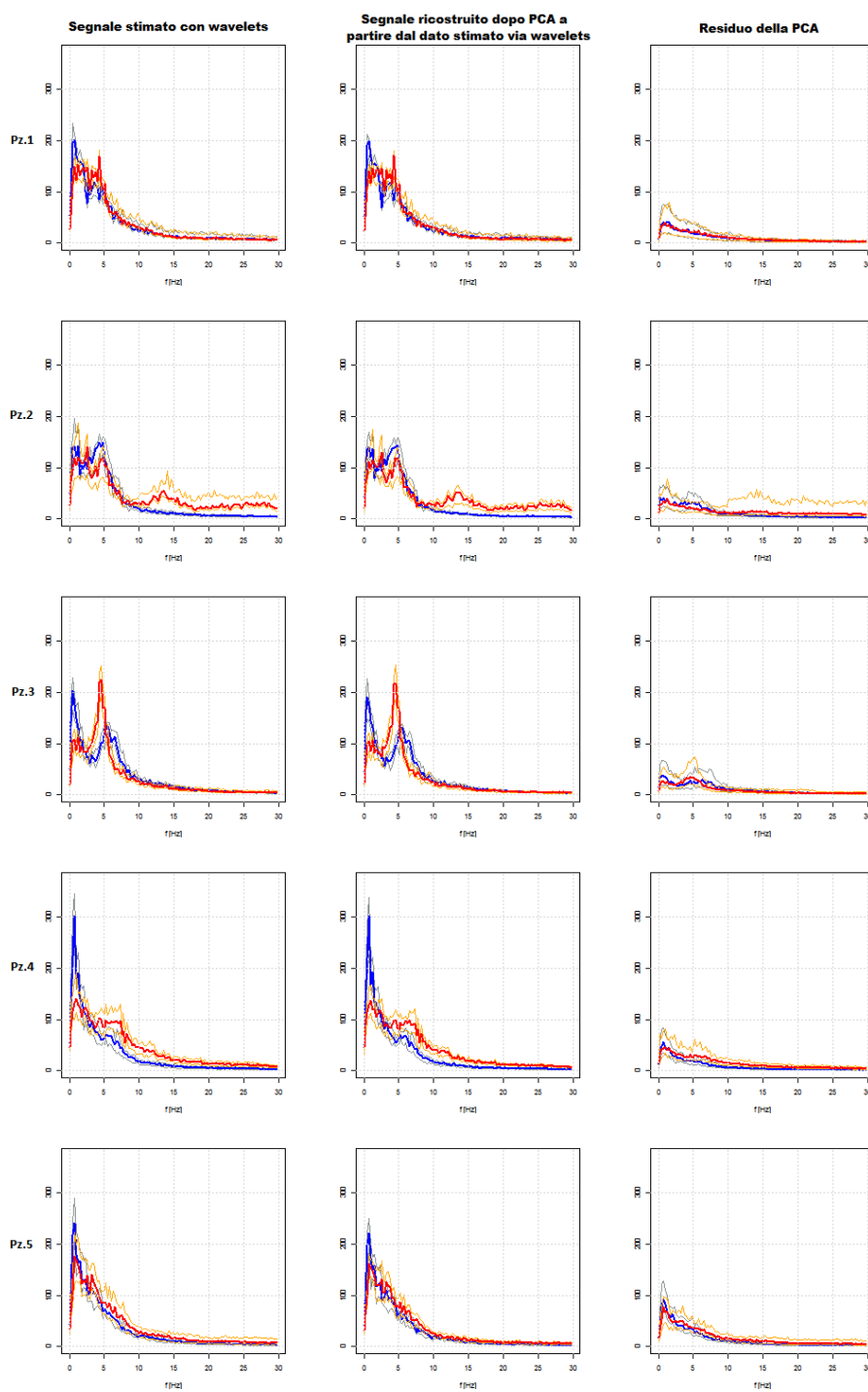
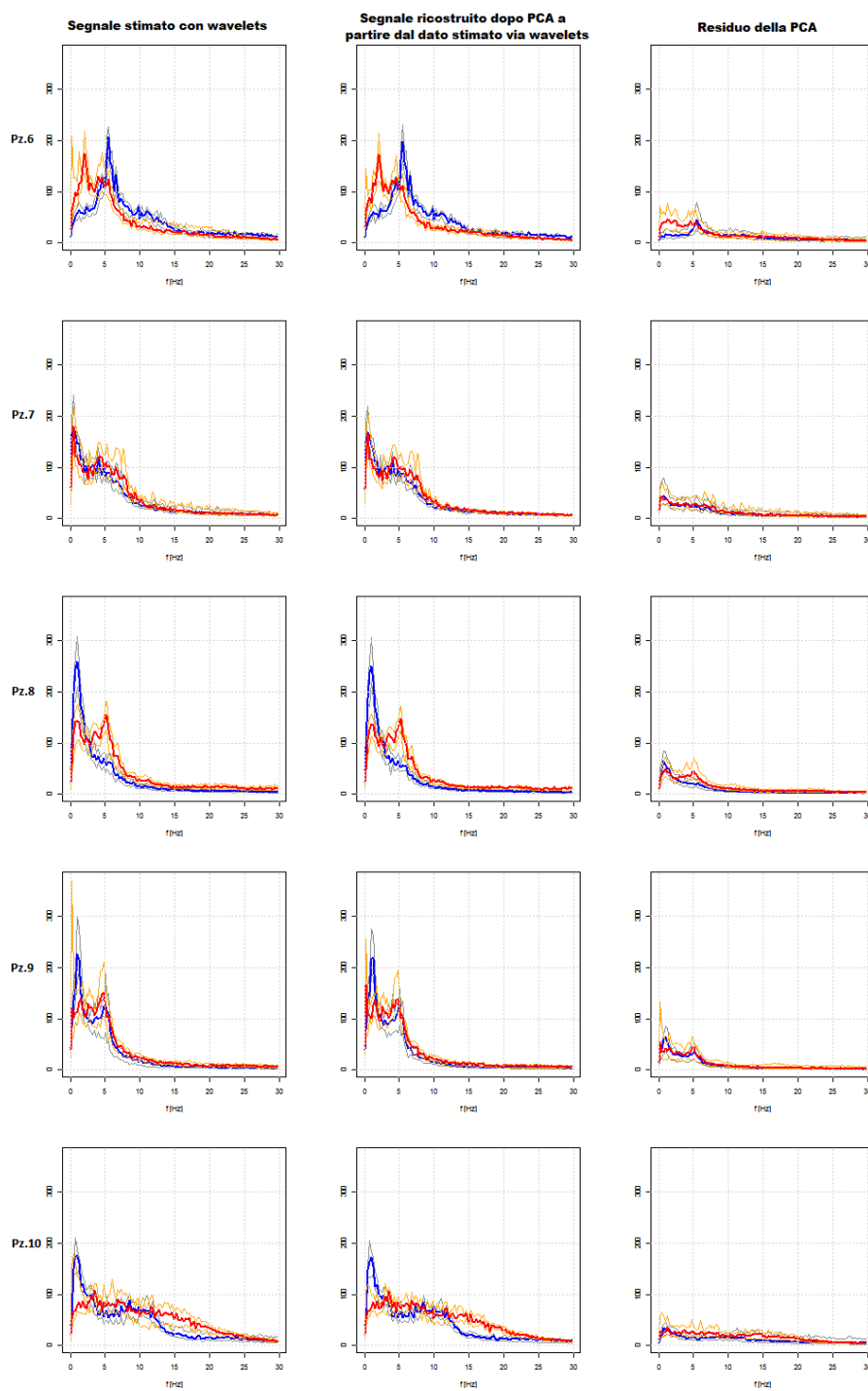


Figura A.6: I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura si riferisce agli EEG registrati in veglia. (Pazienti da 16 a 19)



(a)

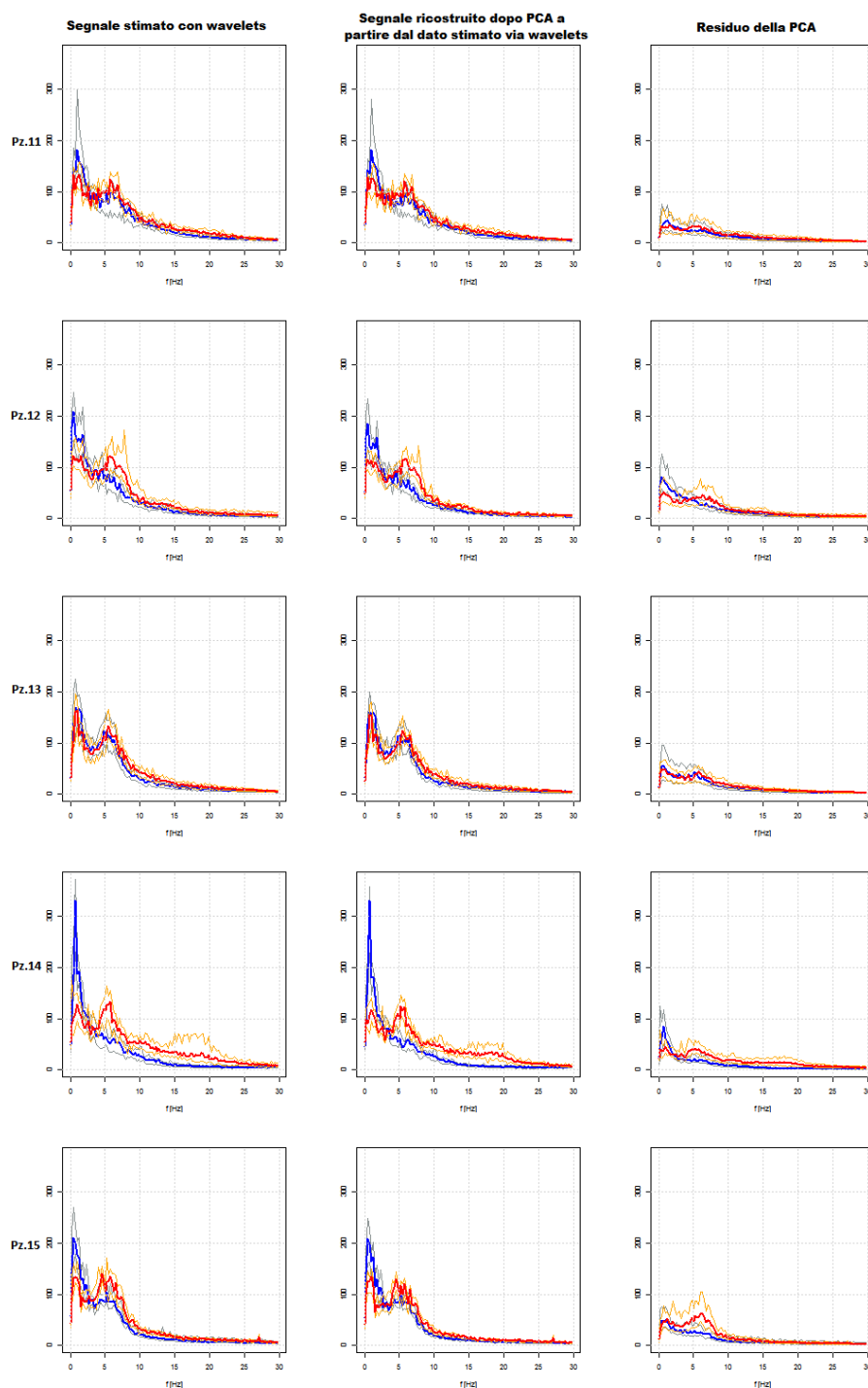
Figura A.7: Intervalli interquartile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 1 a 5)



(b)

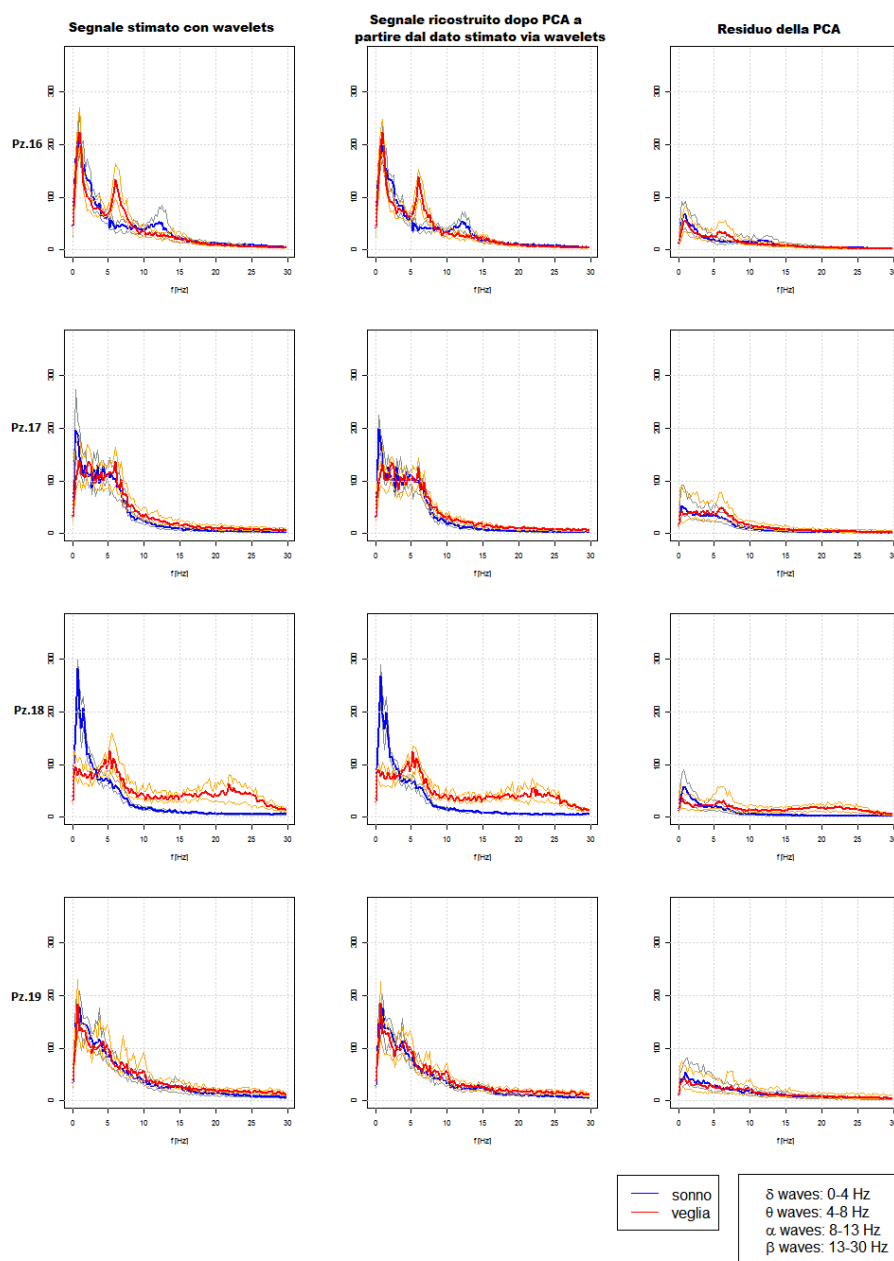
Figura A.7: Intervalli interquartile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 6 a 10)





(c)

Figura A.7: Intervalli interquartile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 11 a 15)



(d)

Figura A.7: Intervalli interquantile al 90% e mediana per gli spettri di ampiezza relativi al segnale stimato via wavelets, al segnale ricostruito dopo la PCA a partire dal dato stimato via wavelets e al segnale residuo della PCA. Gli IC in rosso-arancio si riferiscono al segnale in veglia, quelli in blu-azzurro al segnale in sonno. (Pazienti da 16 a 19)

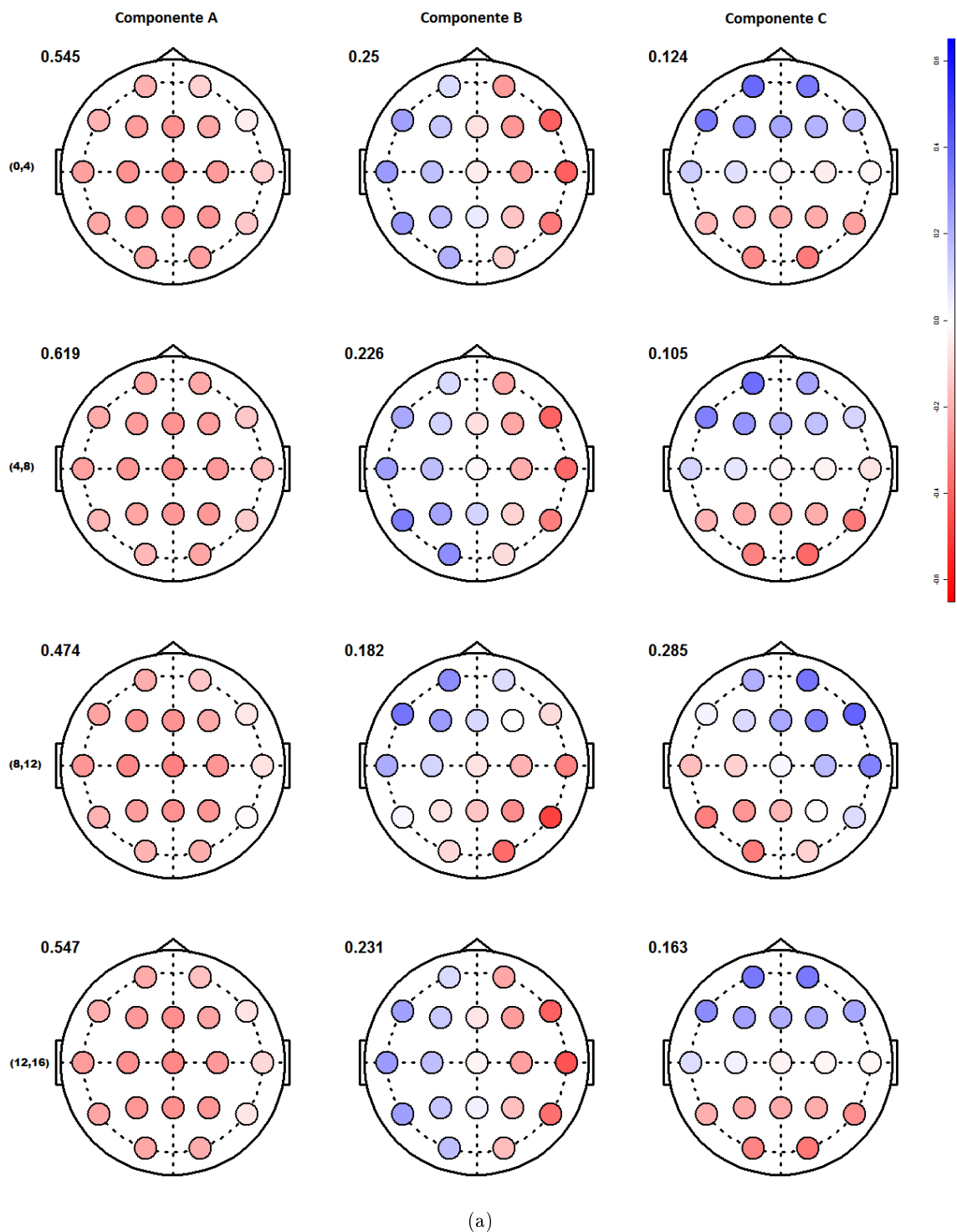


Figura A.8: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (0s,16s))

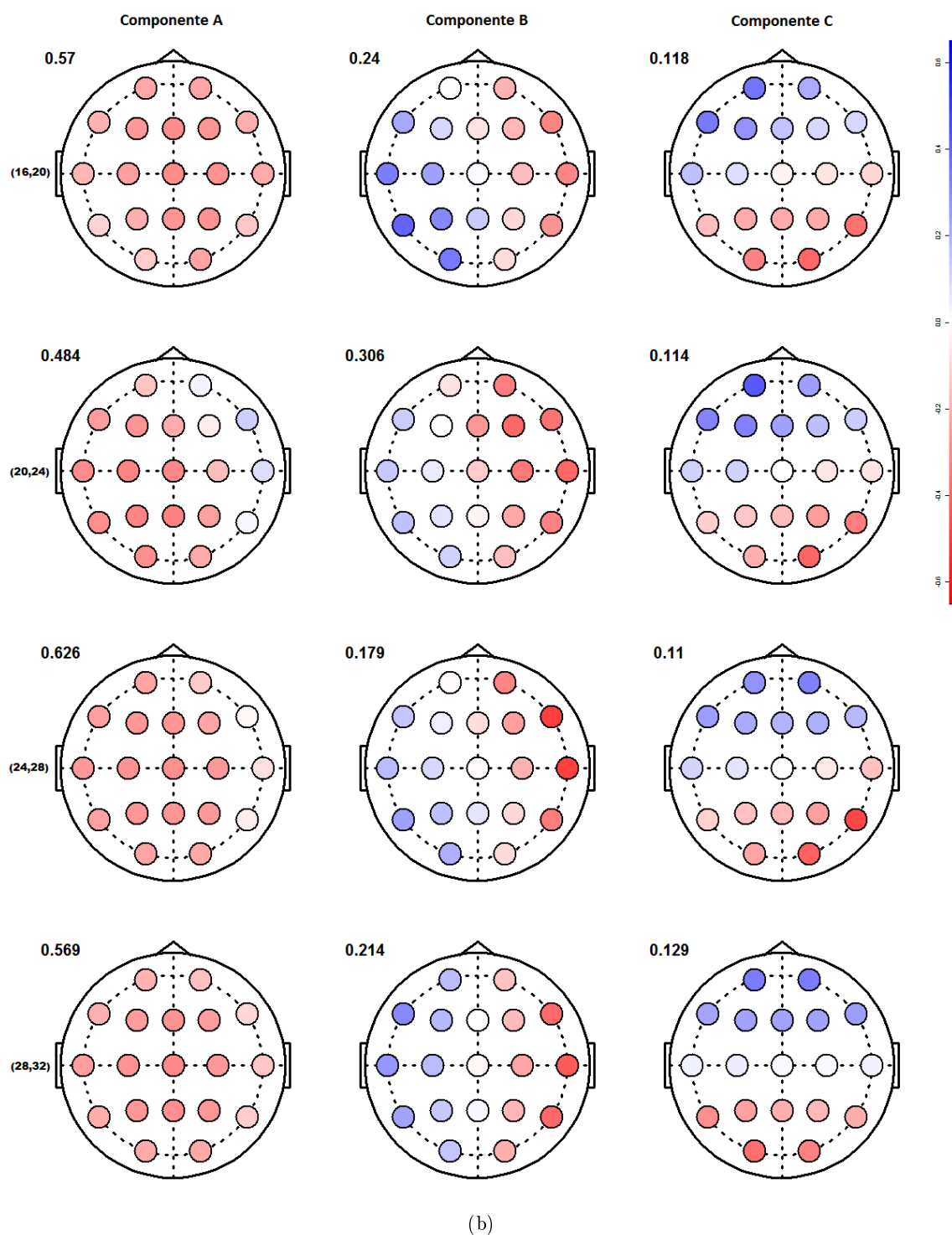


Figura A.8: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (16s,32s))

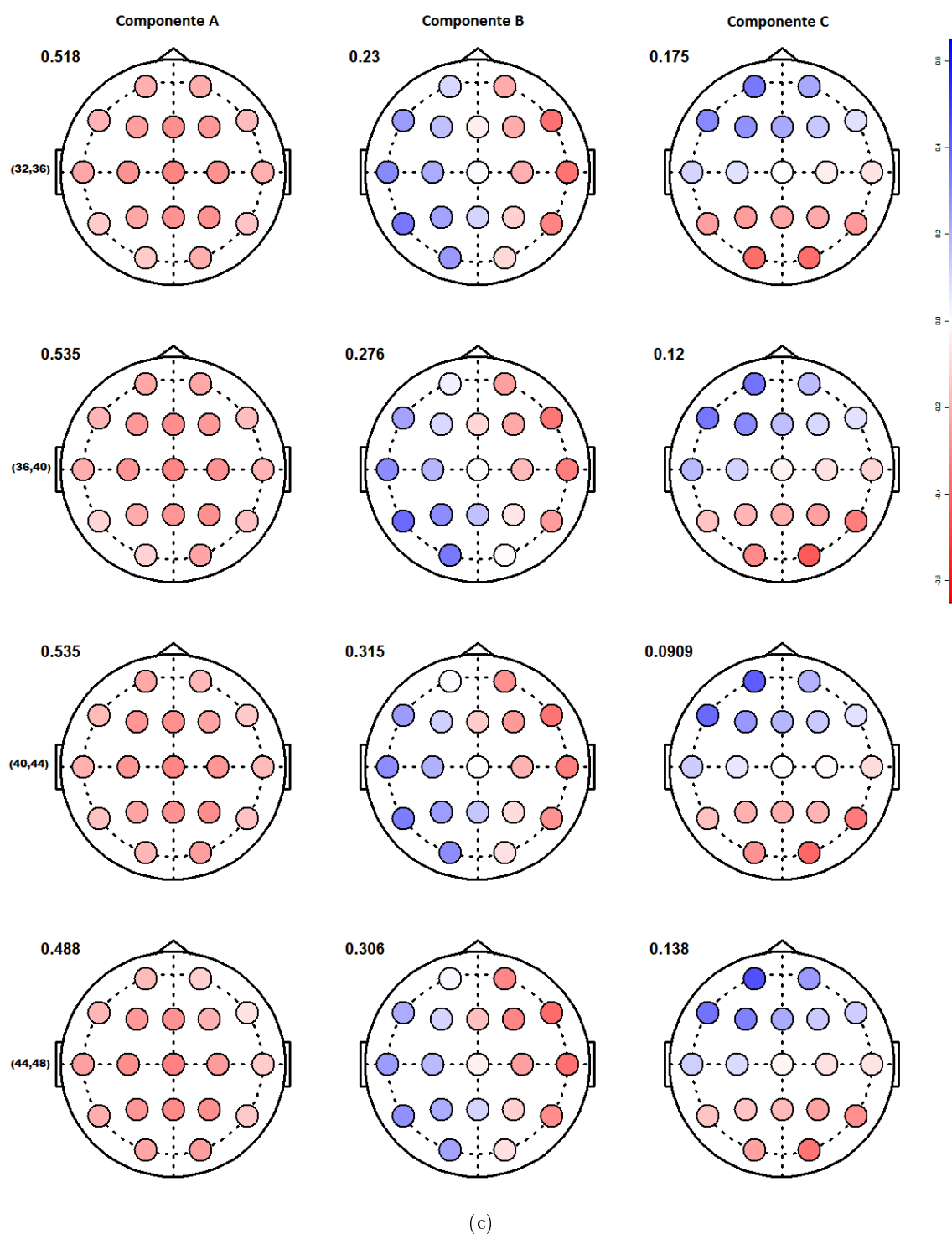


Figura A.8: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (32s,48s))

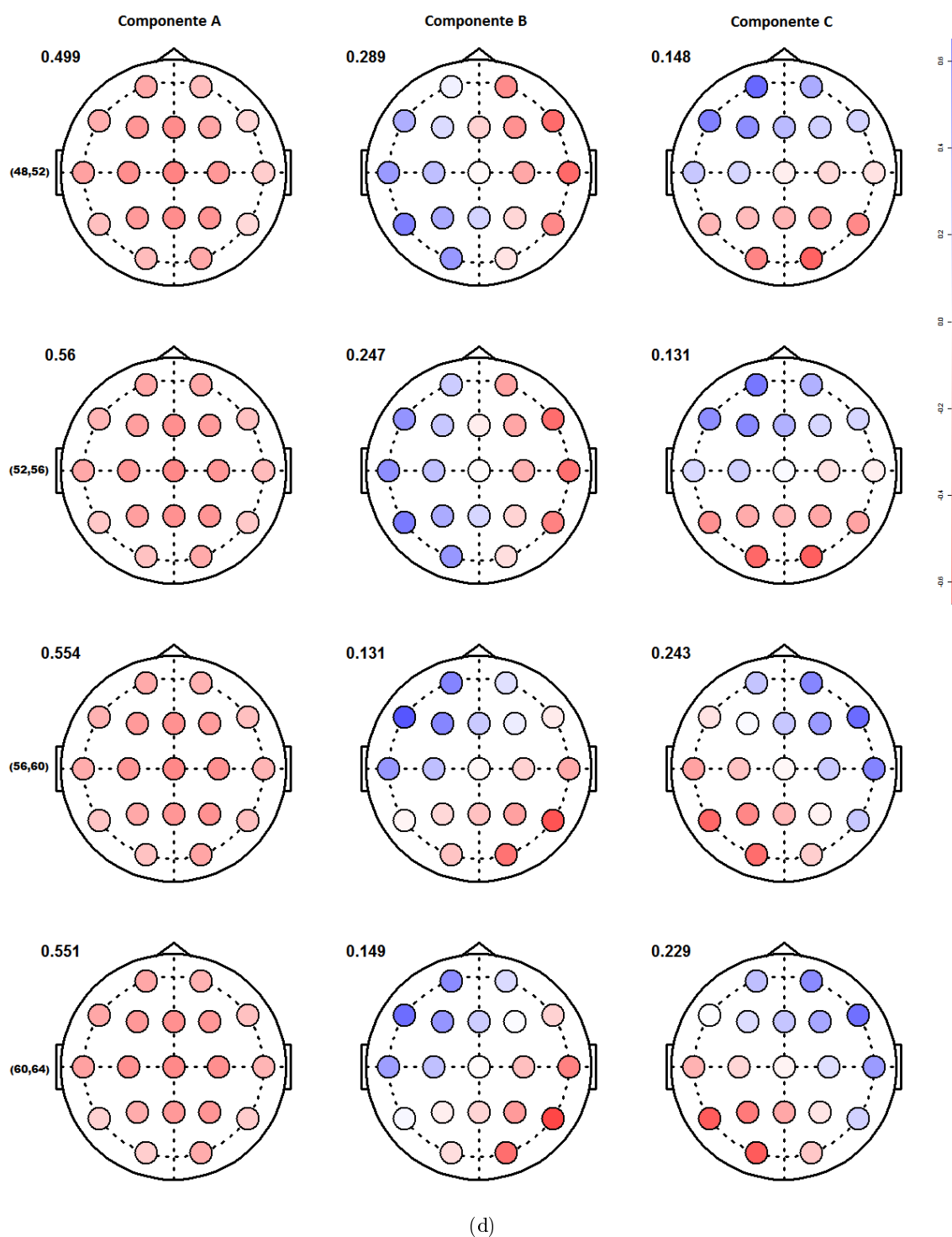


Figura A.8: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in sonno del Pz.10. (Finestre nell'intervallo (48s,64s))

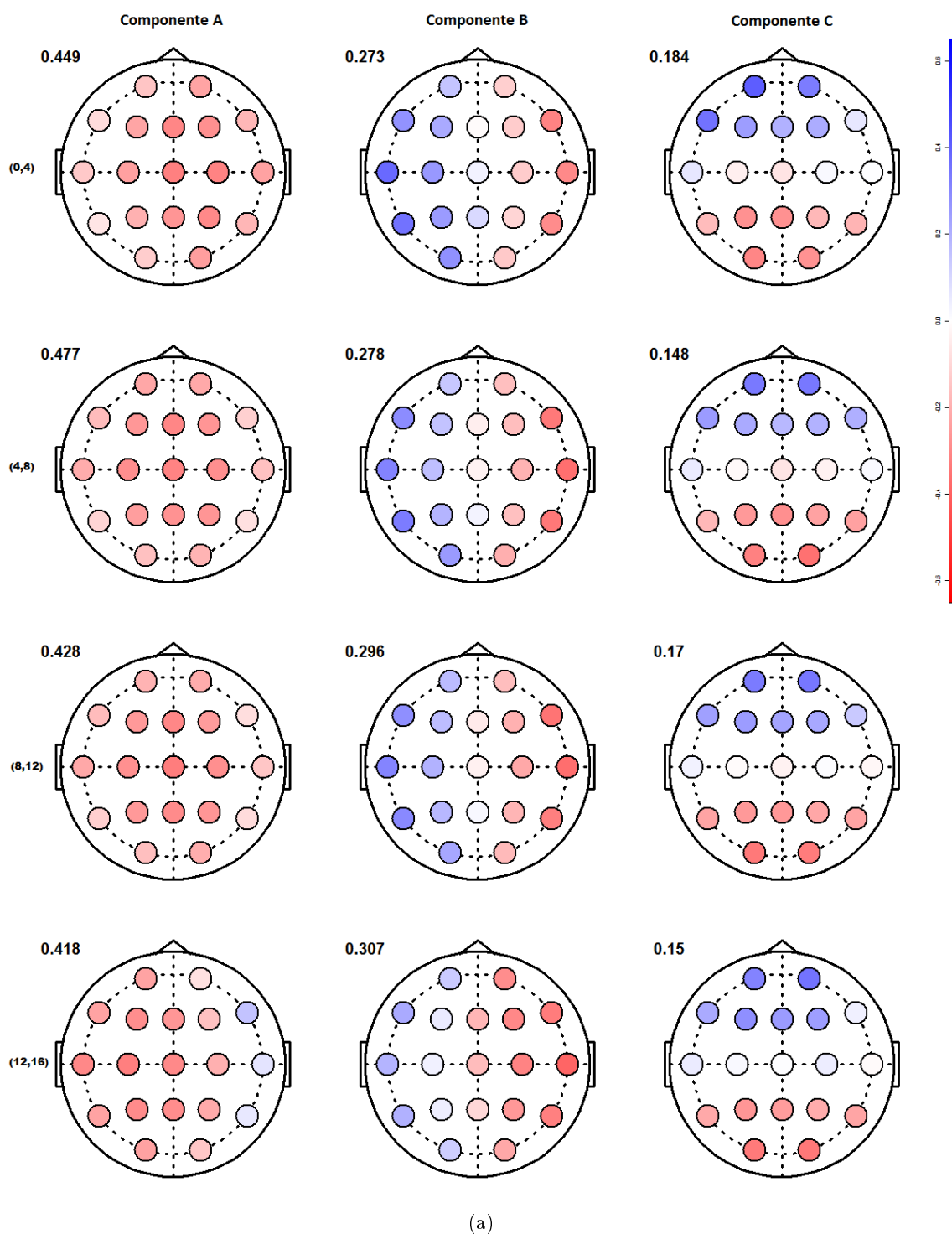


Figura A.9: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (0s,16s))

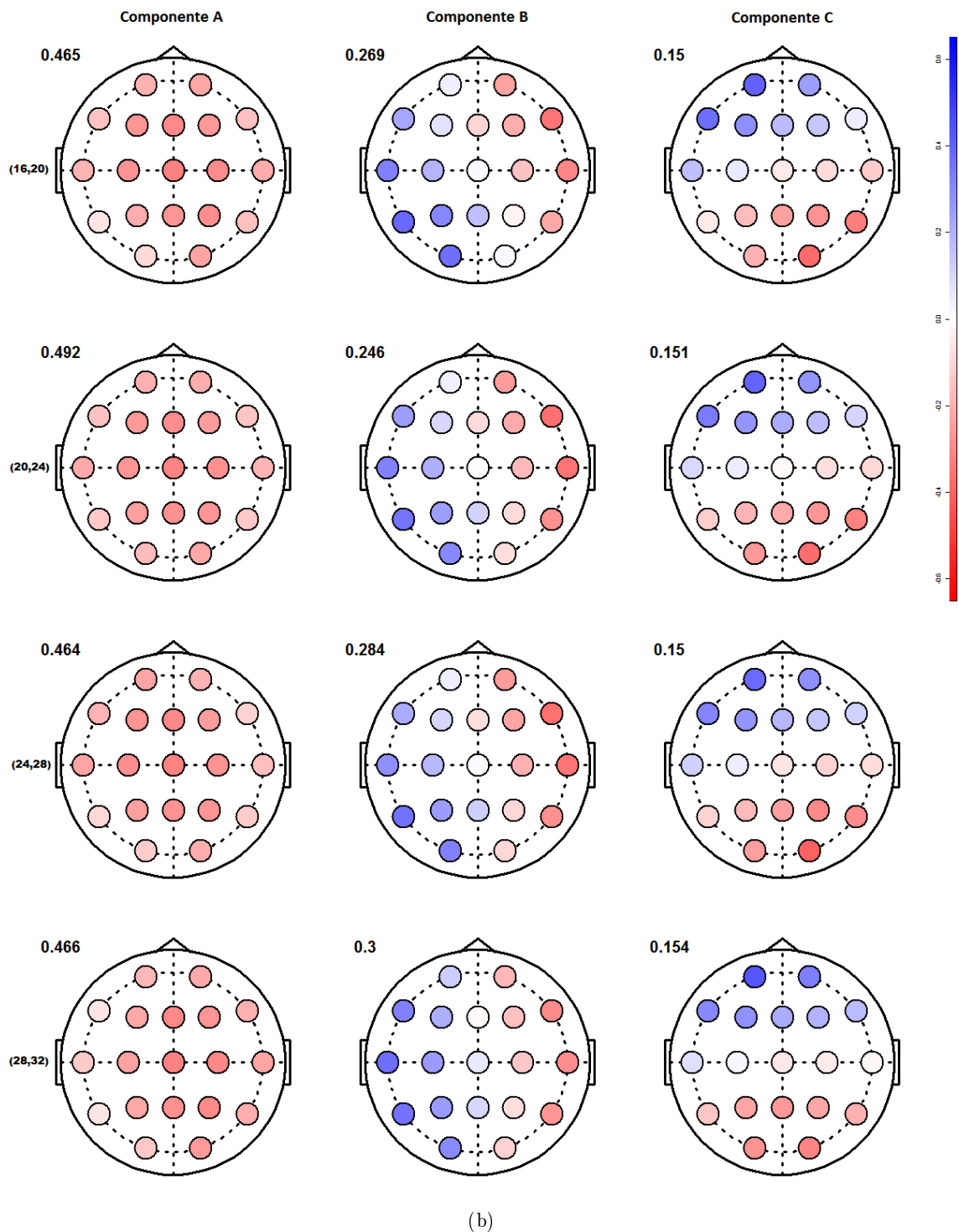


Figura A.9: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (16s,32s))



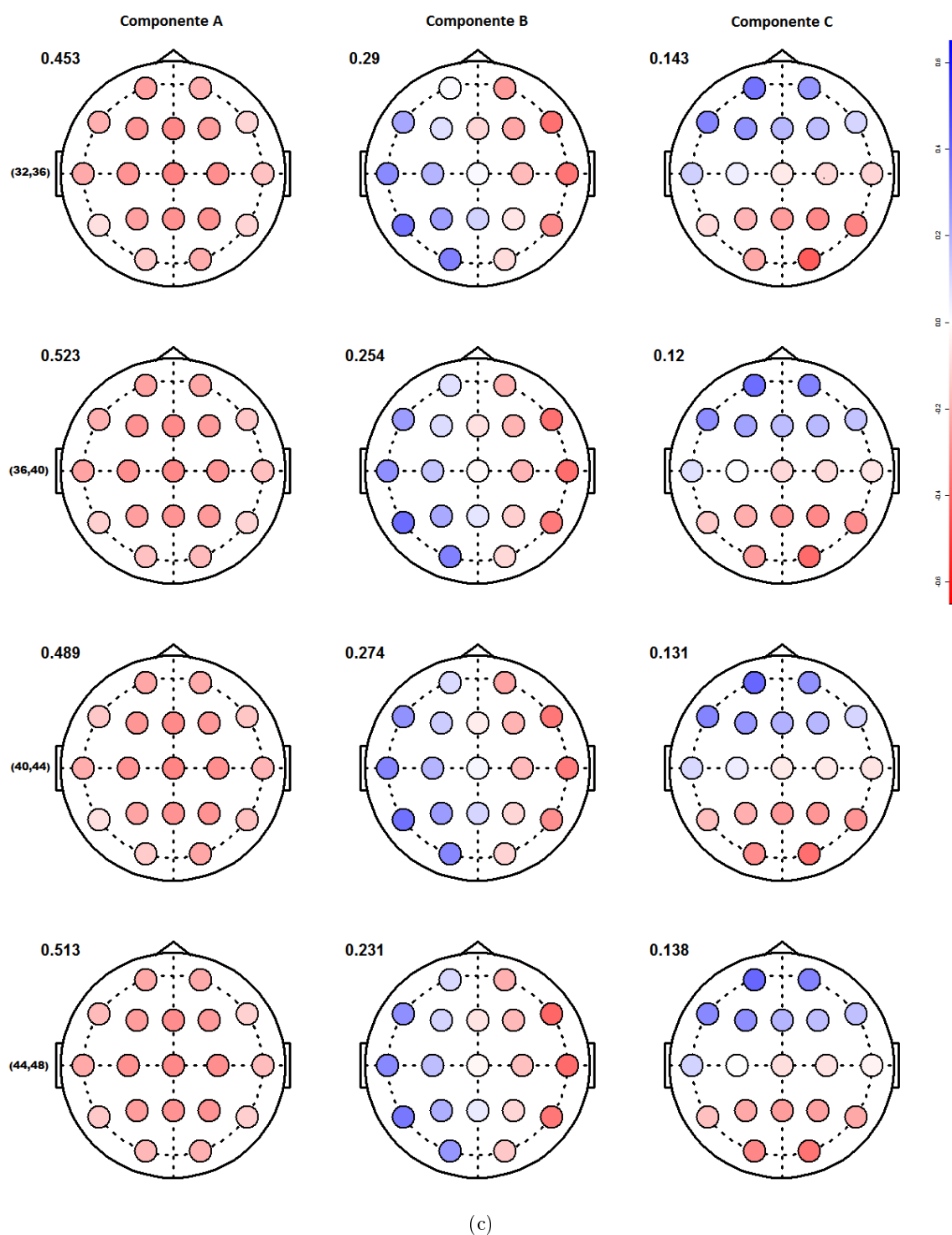


Figura A.9: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (32s,48s))

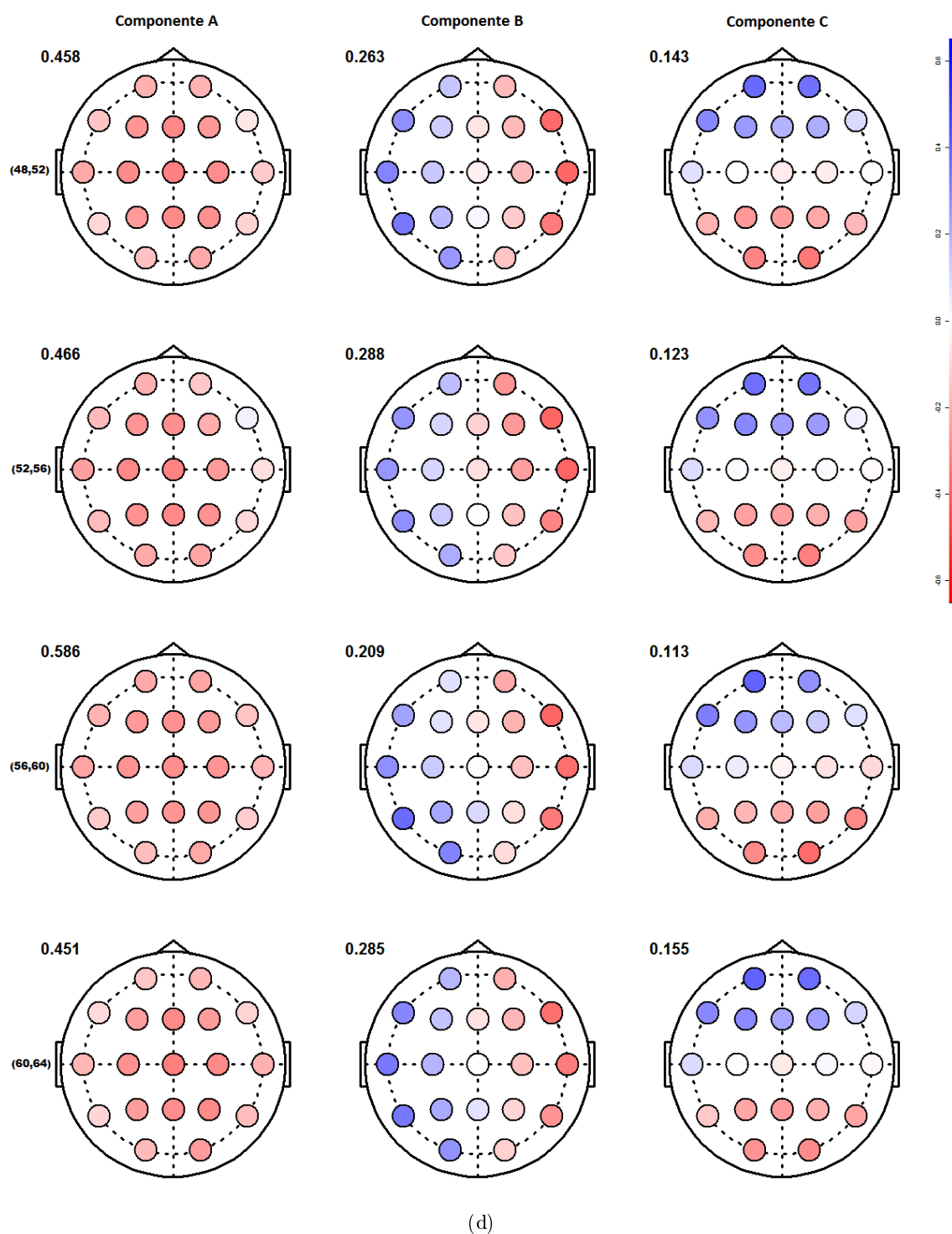


Figura A.9: Sintesi dei risultati ottenuti effettuando la PCA su ogni finestra di 4s in cui viene diviso il dato. I canali sono colorati in base al valore assunto dai loadings della rispettiva componente. La figura fa riferimento all'EEG in veglia del Pz.10. (Finestre nell'intervallo (48s,64s))

# Bibliografia

- [1] G. Bachman, L. Narici, and E. Beckenstein. *Fourier and Wavelet Analysis*. Universitext Series. Springer, 2000.
- [2] G. Beylkin, R. Coifman, and V. Rokhlin. Fast wavelet transforms and numerical algorithms I. *Communications on Pure and Applied Mathematics*, 44:141–183, 1991.
- [3] E.O. Brigham. *The fast Fourier transform and its applications*. Prentice-Hall signal processing series. Prentice Hall, 1988.
- [4] I. Daubechies. *Ten Lectures on Wavelets*. Cbms-Nsf Regional Conference Series in Applied Mathematics. Society for Industrial and Applied Mathematics, 1992.
- [5] D.L. Donoho and I.M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90:1200–1224, 1995.
- [6] Stanley M. Dunn, A. Constantinides, and Prabhas V. Moghe. *Numerical Methods in Biomedical Engineering*. Academic Press, 2006.
- [7] Brian S. Everitt. *An R and S-Plus Companion to Multivariate Analysis*. Springer Texts in Statistics, 2005.
- [8] B.S. Everitt, S. Landau, M. Leese, and D. Stahl. *Cluster Analysis*. Wiley Series in Probability and Statistics. Wiley, 2011.

- 
- [9] J.T. Giacino, S. Ashwal, N. Childs, R. Cranford, B. Jennett, D.I. Katz, J.P. Kelly, J.H. Rosenberg, J. White, R.D. Zafonte, and N.D. Zasler. The minimally conscious state: definition and diagnostic criteria. *Neurology*, 58:349–353, 2002.
- [10] J.T. Giacino, K. Kalmar, and J. Whyte. The JFK Coma Recovery Scale-Revised: measurement characteristics and diagnostic utility. *Arch Phys Med Rehabil*, 85(12):2020–2029, 2004.
- [11] Gianluigi Gigli, Antonio Carolei, Paolo Maria Rossini, and Rachele Zylberman (testo redatto a cura di). *Stato vegetativo e di minima coscienza: epidemiologia, evidenze scientifiche e modelli assistenziali*. Ministero della Salute - Gruppo di lavoro istituito con D.M. 15 ottobre 2008 e presieduto dal Sottosegretario di Stato On. Eugenia Roccella. Documento Finale (4 dicembre 2009). [http://www.salute.gov.it/imgs/C\\_17\\_pubblicazioni\\_1378\\_allegato.pdf](http://www.salute.gov.it/imgs/C_17_pubblicazioni_1378_allegato.pdf).
- [12] Olivia Gosseries, Audrey Vanhaudenhuyse, Marie-Aurélié Bruno, Athena Demertzi, Caroline Schnakers, MélanieM. Boly, Audrey Maudoux, Gustave Moonen, and Steven Laureys. Disorders of consciousness: Coma, vegetative and minimally conscious states. In D. Cvetkovic and I. Cosic, editors, *States of Consciousness*, The Frontiers Collection, pages 29–55. Springer-Verlag Berlin Heidelberg, 2011.
- [13] J.A. Hartigan. *Clustering algorithms*. Wiley series in probability and mathematical statistics: Applied probability and statistics. Wiley, 1975.
- [14] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: Data Mining, Inference, and Prediction, 2nd Ed*. Springer, New York, 2009.
- [15] R.A. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis. 6th Ed*. Pearson Prentice Hall, 2007.
- [16] I.T. Jolliffe. *Principal Component Analysis*. Springer Series in Statistics. Springer, 2002.

- [17] Eric Landsness, Marie-Aur lie Bruno, Quentin Noirhomme, Brady Riedner, Olivia Gosseries, Caroline Schnakers, Marcello Massimini, Steven Laureys, Giulio Tononi, and M lanie Boly. Electrophysiological correlates of behavioural changes in vigilance in vegetative state and minimally conscious state. *Brain*, 134(8):2222–2232, August 2011.
- [18] Steven Laureys, Adrian M. Owen, and Nicholas D. Schiff. Brain function in coma, vegetative state, and related disorders. *The Lancet Neurology*, 3:537–546, September 2004.
- [19] Martin M. Monti, Steven Laureys, and Adrian M. Owen. The vegetative state. *BMJ*, 341:c3765, 8 2010.
- [20] G.P. Nason. *Wavelet Methods in Statistics with R*. Use R! Springer, 2008.
- [21] D. Pigoli. *Stima di curve tridimensionali tramite wavelets, con un’applicazione alla geometria dell’arteria carotide interna*. Tesi di Laurea Specialistica in Ingegneria Matematica, 2009. <http://www1.mate.polimi.it/biblioteca/tesiview.php?id=291&L=i>.
- [22] D. Pigoli and L. M. Sangalli. Wavelets in functional data analysis: Estimation of multidimensional curves and their derivatives. *Computational Statistics and Data Analysis*, 56:1482–1498, 2012.
- [23] A. Quarteroni, R. Sacco, and F. Saleri. *Numerical mathematics, 2nd Edition*. Springer, Berlin, 2007.
- [24] J. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2005.
- [25] J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer, 2002.
- [26] J. Walker. *Fast Fourier Transforms*. CRC Press, Boca Raton, 1991.
- [27] J.G. Webster. *Medical Instrumentation: Application and Design. 4th Ed.* John Wiley & Sons, 2009.

# Ringraziamenti

Un sincero e sentito grazie a Laura e Simone per tutto il tempo dedicatomi, i loro preziosi consigli e per avermi sapientemente guidato e incoraggiato durante tutto il lavoro.

Desidero inoltre ringraziare l'Unità di Ingegneria Biomedica dell'Istituto Neurologico "Carlo Besta" di Milano, e in modo particolare il Dott. Ing. Ferruccio Panzica, Fabio, Giulia e Patrik, per la disponibilità nel fornirmi il materiale richiesto e per tutto il loro aiuto.

Ringrazio tutti compagni di studio e di pausa pranzo che ho conosciuto e che hanno reso ancora più piacevoli questi anni.

Infine un particolare ringraziamento alla mia famiglia per il costante appoggio che mi ha dato.

Marco