Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

TESI DI LAUREA

Sintesi di segnali vocali e classificazione pesata di pazienti affetti da Parkinson

Candidato:
Francesco Gammaraccio

Relatore:
Prof. Alessio Carullo

Correlatore:
Prof. Alberto Vallan

Correlatore:
Ph.D. Alessio Atzori

Anno Accademico 2020-2021

A mia madre e a mio padre.

Abstract

L'elaborazione del segnale vocale permette di ottenere importanti informazioni che possono essere utilizzate per la diagnosi di patologie legate all'apparato fonatorio, come la malattia di Parkinson. Quest'ultima è una malattia neurodegenerativa che provoca prevalentemente sintomi motori, che possono influenzare anche le capacità comunicative del paziente. Presso il DET del Politecnico di Torino si sta sviluppando un progetto che ha come obiettivo la realizzazione di un dispositivo indossabile che fornirà informazioni preziose per il medico, sia per una diagnosi precoce sia per valutare l'efficacia della terapia nei soggetti affetti dalla malattia di Parkinson. In questo lavoro di tesi è proposto un algoritmo di classificazione pesato basato sulla stima dell'errore dell'algoritmo di estrazione delle feature che potrebbe essere utilizzato nel progetto.
I segnali vocali forniti sono stati registrati durante la ripetizione di 3 vocali sostenute /a/ da parte di 57 soggetti sani (HE), da 67 pazienti con patologie del tratto vocale e non affetti da Parkinson (PA) e da 45 pazienti con la malattia di Parkinson (PD). Dato che il dataset fornito è risultato essere sbilanciato per quanto riguarda il genere e l'età dei soggetti (i pazienti PD hanno un'età media che è più del doppio rispetto ai soggetti HE), è stato deciso di selezionare 10 soggetti da ogni classe per ottenere un dataset bilanciato nell'età e nel genere.
L'elaborazione del segnale vocale ha permesso di estrarre parametri per valutare la stabilità in frequenza e in ampiezza, rispettivamente jitter e shimmer, e parametri per quantificare la qualità del segnale, come HNR e CPPS.
Tramite una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings sono state generate sequenze di periodi, ampiezze e CPPS noti che sono state utilizzate per la produzione di vocali artificiali seguendo il metodo dello pseudo-periodo candidato, il metodo del "liftering", il metodo di ricampionamento dei periodi originali ed il metodo di trasformazione del cepstrum. Per ogni singola ripetizione della vocale sostenuta /a/ eseguita dal soggetto, l'algoritmo ha generato 10 nuove vocali artificiali, da cui sono stati estratti i parametri di interesse, aumentando la numerosità del dataset. Inoltre, le sequenze dei parametri estratti dai segnali artificiali sono state confrontate con le sequenze di periodo, ampiezza e CPPS note per stimare l'errore dell'algoritmo di estrazione delle feature.
Le feature estratte dal segnale reale e dal segnale artificiale, insieme alle stime dell'errore, sono state utilizzate per allenare un modello di regressione logistica pesata tramite la funzione Matlab "fitglm" per discriminare i soggetti

HE dai pazienti PD e i pazienti PA dai pazienti PD. I pesi dell'algoritmo di classificazione sono stati definiti in modo da far tenere maggiormente in considerazione al classificatore la combinazione di feature che presentano il minor errore relativo medio: ad errori minori corrispondono pesi maggiori e viceversa. Quindi, utilizzando la classificazione pesata e propagando gli errori stimati, il modello ha fornito una stima della classificazione con un intervallo di fiducia del 95%. Le prestazioni del classificatore sono state valutate in termini di massimo numero di classificati e massima accuratezza tramite una metrica chiamata accuratezza dei classificati che è stata definita come la somma dei veri positivi e dei veri negativi degli elementi ritenuti classificati (elementi la cui forchetta di probabilità data dall'intervallo di fiducia del 95% non interseca la soglia di classificazione) diviso il numero totale degli elementi.
Nel discriminare i pazienti PD dai soggetti HE, sono state raggiunte accuratezze dei classificati del 95%, mentre nel discriminare i pazienti PD dai pazienti PA, sono state ottenute accuratezze dei classificati fino al 85%. In entrambe le classificazioni sono stati utilizzati segnali artificiali generati con sequenze di periodi e ampiezze note per aumentare la numerosità del dataset

Indice

1 Segnale vocale **6**
1.1 Introduzione 6
1.2 La fonazione 6
1.3 Classificazione della voce 7
1.4 Malattia di Parkinson e voce 9

2 Materiali e metodi **11**
2.1 Partecipanti e procedure sperimentali 11
2.2 Acquisizione di dati 13
2.3 Estrazione dei parametri 14
2.3.1 Parametri temporali 14
2.3.2 Separazione dei *frame* 19
2.3.3 Cepstrum 19
2.3.4 Elenco dei parametri 24
2.4 Parametri per soggetti sani e per soggetti parkinsoniani 28

3 La generazione di vocali artificiali **29**
3.1 Introduzione 29
3.2 Generazione di sequenze random con il metodo Monte Carlo . 32
3.3 Concatenazione dei frame 36
3.3.1 Metodo dello pseudo-periodo candidato 36
3.3.2 Il metodo del "liftering" 38
3.3.3 Il metodo di ricampionamento dei periodi candidati . . 43
3.3.4 Il metodo di trasformazione del cepstrum 45

4 Valutazione dell'errore dell'algoritmo di estrazione **50**
4.1 Introduzione 50
4.2 Risultati e discussione 55

5 Classificazione **63**
5.1 Introduzione 63

5.2 Feature selection 63
5.3 La regressione logistica 66
5.3.1 L'algoritmo combinatorio 68
5.3.2 Pesi stimati 68
5.3.3 L'intervallo di fiducia 70

6 Risultati **73**
6.1 Introduzione 73
6.2 Confronto tra voci nella malattia di Parkinson e voci sane . . 76
6.3 Confronto tra voci nella malattia di Parkinson e voci patologiche 87

7 Conclusioni **94**

Capitolo 1

Segnale vocale

1.1 Introduzione

Nel seguente capitolo sarà affrontato il problema della produzione del parlato. Si partirà con una breve descrizione degli organi vocali e dei fonemi coinvolti nella fonazione. Questa parte sarà seguita dalla classificazione dei suoni in vocalizzati e non vocalizzati. Il capitolo si concluderà con un'analisi del parlato nei pazienti affetti dalla malattia di Parkinson.

1.2 La fonazione

L'apparato fonatorio può essere rappresentato come un sistema fluido - meccanico composto da un generatore di pressione continua (i polmoni), un'attuatore (le corde vocali) ed un risuonatore (il vocal tract). La produzione di suoni inizia con i polmoni che si contraggono per espellere l'aria. Quest'aria spinta verso l'alto viaggia nel tratto bronchiale, raggiunge e attraversa la glottide, facendo vibrare le corde vocali. [1]. Le corde vocali (o pliche vocali) sono costituite da un tessuto molto elastico capace di contrarsi e rilassarsi in maniera tale da produrre vibrazioni complesse. Al contrario quando sono aperte l'aria passa senza ostacoli e non vibrano. Successivamente il flusso d'aria passa nella laringe, nella faringe e nelle cavità nasali e orali e giunge all'esterno. Tutte queste cavità che si trovano al disopra della glottide costituiscono l'apparato di risonanza. La Figura 1.1 mostra questo sistema.
In altre parole, quello che accade è che l'energia pressoria continua proveniente dai polmoni viene trasformata in energia vibrazionale dalle corde vocali. Il flusso d'aria modulato risuona nel vocal tract e nella cavità orale che agisce come un filtro. L'alterazione volontaria del vocal tract e del cavo orale consente di articolare i fonemi durante il parlato.

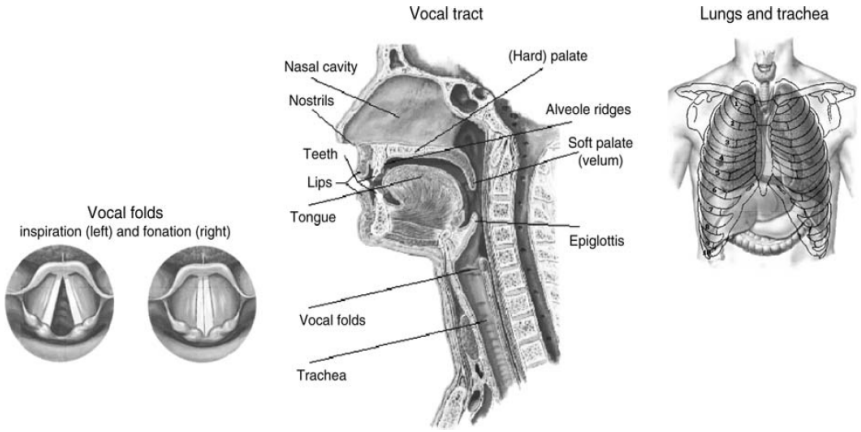

Figura 1.1: Schema dell'apparato fonatorio umano [2]

1.3 Classificazione della voce

I suoni del linguaggio umano possono essere classificati in suoni vocalizzati e in suoni non vocalizzati. I suoni vocalizzati coinvolgono la vibrazione delle corde vocali. Il flusso d'aria è costretto a passare attraverso la glottide, cioè lo spazio che si forma tra le corde vocali, e le fa vibrare. Come risultato del fatto che le pliche vocali ostruiscono il passaggio dell'aria, si hanno una serie di impulsi di pressione quasi periodici che poi passano attraverso il tratto vocale. Un esempio di questi impulsi sono in Figura 1.2. Si può osservare come all'inizio le corde vocali sono chiuse bloccando il flusso d'aria. Il tratto in salita indica che le corde vocali si stanno allontanando, dell'aria lentamente inizia a passare fino a quando le pliche sono completamente staccate (il flusso d'aria è massimo). Poi tornano a chiudersi fino a impedire nuovamente all'aria di passare. La durata di un ciclo è definito periodo del pitch [3].

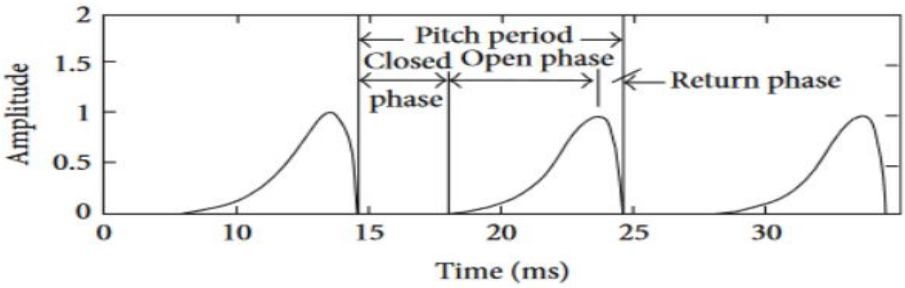

Figura 1.2: Forma d'onda periodica del flusso d'aria glottale [3].

La frequenza fondamentale, ovvero l'inverso del periodo del pitch e le formanti sono le grandezze di interesse nei suoni vocalizzati.
La frequenza fondamentale è la frequenza alla quale le corde vocali si aprono e si chiudono e dipende da diversi fattori, come la rigidità e la lunghezza delle corde vocali. Dato che le corde vocali dei maschi sono più lunghe e più rigide rispetto a quelle della femmine risulta che i maschi hanno una voce più bassa (la frequenza fondamentale parte da 80 Hz) mentre le femmine hanno una frequenza più alta, che parte da 160 Hz [4]. Inoltre il pitch varia con l'età e le condizione di salute.
Le formanti sono relative a ciò che accade nel tratto vocale per ottenere i vari tipi di fonemi. In sostanza il segnale di eccitazione che arriva dalle pliche vocali è filtrato nel tratto vocale che può cambiare conformazione grazie alle parti mobili che lo compongono come la lingua, le labbra, la mandibola.
Dal punto di vista spettrale quello che succede è che il profilo dello spettro del segnale proveniente dalle corde vocali viene modellato nel tratto vocale generando picchi detti formanti. In altre parole, le formanti sono i picchi spettrali in corrispondenza dei massimi dell'inviluppo spettrale. La distribuzione di questi massimi determina le differenze tra i diversi fonemi. Come esempio, si possono osservare tre diversi picchi o formanti nel grafico nel dominio della frequenza di un breve registrazione vocale (Figura 1.3) [1].
I suoni non vocalizzati non coinvolgono la vibrazione delle corde vocali ma si ottengono facendo passare l'aria attraverso una piccola apertura creata da qualche parte lungo tratto vocale. Quello che si ottiene è un segnale che viene modulato dal filtro del vocal tract e che appare quasi come rumore casuale [5]. Esempi di suoni non vocalizzati sono i fonemi sordi come /s/, /f/, /t/.

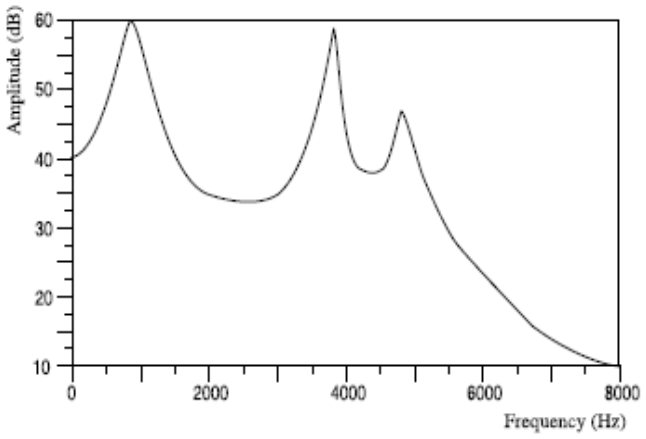

Figura 1.3: Inviluppo di uno spettro di una breve registrazione vocale, che mostra tre distinti picchi o formanti [1].

Si consideri il segnale vocale della parola "si" pronunciata da un maschio adulto, e rappresentato in Figura 1.4. I due fonemeni e i rispettivi intervalli di tempo che costituiscono la parola sono /S/: 1.42-1.44 s; /I/: 1.45-1.5 s. Si possono notare le caratteristiche dei due fonemi nel segnale in Figura 1.4 il fonema /s/ appare come un rumore casuale mentre il fonema /i/ si distingue per il suo comportamento chiaramente oscillante.

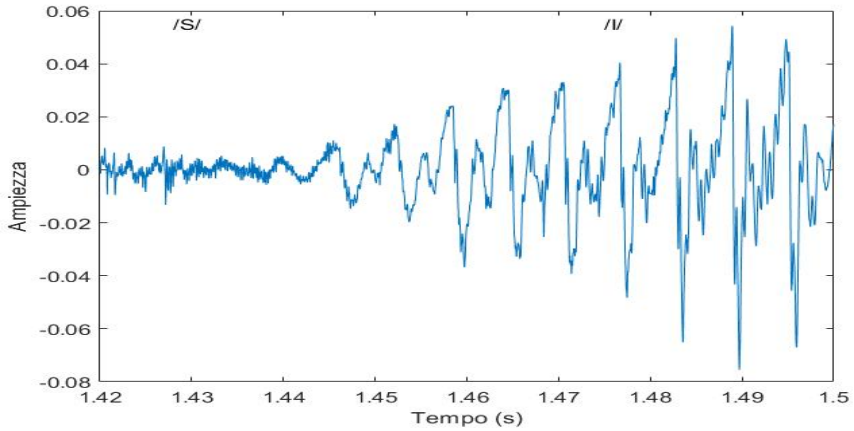

Figura 1.4: Segnale vocale della parola "si".

Si può osservare come ogni volta che la conformazione dell'apparato fonatorio si modifica, le caratteristiche spettrali del segnale cambiano e quindi non è più stazionario. Per questo motivo occorre separare gli eventi fonatori in maniera tale da estrarre le feature adatte al tipo di fonema analizzato.

1.4 Malattia di Parkinson e voce

Il morbo di Parkinson (PD) è una malattia neurodegenerativa che colpisce milioni di persone che hanno un'età media di 65 anni in tutto il mondo [6]. Il morbo di Parkinson è causato dalla degenerazione o dalla lesione di parte dei gangli basali che sono un gruppo di nuclei sottocorticali, ovvero delle regioni di sostanza grigia che si trovano all'interno del cervello, i quali giocano un ruolo fondamentale nelle funzioni motorie. Infatti la loro degenerazione in concomitanza all'assenza di dopamina, un neurotrasmettitore prodotto da un loro componente, provocano i sintomi motori tipici della malattia di Parkinson. I sintomi motori caratteristici della condizione sono tremore, movimenti lenti (bradicinesia), rigidità e instabilità posturale. Nella malattia si manifestano anche sintomi non motori.

Il morbo di Parkinson è classificato usando un certo numero di scale cliniche che permettono di definire la gravità della malattia. La scala di Hoehn e Yahr è comunemente usata per confrontare gruppi di pazienti e per fornire una valutazione della progressione della malattia, che va dalla stadio 0 (nessun segno di malattia) allo stadio 5 (costretto a letto o sulla sedia a rotelle) [7]. La Unified Parkinson's Disease Rating scale (UPDRS) è la scala più utilizzata.
Ad oggi non esiste un test definitivo per la diagnosi della malattia di Parkinson e la malattia deve essere diagnosticata in base a criteri clinici [7]. Gli ultimi risulati della ricerca indicano che approssimativamente il 70-90% dei pazienti con questa malattia ha una forma di disabilità vocale [9]. I disturbi vocali influenzano il linguaggio dei soggetti parkinsoniani in termini di intensità, difficoltà nella pronuncia delle sillabe, voce affannosa, tremore vocale [8]. L'analisi dei loro segnali vocali può risultare utile per la diagnosi e la valutazione della progressione della malattia senza ricorrere a tecniche invasive [6] [9].

Capitolo 2

Materiali e metodi

2.1 Partecipanti e procedure sperimentali

I dati forniti per il lavoro di tesi sono stati raccolti in parte presso l'ambulatorio Parkinson (Città della Salute di Torino), e in parte presso l'ambulatorio di Foniatria (sempre Città della Salute di Torino), dove sono stati certificati i soggetti sani.

I partecipanti che hanno preso parte alla ricerca sono stati:

- 57 soggetti sani (HE): 30 maschi e 27 femmine (età media: 30.7 anni e deviazione tipo: 1.7 anni);
- 67 pazienti patologici e non affetti dalla malattia di Parkinson (PA): 26 maschi e 41 femmine (età media: 52.5 anni e deviazione tipo: 2.3 anni);
- 45 pazienti con il morbo di Parkinson (PD): 27 maschi e 18 (età media: 66.8 anni e deviazione tipo: 1.9 anni).

I soggetti HE e i soggetti PA sono stati sottoposti ad un'analisi endoscopica per certificare la condizione medica dei soggetti.
Il dataset fornito (Tabella 2.1) risulta essere sbilanciato sia per quanto riguarda l'età (Figura 2.1), sia per quanto riguarda il genere dei soggetti (Figura 2.2). Ad esempio i soggetti PD hanno un'età media che è più del doppio rispetto ai soggetti HE. I soggetti sani (HE) sono composti dal 50% di maschi e dal 50% di femmine, i pazienti patologici non affetti dalla malattia di Parkinson (PA) sono composti da una percentuale maggiore di femmine mentre i pazienti affetti dal morbo di Parkinson (PD) sono composti da una percentuale maggiore di maschi.

Tabella 2.1: Numero, genere, età dei partecipanti (dataset sbilanciato).

	HE	PA	PD
Totale partecipanti	57	67	45
Numero maschi	30	26	27
Numero femmine	27	41	18
Età media (anni)	30.7	52.5	66.8
Deviazione tipo (anni)	1.7	2.3	1.9

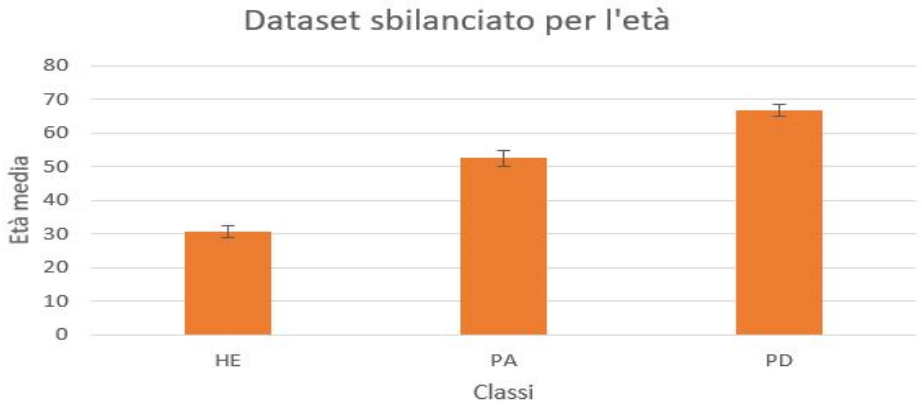

Figura 2.1: Età dei partecipanti che compongono il dataset sbilanciato

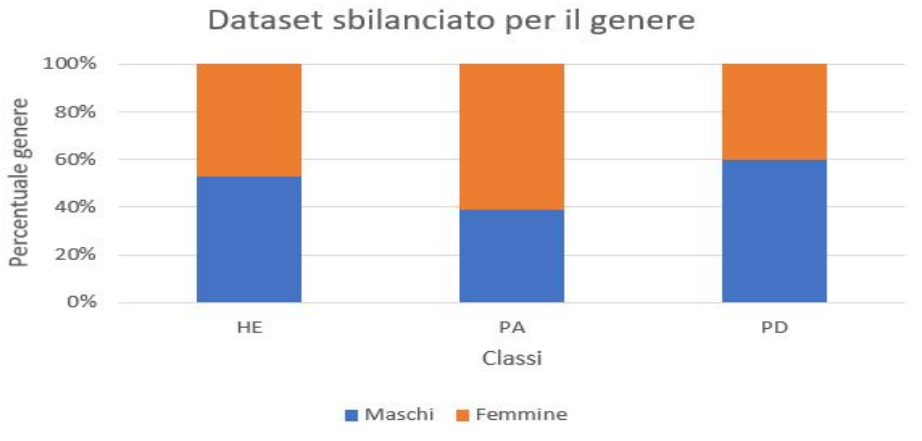

Figura 2.2: Genere dei partecipanti che compongono il dataset sbilanciato.

A tutti i partecipanti è stato chiesto di eseguire i seguenti tre task:

- ripetere tra volte la vocale sostenuta /a/;
- leggere una brano foneticamente bilanciato per circa 60 s;
- libero per circa 60 s.

In questo lavoro di tesi solo il task della voce sostenuta è stato analizzato.

2.2 Acquisizione di dati

Per l'acquisizione dei segnali sono stati utilizzati due microfoni: il microfono in aria e il microfono a contatto. L'utilizzo del microfono a contatto è giustificato dal fatto che è insensibile al rumore ambiente seppur il contenuto spettrale del segnale catturato è inferiore rispetto al microfono in aria. Il microfono in aria, invece, riesce a catturare lo spettro vocale con più risoluzione rispetto al microfono a contatto. Un confronto tra i due segnali è riportato in Figura 2.4.

Figura 2.3: Esempio di posizionamento del microfono ad aria e del microfono a contatto.

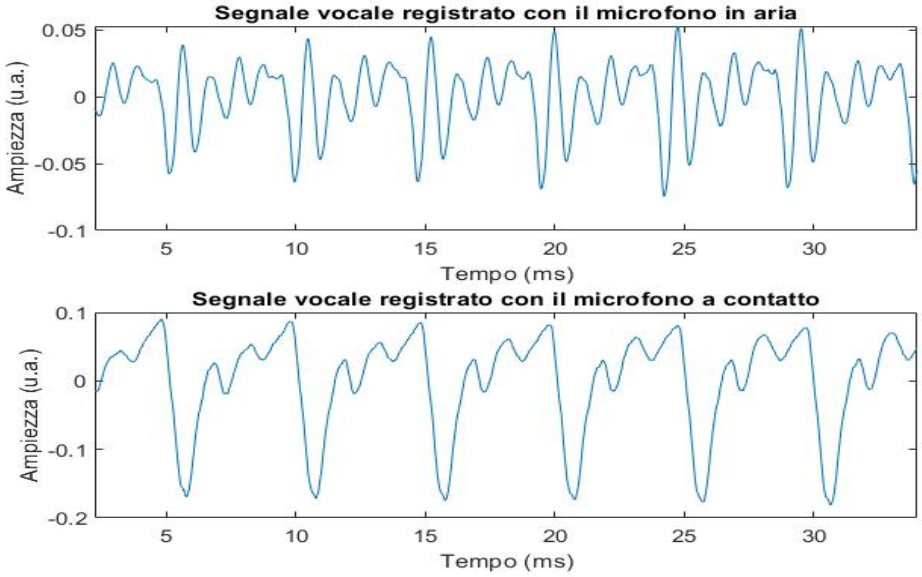

Figura 2.4: Segnale vocale registrato con il microfono ad aria e con il microfono a contatto.

Entrambi i microfoni sono stati connessi all'ingresso stereo di un registratore portatile (Edirol Roland R-09HR). Il registratore campiona il segnale ad una frequenza di 44100 Hz usando un convertitore analogico-digitale caratterizzato da 16 bit di risoluzione.
I segnali vocali utilizzati per il lavoro di tesi sono le tre ripetizioni della vocale /a/ sostenuta registrati dal microfono in aria.

2.3 Estrazione dei parametri

Nel seguente capitolo saranno descritti i parametri che rappresentano il contesto di interesse, che in questo caso sono i soggetti sani, i pazienti parkinsoniani e i pazienti patologici non parkinsoniani.

2.3.1 Parametri temporali

I primi parametri studiati permettono di valutare la stabilità in frequenza e in ampiezza del segnale vocale andando a misurare le variazioni di queste due grandezze da ciclo a ciclo [4] [10]. Quindi è fondamentale stimare correttamente uno pseudo-periodo. Questa stima risulta essere più affidabile per

segnali quasi periodici rispetto a segnali vocali fortemente perturbati. Comunemente vengono misurate la stabilità in periodo ed ampiezza definite da Jitter e Shimmer.

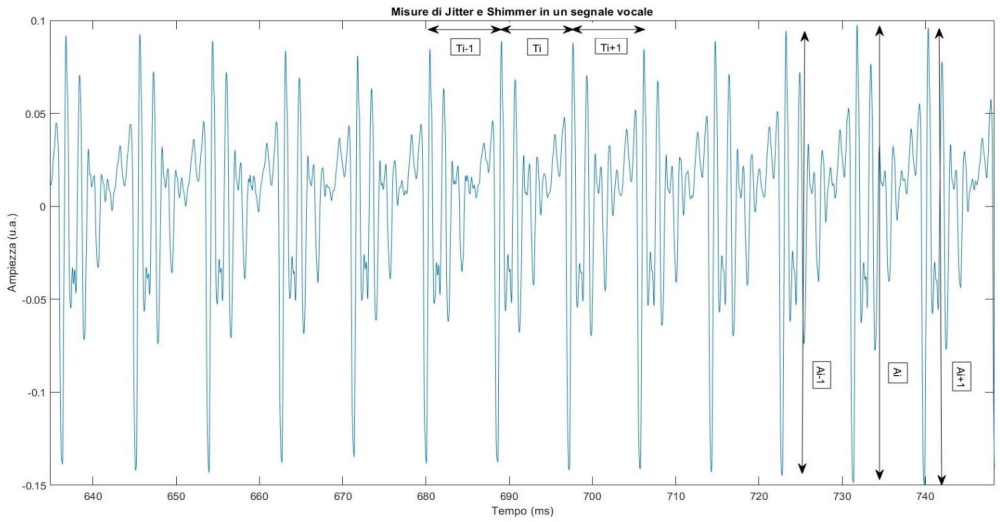

Figura 2.5: Misure di Jitter e Shimmer in un segnale vocale.

La valutazione di Jitter e Shimmer sono composti dalle seguenti famiglie di misure che hanno senso per le vocali sostenute:

- Jitter percentuale (%): misura relativa della variazione del periodo fondamentale da ciclo a ciclo. Si calcola come:

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.1}$$

  dove $T_0^{(i)}, i = 1, 2, ..., N$ sono i pseudo-periodi stimati, N è il numero di pseudo-periodi stimati.

- Jitter assoluto (μs): misura della variazione del periodo fondamentale da ciclo a ciclo. Si calcola come:

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_0^{(i)} - T_0^{(i+1)}| \tag{2.2}$$

  dove $T_0^{(i)}, i = 1, 2, ..., N$ sono i pseudo-periodi stimati, N è il numero di pseudo-periodi stimati.

- RAP (%): misura relativa della variazione del pitch da periodo a periodo con una media lungo 3 periodi successivi. Si calcola come:

$$RAP = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{T_0^{(i-1)}+T_0^{(i)}+T_0^{(i+1)}}{3} - T_0^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.3}$$

  dove $T_0^{(i)}, i = 1, 2, ..., N$ sono i pseudo-periodi stimati, N è il numero di pseudo-periodi stimati.

- PPQ (%): misura relativa della variazione del pitch da periodo a periodo con una media lungo 5 periodi successivi. Si calcola come:

$$PPQ = \frac{\frac{1}{N-4}\sum_{i=1}^{N-4}\left|\frac{1}{5}\sum_{r=0}^{4} T_0^{(i+r)} - T_0^{(i+2)}\right|}{\frac{1}{N}\sum_{i=1}^{N} T_0^{(i)}} \tag{2.4}$$

  dove $T_0^{(i)}, i = 1, 2, ..., N$ sono i pseudo-periodi stimati, N è il numero di pseudo-periodi stimati.

- vF_0 (%): rappresenta la deviazione standard relativa della frequenza fondamentale. Si misura come:

$$vF_0 = \frac{\sigma}{F_0} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{N}\sum_{j=1}^{N} F_0^{(j)} - F_0^{(i)})^2}}{\frac{1}{N}\sum_{i=1}^{N} F_0^{(i)}} \tag{2.5}$$

  dove $F_0 = \frac{1}{N}\sum_{i=1}^{N} F_0^{(i)}$ e $F_0^{(i)} = \frac{1}{T_0^{(i)}}$ sono i valori delle frequenze fondamentali di ciascun pseudo-periodo stimato, $T_0^{(i)}, i = 1, 2, ..., N$ sono i pseudo-periodi stimati, N è il numero di pseudo-periodi stimati.

- Shimmer percentuale (%): misura relativa della variazione dell'ampiezza picco-picco da periodo a periodo. Si calcola come:

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A^{(i)} - A^{(i+1)}|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \tag{2.6}$$

  dove $A^{(i)}, i = 1, 2, ..., N$ sono le ampiezze picco-picco, N è il numero di impulsi stimati.

- Shimmer assoluto (dB): misura della variazione dell'ampiezza picco - picco da periodo a periodo. Si calcola come:

$$ShdB = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|20\,log(A^{(i+1)}/A^{(i)})\right| \tag{2.7}$$

dove $A^{(i)}, i = 1, 2, ..., N$ sono le ampiezze picco-picco, N è il numero di impulsi stimati.

- APQ (%): misura della variazione dell'ampiezza picco-picco da periodo a periodo con una media lungo 11 periodi successivi. Si calcola come:

$$APQ = \frac{\frac{1}{N-10}\sum_{i=1}^{N-10}\left|\frac{1}{11}\sum_{r=0}^{10}A^{(i+r)} - A^{(i+5)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \tag{2.8}$$

dove $A^{(i)}, i = 1, 2, ..., N$ sono le ampiezze picco-picco, N è il numero di impulsi stimati.

- vAm (%): deviazione standard relativa dell'ampiezza picco-picco. Si calcola come:

$$vAm = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{1}{N}\sum_{j=1}^{N}A^{(j)} - A^{(i)})^2}}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \tag{2.9}$$

dove $A^{(i)}, i = 1, 2, ..., N$ sono le ampiezze picco-picco, N è il numero di impulsi stimati.

Il rapporto armoniche-rumore (HNR) è un parametro che permette di quantificare l'armonicità del segnale e può essere calcolato a partire dall'autocorrelazione.
In teoria dei segnali la funzione di autocorrelazione $R_x[n]$ nel discreto è definta come:

$$R_x[n] = \sum_{k=-\infty}^{+\infty} x^*[k]x[k+n] \tag{2.10}$$

dove n è il lag (ritardo). Nella funzione di autocorrelazione si ha il massimo nell'origine, che coincide con la potenza media del segnale. La funzione di autocorrelazione normalizzata $R_x^{'}[n]$:

$$R_x^{'}[n] = \frac{R_x[n]}{R_x[0]} \tag{2.11}$$

Si consideri il modello di rumore additivo in cui si suppone di avere un segnale periodico $h[n]$ al quale si somma un rumore $N[n]$: $x[n] = h[n] + N[n]$. La funzione di autocorrelazione normalizzata al ritardo $[n] = [\tau_{max}]$ rappresenta la potenza relativa della componente periodica (o armonica) del segnale, e il suo complemento rappresenta la potenza relativa della componente di rumore [11]:

$$R_x^{'}[\tau_{max}] = \frac{R_H[0]}{R_x[0]} \;\; ; \;\; 1 - R_x^{'}[\tau_{max}] = \frac{R_N[0]}{R_x[0]} \tag{2.12}$$

Si definisce il rapporto armoniche-rumore (HNR) come:

$$HNR = 10\, log \frac{R_x^{'}[\tau_{max}]}{1 - R_x^{'}[\tau_{max}]} \tag{2.13}$$

Si osservi come valori elevati di HNR indicano una componente periodica nel segnale maggiore della componente di rumore mentre bassi valori di HNR indicano una componente di rumore maggiore rispetto la componente periodica. Il caso limite si ha per segnali perfettamente periodici in cui il valore del rapporto armoniche-rumore è infinito.

Un altro parametro basato sull'autocorrelazione è la frequenza fondamentale F_0. Questo paramentro è particolarmente importante dato che a partire dalla sua stima si basano anche i parametri relativi alla stabilità del periodo e dell'ampiezza. Il metodo che è stato scelto per stimare la durata di uno pseudo-periodo T_0 (o della frequenza fondamentale $F_0 = \frac{1}{T_0}$) è il metodo dell'autocorrelazione [11].

Questi parametri hanno senso sia per vocali sostenute sia per il parlato (eloquio libero o lettura di un brano).

2.3.2 Separazione dei *frame*

Gli intervalli del segnale sono stati analizzati alla cerca di intervalli silenziosi, intervalli non vocalizzati (contengono suoni non vocalizzati) e intervalli vocalizzati (contengono suoni vocalizzati). L'algoritmo che ha permesso questa classificazione degli intervalli è stato il seguente:

1. Si calcola il valore efficace dell'intervallo $RMS_{intervallo}$ lo si confronta con la metà del valore efficace dell'intero segnale $RMS_{segnale}$
   (a) se $RMS_{intervallo} < \frac{RMS_{segnale}}{2}$, allora l'intervallo considerato è un frame di silenzio;
   (b) altrimenti si va al passo successivo.
2. Se la condizione logica:
   (a) $\left[\frac{|F0_i - F0_{i-1}|}{F0_{i-1}} < 0.5\right] AND\, [HNR_i > 0\, dB]$ è falsa, allora l'intervallo è non vocalizzato;
   (b) altrimenti l'intervallo è vocalizzato.

Sulla base di questo algoritmo è stato possibile calcolare un altro paramentro di interesse che è il rapporto tra suoni vocalizzati e suoni non vocalizzati:

- $(V/uV)(\%)$.

Questo parametro ha senso sia per vocali sostenute sia per il parlato (eloquio libero o lettura di un brano).

2.3.3 Cepstrum

Il concetto di cepstrum fu introdotto nel 1963 da Bogert et al. [12] e il nome è stato ottenuto anagrammando la parola "spectrum" (spettro in italiano). Il cepstrum può essere definto come uno spettro di uno spettro logaritmico: se il primo spettro (spettro di potenza) indica l'energia del segnale attorno a ciascuna frequenza, il secondo spettro indica quanto sono periodiche le componente armoniche nello spettro [4].
Dal punto di vista matematico il cepstrum $C(\tau)$ è definito come [13]:

$$C(\tau) = \left|\mathcal{F}\left\{log\left(|\mathcal{F}\{f(t)\}|^2\right)\right\}\right|^2 \tag{2.14}$$

dove $\mathcal{F}$ è la trasformata di Fourier, $|\mathcal{F}\{f(t)\}|^2$ è lo spettro di potenza e $f(t)$ è il segnale in funzione del tempo. La variabile τ è chiamata "quefrency" (parola che deriva dall'anagramma del termine "frequency", in italiano frequenza) ed ha le dimensioni di tempo.
L'analisi cepstrale è una tecnica che trova impiego in diversi campi, tra cui nell'analisi del parlato.

Dal punto di vista matematico il sistema sorgente-filtro del sistema fonatorio può essere schematizzato dalla seguente relazione:

$$S(z) = H(z) \cdot U(z) \tag{2.15}$$

dove $H(z)$ è la funzione di trasferimento del vocal tract, $U(z)$ è lo spettro del segnale di eccitazione, ovvero lo spettro del segnale che arriva dalle corde vocali e $S(z)$ è lo spettro del suono risultante.
Per le proprietà del teorema di convoluzione, la (2.15), nel dominio del tempo, diventa:

$$s(t) = h(t) * u(t) \tag{2.16}$$

La proprietà del logaritmo afferma che il logaritmo del prodotto di due numeri è la somma dei logaritmi dei numeri stessi. Allora la (2.15), nel dominio del cepstrum, diventa:

$$log|S(z)| = log|H(z)| + log|U(z)| \tag{2.17}$$

In altre parole, nel dominio della frequenza la componente del periodo fondamentale e del tratto vocale sono correlate in modo moltiplicativo mentre nel dominio del cepstrum sono correlate in modo additivo. Come mostrato nel grafico del cepstrum in Figura 2.6, il picco cepstrale è ben visibile e separato dalla componente del tratto vocale e corrisponde al periodo fondamentale [1]. L'utilizzo del cepstrum nell'analisi del parlato permette di misurare il pitch della voce. Infatti la quefrency a cui si presenta il picco cepstrale fornisce un'informazione riguardo la frequenza fondamentale media del frame considerato, mentre la prominenza rispetto al fondo di rumore indica l'armonicità del segnale. Oltre a questo il cepstrum è utile anche per ottenere informazioni sull'inviluppo spettrale per l'analisi del parlato [1].

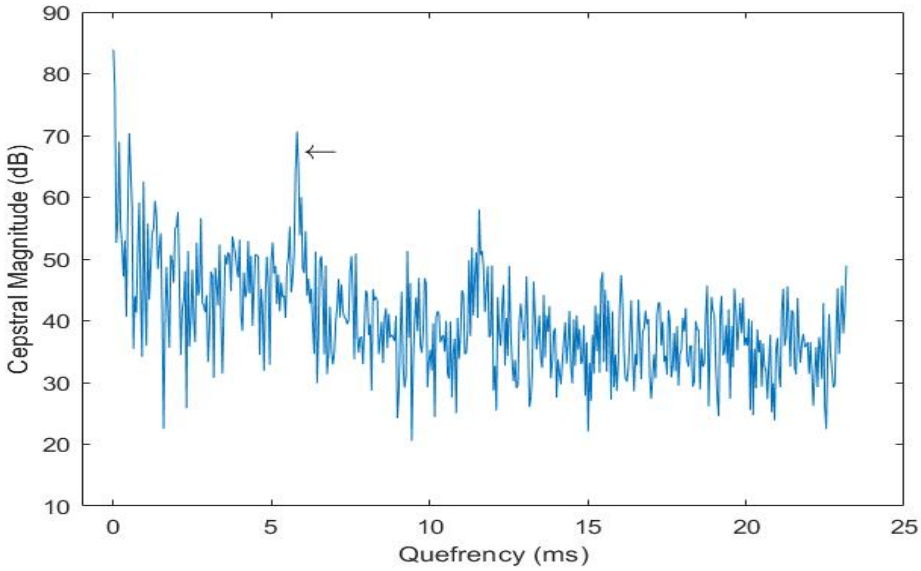

Figura 2.6: Grafico del cepstrum di un tratto di una vocale pronunciata da un adulto maschio. La freccia indica la componente del periodo fondamentale ben visibile.

I parametri che possono essere estratti dal cepstrum sono il CPP (Cepstral Peak Prominence) e il CPPS (CPP-smoothed) [14]. Quest'ultimo permette di stimare correttamente la prominenza del picco cepstrale utilizzando algoritmi di smoothing per ridurre l'effetto del rumore di fondo sulla misura del picco. Il CPP (Cepstral Peak Prominence) è la differenza tra l'ampiezza del picco cepstrale e il valore della retta di regressione in corrispondenza del picco e viene misurata in dB [15]. Un esempio è riportato in Figura 2.7 .

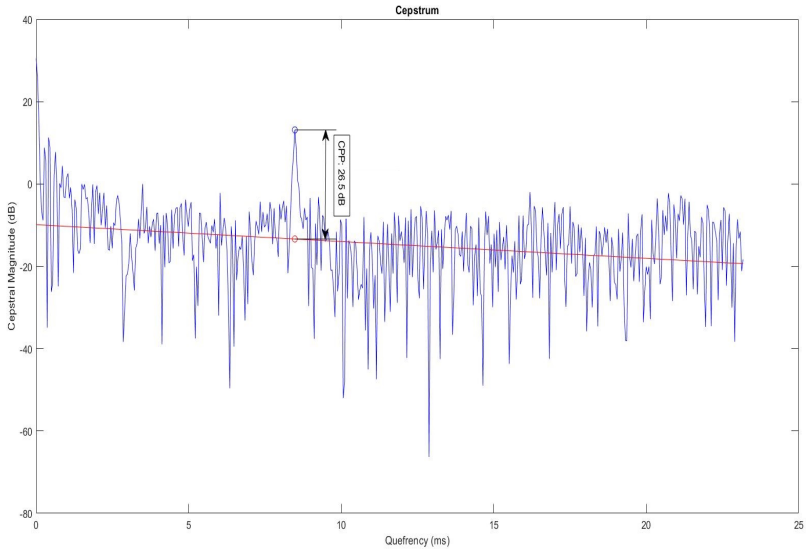

Figura 2.7: Grafico della misura del CPP.

Il CPPS (CPP-smoothed) misura il CPP su una versione mediata del cepstrum. Lo smoothing consiste in due fasi [14]. Nella prima fase i cepstra sono mediati nel tempo: il cepstrum corrente è mediato con un certo numero di cepstra precedenti e successivi al cepstrum considerato. Il cepstrum raffigurato al centro della Figura 2.8 è stato ottenuto mediando 7 frame. Nella seconda fase i cepstra sono mediati lungo la quefrency. Il cepstrum raffigurato in basso della Figura 2.8 è stato ottenuto mediando i valori del cepstrum su 7 bin. Su quest'ultimo cepstrum si misura il parametro di interesse.

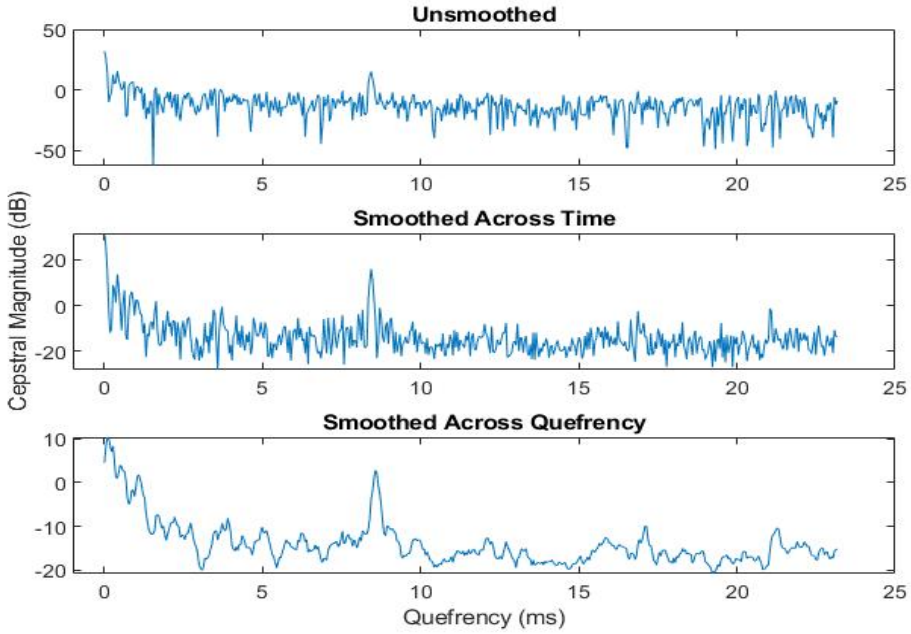

Figura 2.8: Grafico della misura del CPPS.

Per misurare il CPPS è necessario conoscere la posizione e il valore del picco cepstrale. Per misurare queste due quantità è stato considerato un intervallo di cepstrum che va, approssimativamente, da 3.3 ms a 16.7 ms, che corrisponde al range di frequenze che va da 60 Hz a 300 Hz. Infatti, approssimativamente, all'interno di questo intervallo può cadere la frequenza fondamentale F_0 della voce umana. Per la misura del parametro cepstrale è stato sottocampionato il segnale vocale di un fattore 2 e successivamente si è misurato il CPPS da frame lunghi 1024 campioni (46.4 ms) e sovrapposti di 44 campioni (2 ms). In questo modo si riesce ad avere sufficiente copertura del segnale misurato.
I parametri cepstrali sono significativi sia per vocali sostenute sia per il parlato (eloquio libero e lettura).

2.3.4 Elenco dei parametri

I parametri descritti in questa sezione saranno identificati da un numero nei capitoli successivi secondo i riferimenti proposti in Tabella 2.2, Tabella 2.3, Tabella 2.4, Tabella 2.5, Tabella 2.6, Tabella 2.7.

Tabella 2.2: Schema di riferimento per i parametri temporali.

Parametro	**Riferimento**
Jitt	1
Jita	2
RAP	3
PPQ	4
vF_0	5
Shim	6
ShdB	7
APQ	8
vAm	9

Le sequenze temporali di HNR vengono quantizzate e trasformate in distribuzioni statistiche da cui si estraggono i seguenti parametri:

- HNR: media, mediana, moda, deviazione standard, range, 5° e 95° percentile, skewness e kurtosis.

Tabella 2.3: Schema di riferimento per la distribuzione HNR.

Parametro	**Riferimento**
HNR media	10
HNR mediana	11
HNR moda	12
HNR dev. std.	13
HNR range	14
HNR 5° percentile	15
HNR 95° percentile	16
HNR skewness	17
HNR kurtosis	18

Le sequenze temporali di F_0 vengono quantizzate e trasformate in distribuzioni statistiche da cui si estraggono i seguenti parametri:

- F_0: media, mediana, moda, deviazione standard, range, 5° e 95° percentile, skewness e kurtosis.

Tabella 2.4: Schema di riferimento per la distribuzione F_0.

Parametro	**Riferimento**
F_0 media	19
F_0 mediana	20
F_0 moda	21
F_0 dev. std.	22
F_0 range	23
F_0 5° percentile	24
F_0 95° percentile	25
F_0 skewness	26
F_0 kurtosis	27

Le sequenze temporali di RMS vengono quantizzate e trasformate in distribuzioni statistiche da cui si estraggono i seguenti parametri:

- RMS: media, mediana, moda, deviazione standard, range, 5° e 95° percentile, skewness e kurtosis.

Tabella 2.5: Schema di riferimento per la distribuzione RMS.

Parametro	**Riferimento**
RMS media	28
RMS mediana	29
RMS moda	30
RMS dev. std.	31
RMS range	32
RMS 5° percentile	33
RMS 95° percentile	34
RMS skewness	35
RMS kurtosis	36

Le sequenze temporali di CPPS vengono quantizzate e trasformate in distribuzioni statistiche da cui si estraggono i seguenti parametri:

- CPPS: media, mediana, moda, deviazione standard, range, 5° e 95° percentile, skewness e kurtosis.

Tabella 2.6: Schema di riferimento per la distribuzione CPPS.

Parametro	**Riferimento**
CPPS media	37
CPPS mediana	38
CPPS moda	39
CPPS dev. std.	40
CPPS range	41
CPPS 5° percentile	42
CPPS 95° percentile	43
CPPS skewness	44
CPPS kurtosis	45

Tabella 2.7: Schema di riferimento per il rapporto tra suoni vocalizzati e suoni non vocalizzati.

Parametro	**Riferimento**
v/uV	46

2.4 Parametri per soggetti sani e per soggetti parkinsoniani

Diversi studi hanno analizzato i parametri descritti in questo capitolo per soggetti sani e per soggetti parkinsoniani. I risultati ottenuti mostrano come i soggetti parkinsoniani presentano Jitter più alto e HNR più basso rispetto i soggetti sani [8] [16]. Un esempio è riportato in Figura 2.9 che rappresenta valori di Jitter estratti da vocali sostenute eseguite sia dai soggetti sani, sia dai soggetti parkinsoniani.

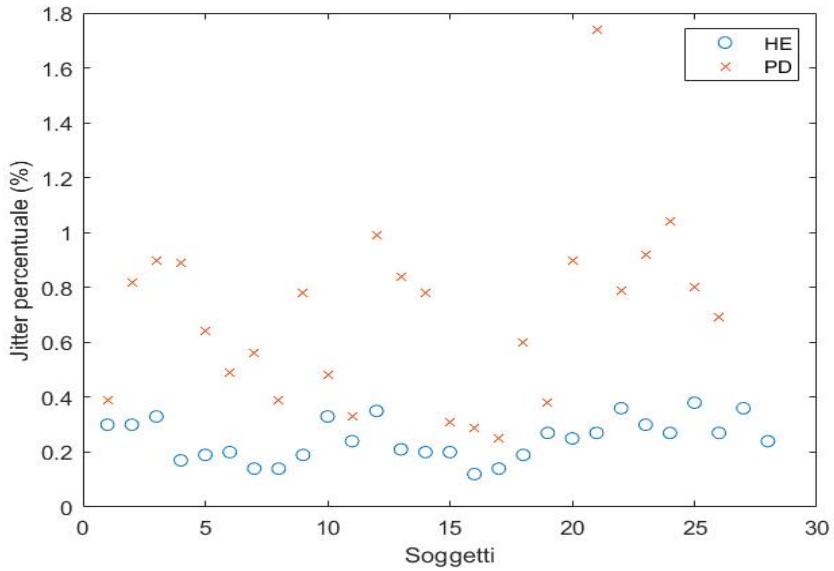

Figura 2.9: Jitter per soggetti sani e per soggetti parkinsoniani.

Capitolo 3

La generazione di vocali artificiali

3.1 Introduzione

In questo capitolo saranno descritti i metodi che hanno permesso di generare segnali vocali artificiali con periodi, ampiezze e CPPS noti. In seguito questi segnali artificiali sono stati utilizzati sia per stimare l'errore dell'algoritmo di estrazione dei parametri, sia per aumentare la numerosità del dataset.
Il dataset fornito è risultato essere sbilanciato per quanto riguarda sia l'età, sia il genere dei soggetti. Il problema nasce dal fatto che i parametri estratti dipendono da questi due fattori e per questo motivo è necessario bilanciare il dataset nell'età e nel genere. Ad esempio, la differenza tra le distribuzioni F_0 che sono state ricavate da un segnale vocale registrato da una femmina sana e da un maschio sano, aventi la stessa età, può essere osservata in Figura 3.1.

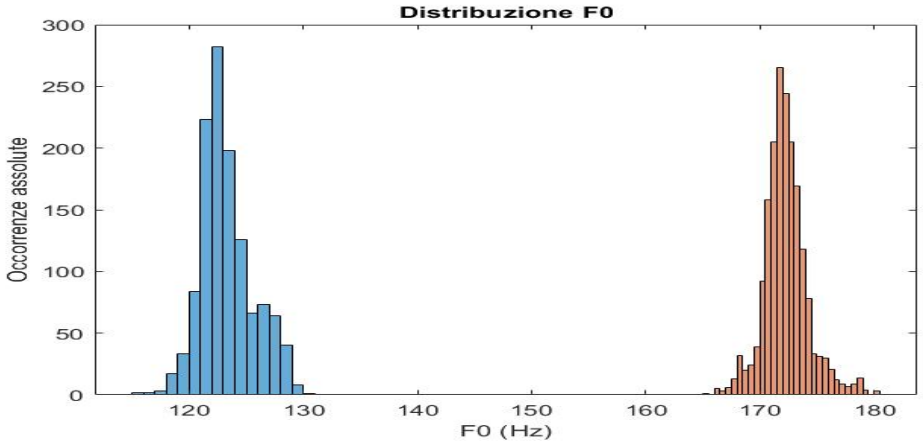

Figura 3.1: Distribuzione F_0 per una voce maschile sana in blu e per una voce femminile sana in rosso (soggetti con la stessa età).

Un altro esempio è riportato per le distribuzioni CPPS ricavate da un segnale vocale registrato da un maschio giovane sano e da un maschio anziano sano (Figura 3.2) che sono state confrontate con la funzione Matlab "kstest2" (Two-sample Kolmogorov-Smirnov test). Il test ha rigettato l'ipotesi nulla che le due distribuzioni sono simili con un livello di significatività minore del 5%.

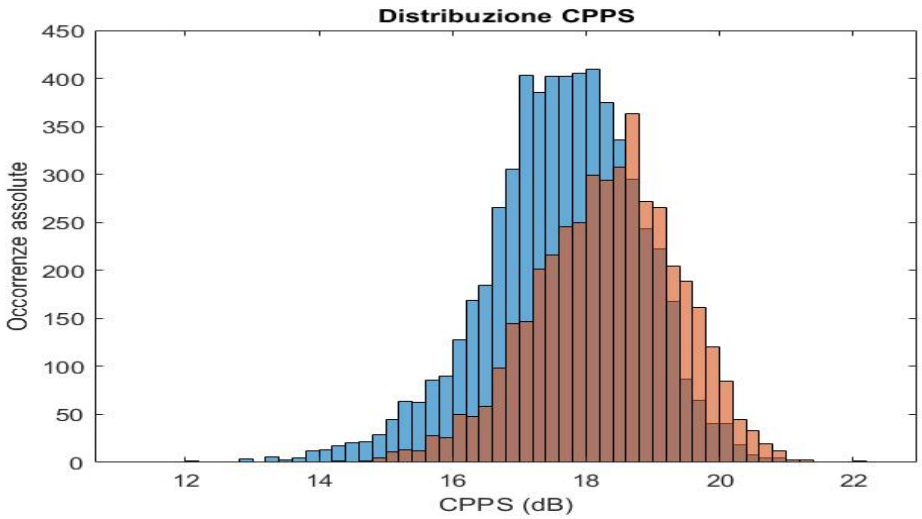

Figura 3.2: Distribuzione CPPS per un maschio giovane sano in blu e per un maschio anziano sano in rosso.

Quindi è stato deciso di selezionare 10 soggetti da ogni classe per ottenere dataset bilanciati (Tabella 3.1) sia nell'età, sia nel genere. In particolare sono stati scelti 5 maschi e 5 femmine per ogni classe ed, ad esempio, sono stati selezionati i soggetti più anziani della classe HE e i pazienti più giovani della classe PD.

Tabella 3.1: Numero, genere, età dei partecipanti (dataset bilanciato).

	HE	PA	PD
Totale partecipanti	10	10	10
Numero maschi	5	5	5
Numero femmine	5	5	5
Età media (anni)	50.8	51.8	52.3
Deviazione tipo (anni)	4.2	3.9	3.8

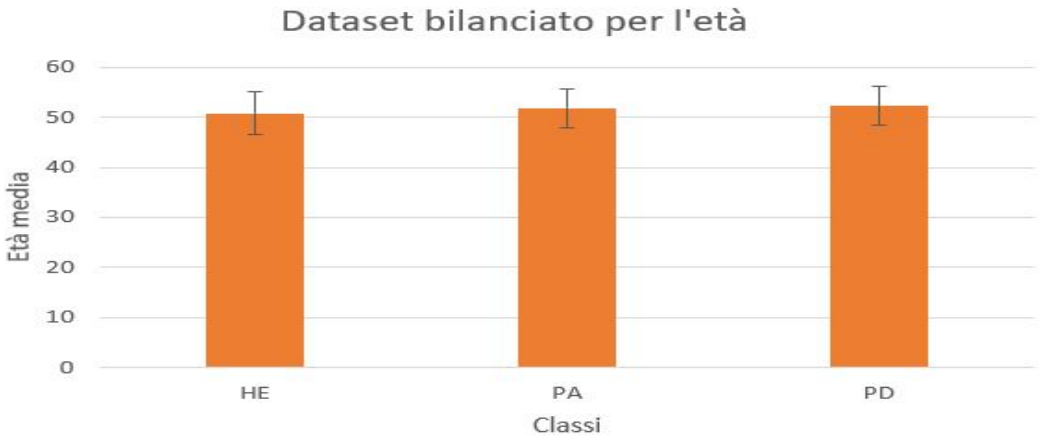

Figura 3.3: Età dei partecipanti che compongono il dataset bilanciato

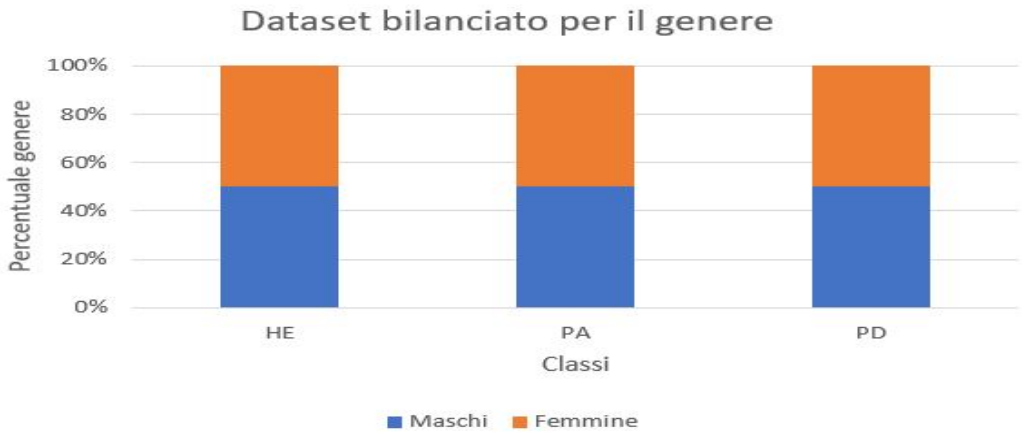

Figura 3.4: Genere dei partecipanti che compongono il dataset bilanciato.

3.2 Generazione di sequenze random con il metodo Monte Carlo

Il metodo Monte Carlo (MC) è un'ampia classe di algoritmi basati sulla generazione di variabili casuali per la risoluzione di problemi matematici a partire da distribuzioni di probabilità assegnate [17]. In questo contesto è stato utilizzato per generare sequenze temporali che hanno distribuzioni statistiche compatibili con le distribuzioni statistiche di riferimento. Uno dei più famosi algoritmi MC è l'algoritmo di Metropolis-Hastings [18].
Le vocali artificiali sono state generate utilizzando sequenze di periodi, ampiezze e CPPS noti ottenute tramite una simulazione MC con l'algoritmo di Metropolis-Hastings.
Di seguito sarà riportato come esempio l'algoritmo che ha permesso di generare sequenze di periodi noti (i passi che sono stati seguiti per ottenere sequenze di ampiezze e CPPS sono gli stessi). Si consideri la sequenza di periodi calcolati da un segnale vocale reale e la sua distribuzione D_t. Si ottenga una sequenza composta dalle differenze tra periodi consecutivi e se ne calcoli la funzione cumulativa empirica F_t.

1. Tramite il metodo MC diretto, si propone un salto casuale calcolando l'inverso della funzione cumulativa F_t rispetto ad una frequenza scelta in maniera random tra 0 ed 1.

2. Si aggiunge il salto proposto al passo precedente e si ottiene il nuovo valore $D_t(i)$.

3. Si genera la proposta tramite il metodo Monte Carlo con l'algoritmo di Metropolis-Hasting come:

$$A = min\left(1, \frac{D_t(i)}{D_t(i-1)}\right). \tag{3.1}$$

4. Si accetta o si rigetta::

   (a) si genera un numero uniforme casuale u, dove $u \in [0; 1]$;

   (b) si accetta il nuovo campione se $u \leq A$;

   (c) si rigetta se $u > A$.

In riferimento al passo 2 dell'algoritmo appena descritto, sono stati proposti due diversi metodi di concatenazione. Il primo metodo, al quale ci si riferirà nei capitoli successivi come al metodo del passo indipendente, consiste

nel sommare il salto proposto al periodo generato casualmente nell'iterazione precedente. In questo caso vengono prodotte sequenze temporali molto lunghe ma con distribuzioni statistiche compatibili con le distribuzioni statistiche di riferimento. I primi valori della sequenza non sono affidabili dato che non provengono da zone di massima densità delle distribuzioni statistiche di riferimento (queste zone non possono essere note a priori), ma sono in una zona scelta casualmente all'inizio. Quindi è necessario tagliare la parte iniziale della sequenza, chiamata di warm-up, ottenendo vettori di lunghezza pari alle sequenze estratte dai segnali reali.
Un esempio di evoluzione temporale dei periodi è mostrato in Figura 3.5 e la loro distribuzione è riportata in Figura 3.6.

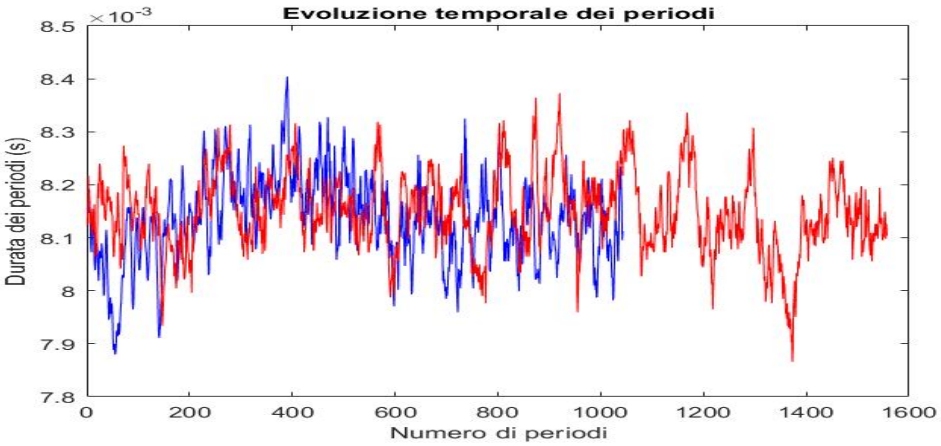

Figura 3.5: Evoluzione temporale di periodi estratti dal segnale reale in blu ed evoluzione temporale di periodi generati con il passo indipendente in rosso.

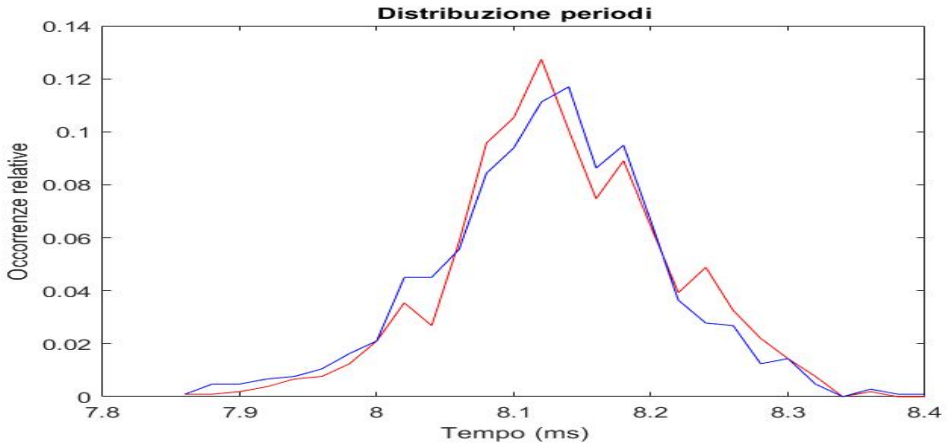

Figura 3.6: Distribuzione dei periodi estratti dal segnale reale in blu e distribuzione dei periodi generati con il passo indipendente in rosso.

L'altro metodo, al quale ci si riferirà nei capitoli successivi come al metodo del passo dipendente, consiste nel sommare il salto proposto al periodo estratto dal segnale reale. Un esempio di evoluzione temporale dei periodi è mostrato in Figura 3.7 e la loro distribuzione è riportata in Figura 3.8.

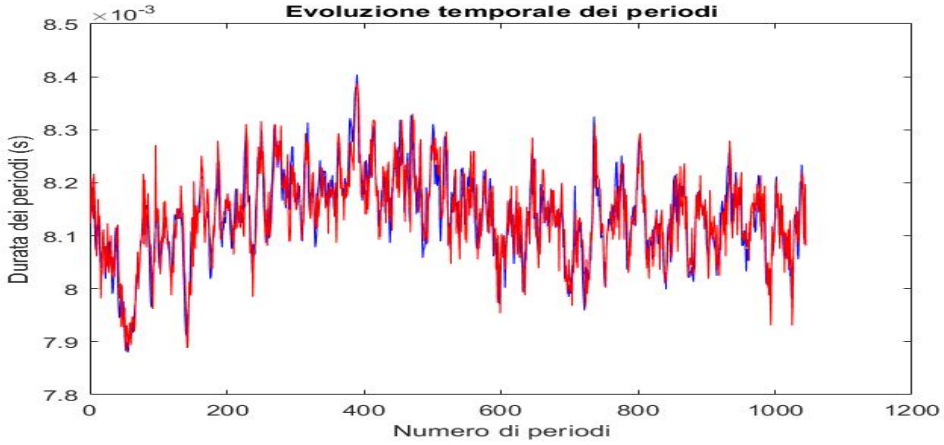

Figura 3.7: Evoluzione temporale di periodi estratti dal segnale reale in blu ed evoluzione temporale di periodi generati con il passo dipendente in rosso.

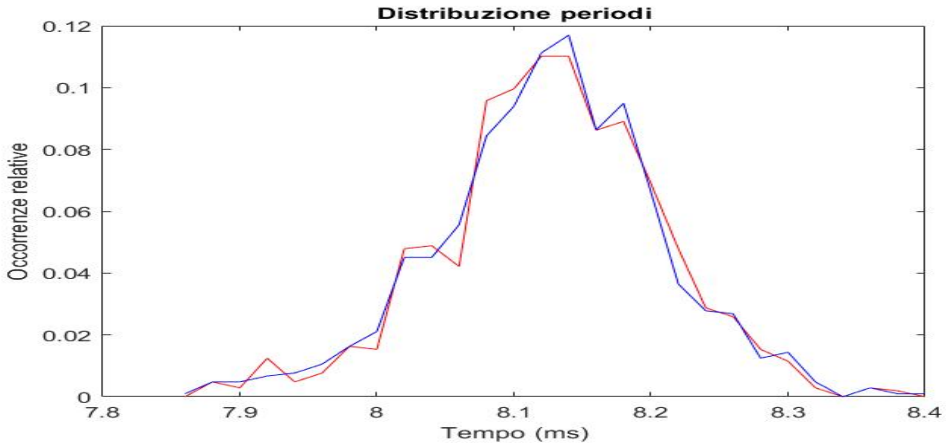

Figura 3.8: Distribuzione dei periodi estratti dal segnale reale in blu e distribuzione dei periodi generati con il passo dipendente in rosso.

Per confrontare la distribuzione ottenuta tramite una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings e la distribuzione proveniente dai segnali registrati dai pazienti è stato utilizzato il test di Kolmogorov - Smirnov. Nel caso in cui il test non ha rigettato l'ipotesi nulla, ovvero le due distribuzioni sono da considerarsi simili, la nuova distribuzione è stata quella che è stata utilizzata successivamente per produrre un segnale artificiale. Nel caso in cui il test ha rigettato l'ipotesi nulla con un livello di significatività minore dell'1%, la generazione MC continua.
Utilizzando la funzione Matlab "kstest2", il test non ha rigettato l'ipotesi nulla con un livello di significatività minore dell'1% per le distribuzioni riportate in Figura 3.6 e in Figura 3.8.
Un esempio di CPPS ottenute con l'algoritmo di Metropolis-Hastings e di distribuzioni calcolate dal segnale reale è riportato in Figura 3.9.

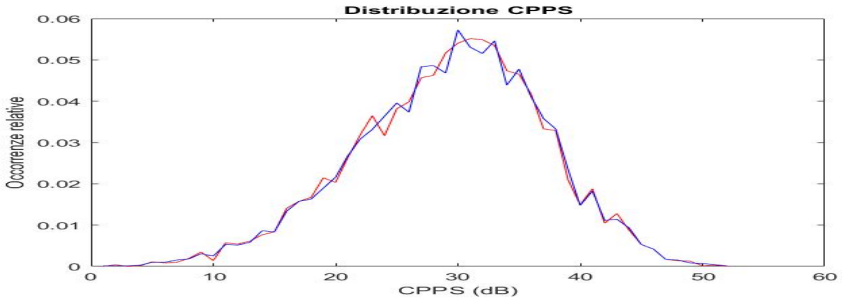

Figura 3.9: Distribuzione di CPPS ottenuta con l'algoritmo di Metropolis-Hastings in rosso e distribuzione di CPPS calcolata dal segnale reale in blu.

3.3 Concatenazione dei frame

Le sequenze di periodi, ampiezze e CPPS noti ottenute con una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings basate sia sul passo indipendente, sia sul passo dipendente, sono state utilizzate per produrre nuovi segnali artificiali. Per ogni singola ripetizione della vocale sostenuta /a/ eseguita dai 10 soggetti che appartengono alle tre classi, sono state generate 10 vocali artificiali con metodi diversi che saranno trattati nel dettaglio nelle sezioni che seguono. In particolare, le prime tre sono relative alle alterazioni di periodi e ampiezze mentre la quarta alle modifiche di CPPS.

3.3.1 Metodo dello pseudo-periodo candidato

Il segnale vocale registrato da un paziente è stato diviso in finestre temporali della durata di un periodo, stimato con il metodo dell'autocorrelazione, una delle tecniche più tradizionali tra quelle riportate da Rabiner et al [19]. È necessario osservare come la stima del periodo è possibile per gli intervalli che non sono silenziosi e che non contengono suoni non vocalizzati. Questi frame sono stati scartati a priori utilizzando l'algoritmo descritto nella Sezione 2.3.2 [20]. Per ogni finestra, oltre al periodo, è stata calcolata anche l'ampiezza, in seguito le singole finestre sono state analizzate alla ricerca della finestra candidata da utilizzare per le elaborazioni successive. La scelta ricade sulla finestra che meglio si concatena a sé stessa. Tale pseudo-periodo candidato è caratterizzato dalla minor distanza verticale tra il primo e l'ultimo campione e dalla minore differenza tra le pendenze di inizio e fine periodo. Un esempio di finestra candidata è riportata in Figura 3.10.

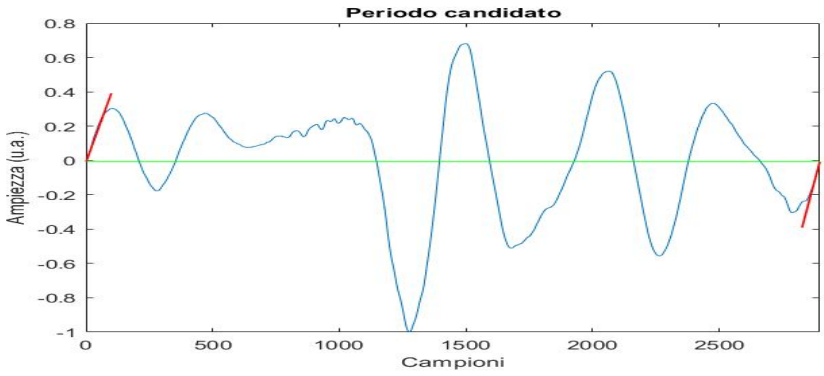

Figura 3.10: Esempio di pseudo-periodo candidato in blu. La linea verde congiunge il primo campione con l'ultimo e le linee rosse rappresentano le derivate iniziali e finali.

Dopo aver scelto lo pseudo-periodo candidato, lo si concatena alterando la sua lunghezza e la sua ampiezza in accordo con la sequenza generata.
La lunghezza del periodo candidato è modificata attraverso la funzione Matlab "linspace" che opera un ricampionamento lineare sul vettore del periodo. Dopo aver modificato il periodo si cambia l'ampiezza, normalizzando prima il vettore dello pseudo-periodo candidato per il suo valore di ampiezza picco-picco, per poi moltiplicarlo per il nuovo valore di ampiezza:

$$y_1(t) = \frac{y_0(t)A_1}{A_0} \tag{3.2}$$

dove $y_1(t)$ è il vettore dello pseudo-periodo candidato con i nuovi valori di periodo e ampiezza, $y_0(t)$ è il vettore dello pseudo-periodo candidato con il nuovo valore di periodo, A_1 è il valore di ampiezza picco-picco generato dal Monte Carlo con l'algoritmo di Metropolis-Hastings, A_0 è il valore di ampiezza picco-picco precedentemente misurato.
Infine i nuovi frame sono stati concatenati uno dopo l'altro generando un segnale artificiale.
Un confronto tra il segnale vocale registrato da un paziente patologico non parkinsoniano e il segnale vocale artificiale, nel dominio del tempo, è riportato in Figura 3.11, mentre una loro rappresentazione tempo-frequenza è riportata in Figura 3.12.

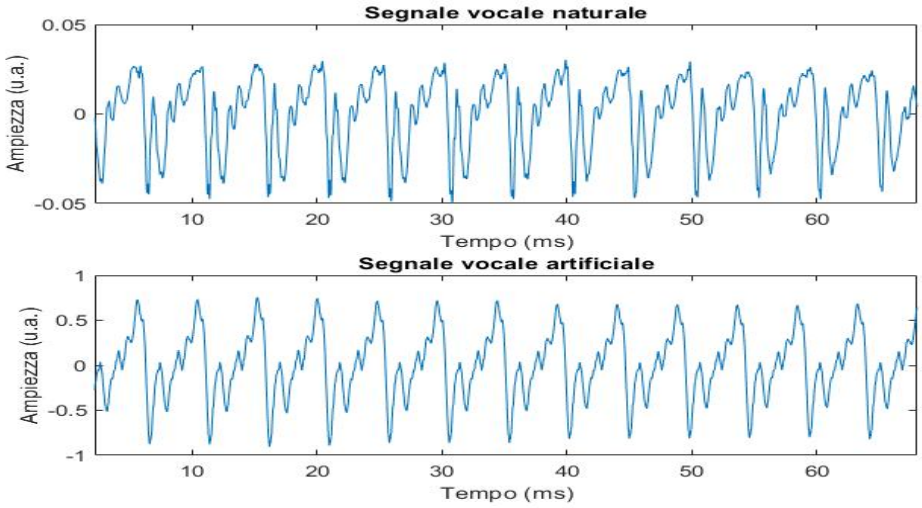

Figura 3.11: In alto è raffigurato il segnale vocale registrato da un soggetto patologico non parkinsoniano e in basso il segnale vocale artificiale ottenuto con il metodo dello pseudo-periodo candidato nel dominio del tempo.

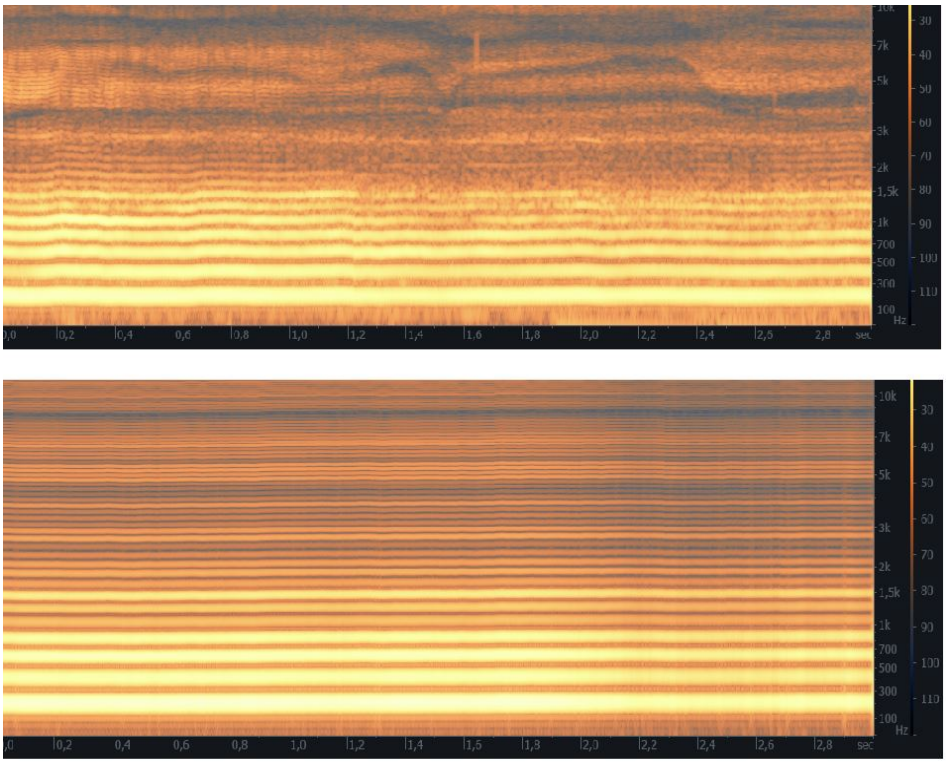

Figura 3.12: Rappresentazione tempo-frequenza di un segnale vocale registrato da un soggetto patologico non parkinsoniano in alto e rappresentazione tempo-frequenza di un segnale vocale artificiale ottenuto con il metodo dello pseudo-periodo candidato in basso.

Dalla rappresentazione tempo-frequenza riportata in Figura 3.12, si può osservare come nel segnale vocale registrato dal paziente ci siano modulazioni dello spettro nel tempo. Invece nel segnale artificiale si nota come le variazioni spettrali siano del tutto assenti, che è una situazione impossibile per il sistema fisico reale che si sta studiando. Quindi un limite di questo metodo consiste nel fatto che la vocale artificiale non presenta gli effetti del filtro tipico del tratto vocale.

3.3.2 Il metodo del "liftering"

Per superare il limite del metodo dello paseudo-periodo candidato, si è cercato di restituire la variazione della modulazione spettrale tramite un processo di separazione cepstrale noto come "liftering". Con il termine liftering, parola ottenuta anagrammando la parola filtering, si intende un filtraggio nel dominio del cepstrum [21]. L'operazione di liftering consiste nel moltiplicare

il cepstrum per un'opportuna finestra in modo da separare le basse quefrenze, dove risiede l'informazione dell'inviluppo dello spettro, dalle alte quefrenze, che caratterizzano la componente armonica del segnale. In base alla finestra selezionata si classifica il liftering in: liftering della parte alta della quefrency e liftering della parte bassa della quefrency.

Il liftering della parte alta della quefrency permette di evidenziare il segnale di eccitazione che arriva dalle corde vocali. La finestra da utilizzare è la seguente:

$$w_l[n] = \begin{cases} 0 & \text{se } 0 \leq n \leq Q_c \\ 1 & \text{se } Q_c < n \leq N/2 \end{cases} \tag{3.3}$$

dove Q_c è la quefrency di taglio e N è la lunghezza del cepstrum.

In alto nella Figura 3.13 è rappresentato il cepstrum di un tratto di segnale vocale in verde e la finestra tipica del liftering della parte alta della quefrency in blu. In basso nella Figura 3.13 è mostrato lo spettro in scala logaritmica del cestrum filtrato. Si osservi come lo spettro, tipico del segnale di eccitazione, ha picchi che decadono verso le alte frequenze e porta l'informazione riguardo l'aria emessa dai polmoni e riguardo il tono.

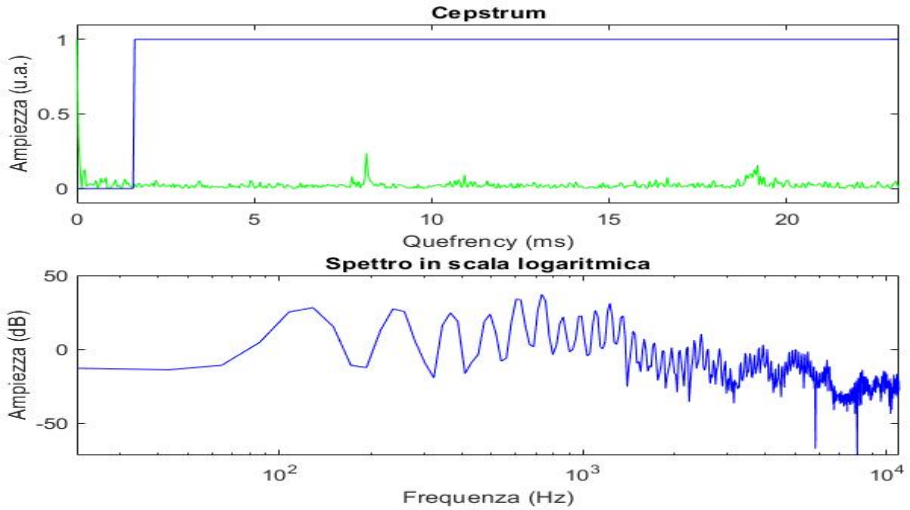

Figura 3.13: Nel grafico in alto è riportato il cepstrum in verde e la finestra usata per il liftering della parte alta della quefrency in blu. In basso lo spettro in scala logaritmica del cepstrum ottenuto dopo il liftering della parte alta della quefrency.

Il liftering della parte bassa della quefrency mette in rilievo la funzione di trasferimento del tratto vocale e quindi permette di separare le formanti. La finestra da utilizzare è la seguente:

$$w_h[n] = \begin{cases} 1 & \text{se } 0 \leq n \leq Q_c \\ 0 & \text{se } Q_c < n \leq N/2 \end{cases} \tag{3.4}$$

dove Q_c è la quefrency di taglio e N è la lunghezza del cepstrum.

Un esempio di cepstrum e di finestra caratteristica del liftering della parte bassa della quefrency è mostrato in alto nella Figura 3.14. In basso nella Figura 3.14 è mostrato lo spettro in scala logaritmica del cestrum filtrato .

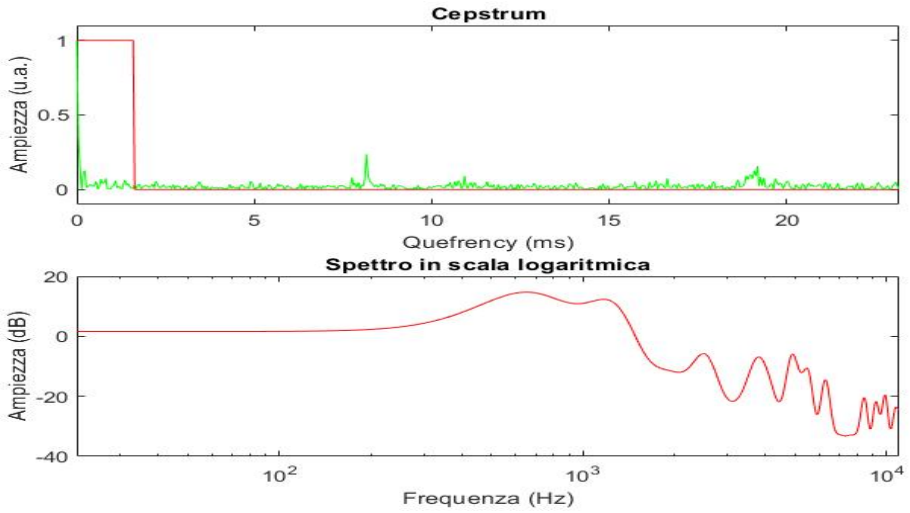

Figura 3.14: Nel grafico in alto è riportato il cepstrum in verde e la finestra usata per il liftering della parte bassa della quefrency in rosso. Nel grafico in basso è rappresentato lo spettro in scala logaritmica del cepstrum ottenuto dopo il liftering della parte bassa della quefrency.

Lo spettro in scala logaritmica del cepstrum originale è riportato in Figura 3.15.

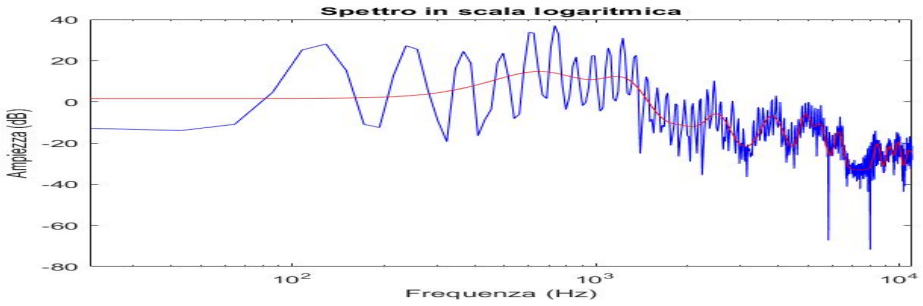

Figura 3.15: Lo spettro in scala logaritmica del cepstrum originale in blu e lo spettro in scala logaritmica del cepstrum ottenuto dopo il liftering della parte bassa della quefrency in rosso.

La modifica cepstrale è stata eseguita sui segnali artificiali ottenuti con il metodo dello pseudo-periodo candidato. L'algoritmo per produrre segnali vocali artificiali con il metodo del "liftering" è descritto di seguito. Il segnale vocale reale e il segnale artificiale campionati ad una frequenza di 44100 Hz, sono stati sotto-campionati di un fattore 2. Successivamente sono state prese finestre della durata di 92.8 ms, che corrispondono a 2048 campioni e sovrapposte per 2 ms. Per ogni singolo frame è stato calcolato il cepstrum che è stato moltiplicato per la finestra tipica del liftering della parte bassa della quefrency. La quefrenza di taglio è stata scelta pari ad un quinto del valore di quefrency a cui corrisponde il picco cepstrale. Il cesptrum non filtrato del segnale artificiale è stato normalizzato per il suo cepstrum filtrato e poi moltiplicato per il cepstrum filtrato del segnale originale del paziente:

$$C_2(\tau) = \frac{C_1(\tau)C_{ho}(\tau)}{C_{h1}(\tau)} \tag{3.5}$$

dove $C_2(\tau)$ è il cepstrum del nuovo segnale vocale artificiale, $C_1(\tau)$ è il cepstrum del segnale vocale artificiale, $C_{h0}(\tau)$ è il cepstrum filtrato del segnale vocale reale e $C_{h1}(\tau)$ è il cepstrum filtrato del segnale vocale artificiale. In questo modo si sono ottenuti cepstrum di segnali artificiali che hanno mantenuto le proprie caratteristiche dei segnali di eccitazione ma che presentano funzioni di trasferimento del tratto vocale dei soggetti reali.

Successivamente per ottenere i nuovi segnali artificiali si è passati dal dominio del cepstrum al dominio del tempo con le seguenti operazioni:

1. $X(f) = \mathcal{F}^{-1}\{C(\tau)\}$, dove $X(f)$ è lo spettro in scala logaritmica, $\mathcal{F}^{-1}$ è la trasformata inversa di Fourier e $C(\tau)$ è il cepstrum;
2. $X_1(f) = 10^{X(f)}$, dove $X_1(f)$ è lo spettro in scala lineare;
3. $x(t) = \mathcal{F}^{-1}\{X_1(f)\}$, dove $x(t)$ è il segnale nel dominio del tempo.

L'ultimo passaggio consiste nel modificare le ampiezze del segnale usando l'espressione 3.2 in accordo con le sequenze generate dall'algoritmo MC.
Un confronto tra il segnale vocale registrato da un paziente patologico non parkinsoniano e il segnale vocale artificiale ottenuto con il metodo del "liftering" nel dominio del tempo è riportato in Figura 3.16, mentre una loro rappresentazione tempo-frequenza è riportata in Figura 3.17.

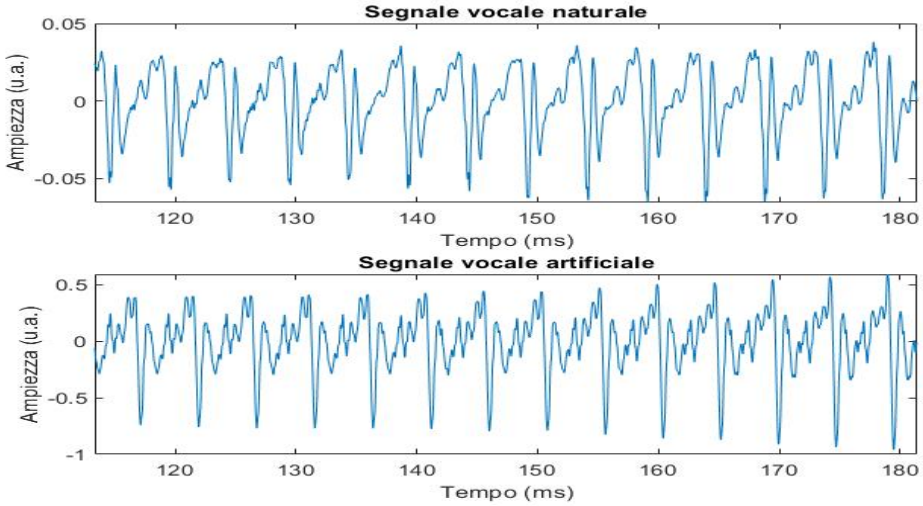

Figura 3.16: In alto è raffigurato il segnale vocale registrato da un soggetto patologico non parkinsoniano e in basso il segnale vocale artificiale ottenuto con il metodo del "liftering".

Per quanto riguarda gli andamenti temporali riportati in Figura 3.16, si noti come la forma d'onda che si ripete periodicamente nella parte di segnale presa in analisi è molto diversa tra il segnale vocale reale e il segnale artificiale.

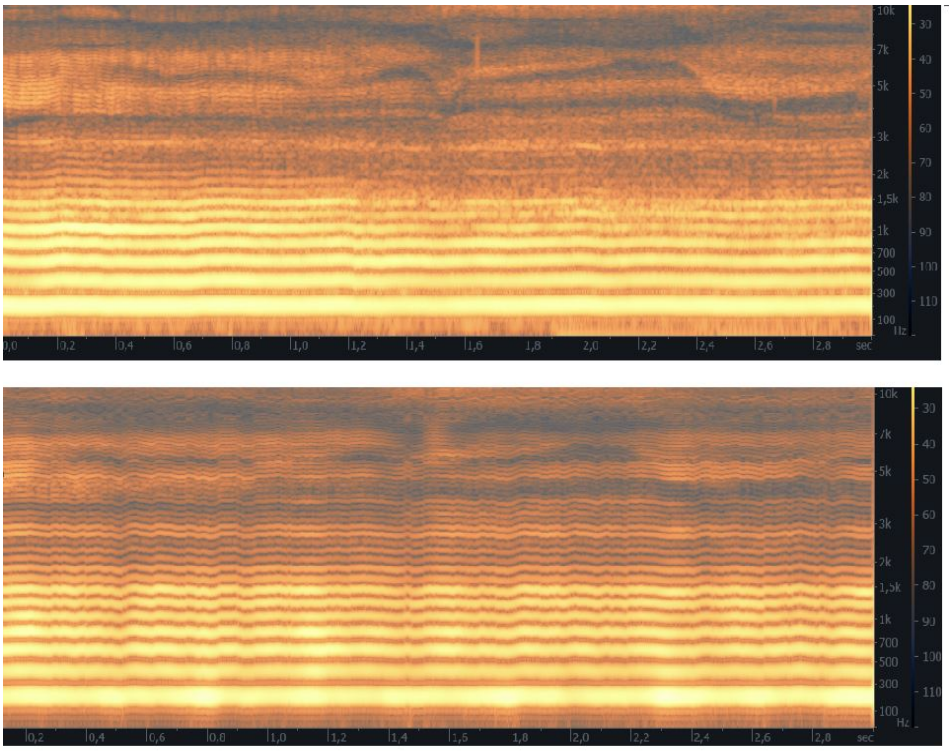

Figura 3.17: Rappresentazione tempo-frequenza di un segnale vocale registrato da un soggetto patologico non parkinsoniano in alto e rappresentazione tempo-frequenza di un segnale vocale artificiale ottenuto con il metodo del "liftering".

Come si può osservare dalle distribuzioni tempo-frequenza in Figura 3.17, nel segnale artificiale le caratteristiche dell'inviluppo spettrale sono più simili a quelle relative al segnale reale, a differenza del caso dello pseudo-periodo candidato.

3.3.3 Il metodo di ricampionamento dei periodi candidati

Un altro metodo proposto per produrre segnali artificiali con caratteristiche di periodo e ampiezza generate con l'algoritmo di Metropolis-Hastings e che presentano gli effetti del tratto vocale è il metodo di ricampionamento dei periodi originali. Si consideri il segnale vocale reale. Questo è stato diviso in pseudo-periodi e non è stato scelto uno pseudo-periodo candidato ma ad ogni pseudo-periodo sono state modificate le caratteristiche di ampiezza e periodo. La modifica della lunghezza dei singoli periodi avviene tramite

un ricampionamento lineare utilizzando la funzione Matlab "linspace" mentre la modifica delle ampiezze avviene tramite la formula (3.2). Il segnale vocale registrato da una paziente patologico e non affetto dalla malattia di Parkinson e il corrispondente segnale artificiale generato con il metodo di ricampionamento dei periodi originali è riportato in Figura 3.18. Si osservi come la morfologia della forma d'onda contenuta in un periodo si mantiene sostanzialmente uguale tra i due segnali.

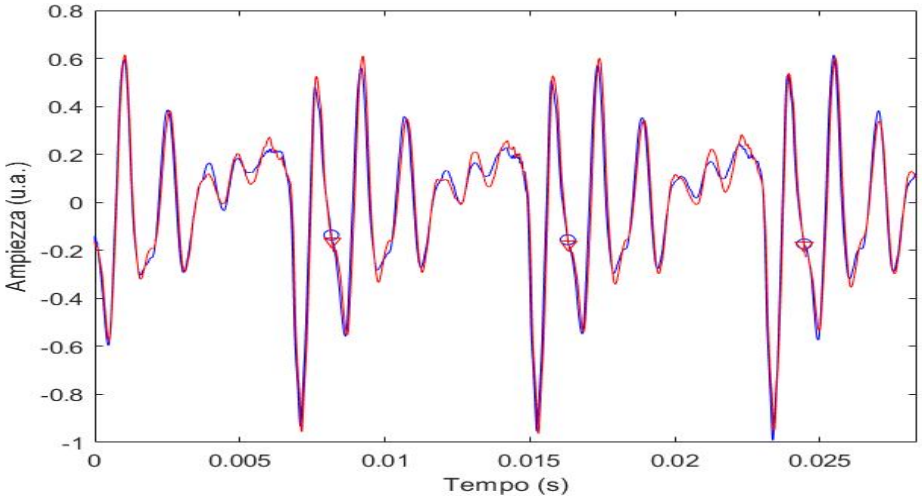

Figura 3.18: In rosso è raffigurato il segnale vocale registrato da un soggetto patologico non parkinsoniano e in blu il segnale vocale artificiale ottenuto con il metodo di ricampionamento dei periodi originali. Il triangolo indica gli pseudo-periodi per il segnale naturale ed il cerchio gli pseudo-periodi per il segnale artificiale.

Degli stessi segnali rappresentati in Figura 30 è stata fatta un'analisi tempo-frequenza, le cui rappresentazioni sono state riportare in Figura 3.19.

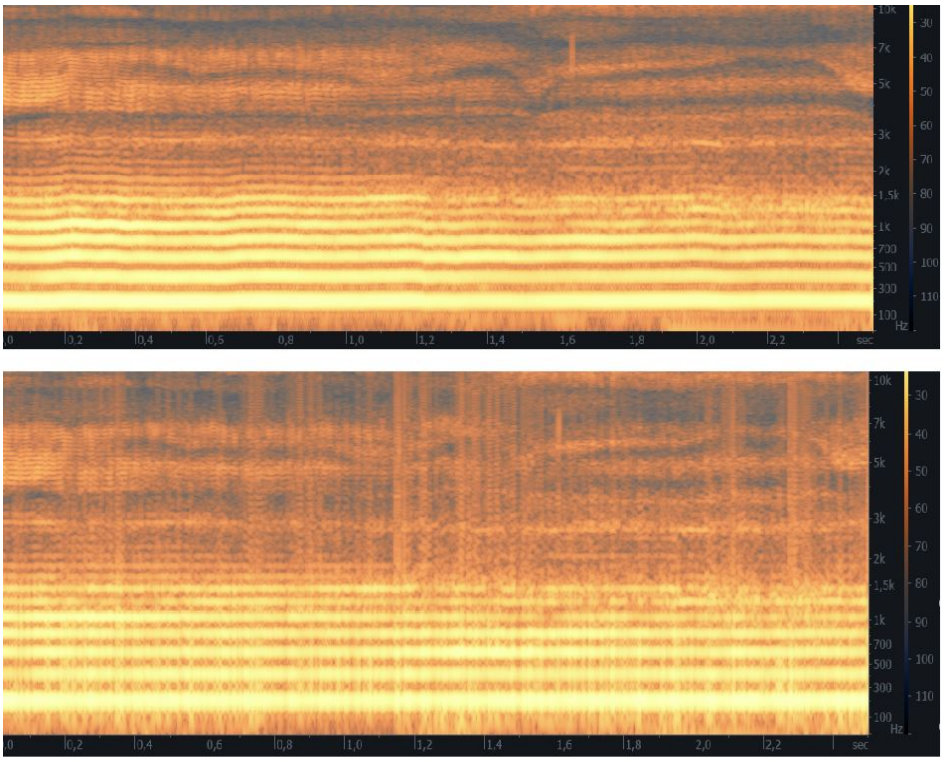

Figura 3.19: Rappresentazione tempo-frequenza di un segnale vocale registrato da un soggetto patologico non parkinsoniano in alto e rappresentazione tempo-frequenza di un segnale vocale artificiale ottenuto con il metodo di ricampionamento dei periodi originali.

Dalle rappresentazioni tempo-frequenza risulta evidente che il segnale vocale artificiale ha mantenuto le stesse caratteristiche spettrali a banda larga. Quindi il segnale vocale artificiale generato con il metodo del ricampionamento dei segnali originali è simile ad una reale situazione fisica.

3.3.4 Il metodo di trasformazione del cepstrum

I segnali vocali artificiali che sono stati prodotti hanno periodi e ampiezze noti. Il passo successivo è stato quello di studiare gli eventuali cambiamenti delle sequenze di CPPS tra i segnali reali e artificiali che potrebbero essere stati apportati durante produzione delle nuove vocali.
L'analisi delle sequenze del parametro cepstrale è stata eseguita come segue. Si consideri un segnale vocale reale che è stato utilizzato per generare nuovi segnali vocali artificiali utilizzando i tre metodi descritti precedentemente. Sia per i segnali reali, sia per i segnali artificiali sono state calcolate

le sequenze di CPPS e le loro distibuzioni statistiche discrete, utilizzando la funzione Matlab "histcounts" ed il numero ottimale di bin è stato scelto con il criterio Friedman-Draconis, specificando la flag "fd" tra gli argomenti della funzione "histcounts". Le distribuzioni provenienti dai segnali reali e le distribuzioni provenienti dai segnali artificiali sono state confrontate con il test di Kolmogorv-Smirnov.
Un esempio di confronto tra la distribuzione di CPPS ottenuta da un segnale registrato da un soggetto sano e da un corrispondente segnale artificiale prodotto con il metodo dello pseudo-periodo candidato è mostrato in Figura 3.20, dove si può osservare la differenza tra le due distribuzioni.

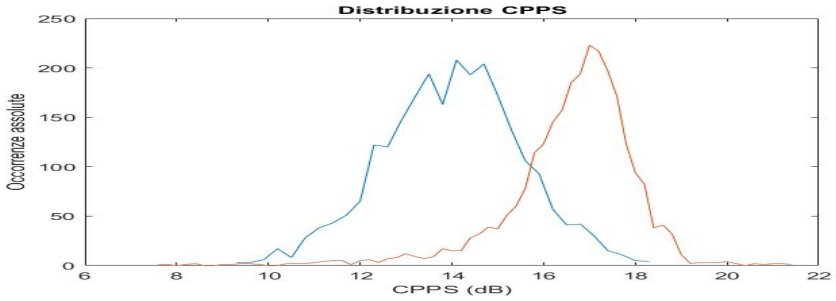

Figura 3.20: Distribuzione di CPPS calcolata dal segnale registrato da un soggetto sano in blu e distribuzione di CPPS calcolata dal segnale artificiale prodotto con il metodo dello pseudo-periodo candidato in rosso.

Un altro esempio è stato riportato per la distribuzione di CPPS ottenuta da un segnale registrato da un soggetto sano e da un corrispondente segnale artificiale prodotto con il metodo del ricampionamento del periodo originale (Figura 3.21). Anche in questo caso si può notare la differenza tra le due distribuzioni.

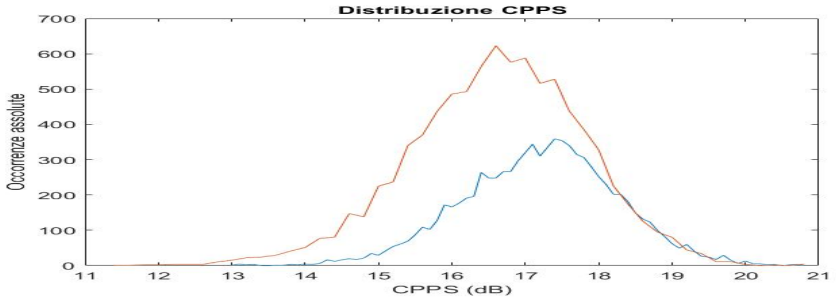

Figura 3.21: Distribuzione di CPPS calcolata dal segnale registrato da un soggetto sano in blu e distribuzione di CPPS calcolata dal segnale artificiale prodotto con il metodo del ricampionamento del periodo originale in rosso.

Le distribuzioni di CPPS calcolate dai segnali artificiali prodotti con il metodo dello pseudo-periodo candidato, con il metodo del "liftering" e con il metodo di ricampionamento del segnale originale modificano le distribuzioni di CPPS calcolate dai segnali vocali reali. Per questo motivo è stato implementato un metodo di trasformazione del cepstrum che ha permesso di generare segnali vocali non solo con sequenze di periodo e ampiezza note ma anche con sequenze di CPPS note.
Questo metodo si basa su un algoritmo che ha permesso di imporre i valori di CPPS desiderati ad un segnale vocale arbitrario. I passi principali sono i seguenti:

1. Si sottocampiona il segnale vocale di un fattore 2, si sceglie una finestra di analisi di 1024 campioni (46.4 ms) e un overlap di 44 campioni (2 ms).

2. Si moltiplica ogni frame per una finestra di Hamming delle stesse dimensioni del frame.

3. Si calcola il cepstrum lineare per ogni frame come:

$$C(\tau) = \left|\mathcal{F}\left\{20 \cdot log\left(|\mathcal{F}\{f(t)\}|\right)\right\}\right| \tag{3.6}$$

   dove $\mathcal{F}$ è la trasformata di Fourier, $x(t)$ è il segnale in funzione del tempo ottenuto al passo 2.

4. Si stima la posizione del picco cepstrale.

5. Si modifica il valore del picco cepstrale e i valori del cepstrum che si trovano 10 bin precedenti e 10 bin successivi la posizione del picco cepstrale con la seguente formula:

$$C_1(\tau) = C_0(\tau) \cdot 10^{(\triangle B/20)} \tag{3.7}$$

   dove $C_1(\tau)$ è il vettore del picco cepstrale trasformato, $C_0(\tau)$ è il vettore del picco cepstrale originale e $\triangle B = CPPS_1 - CPPS_0$ con $CPPS_1 =$ valore di CPPS desiderato e $CPPS_0 =$ valore di CPPS originale.

Al fine di dimostrare l'efficacia dell'algoritmo proposto, si riporta l'esempio per il caso in cui si vogliono aumentare tutti i valori di CPPS estratti da un segnale vocale di una quantità nota e pari a 6 dB ($\triangle B = 6\,dB$). Dopo aver applicato l'algoritmo ad un segnale vocale registrato da un soggetto sano, è stato fatto un confronto tra il cepstrum del segnale reale e il cepstrum trasformato. I due cepstra sono riportati in Figura 3.22. Nel grafico in alto si può osservare come i due cepstra cambiano solamente nell'intervallo

selezionato dall'algoritmo. In particolare si noti nel grafico in basso come effettivamente i valori dei due picchi cepstrali differiscono di una quantità pari a 6 dB, senza alterare la forma della distribuzione.

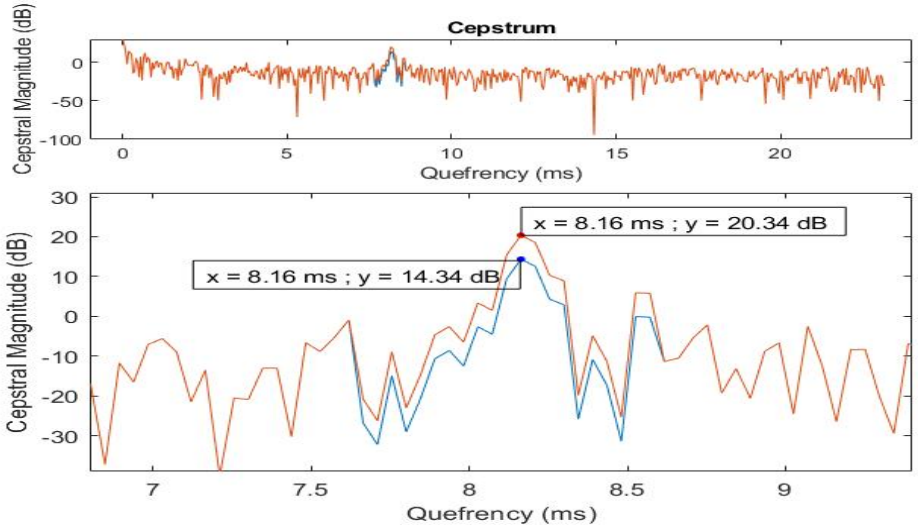

Figura 3.22: In alto grafico del cestrum originale in blu e grafico del cepstrum trasformato in rosso. In basso porzione dei due cepstra che contengono i picchi cepstrali.

Successivamente sia per i cepstra originali, sia per i cepstra trasformati è stato calcolato il CPPS e di queste sequenze sono state ottenute le loro distribuzioni con la funzione Matlab "histcounts". Si può vedere dalla Figura 3.23 come la distribuzione di CPPS ottenuta dai cepstra trasformati è la distribuzione di CPPS ottenuta dai cepstra originali ma traslata di 6 dB.

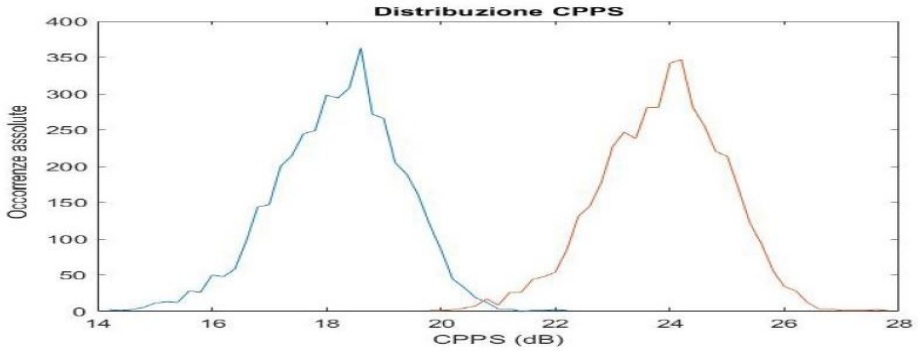

Figura 3.23: Distribuzione CPPS ottenuti dai cepstra originali in blu e distribuzione CPPS ottenuti dai cepstra trasformati in rosso.

Inoltre mediante uno studio nel dominio della frequenza sono state analizzate le modifiche sullo spettro del segnale a seguito della trasformazione del cepstrum. Lo studio ha evidenziato come lo spettro calcolato dal cepstrum trasformato non presenta picchi spettrali spuri ma mostra una struttura armonica più definita rispetto lo spettro calcolato dal cepstrum originale. Questo è dovuto al fatto che il valore del picco cepstrale è stato aumentato rispetto a quello assunto nel cepstrum originale. Un esempio di spettro in scala logaritmica è in Figura 3.24.

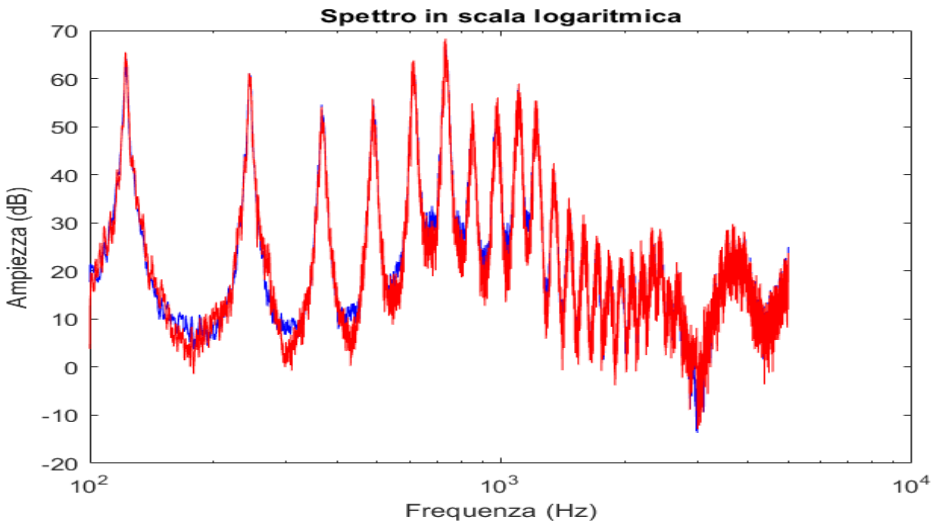

Figura 3.24: Spettro in scala logaritmica calcolato dal cepstrum originale in blu e spettro in scala logaritmica calcolato dal cepstrum trasformato in rosso.

Capitolo 4

Valutazione dell'errore dell'algoritmo di estrazione

4.1 Introduzione

In questa prima parte del capitolo sarà descritto come è stata ottenuta la stima dell'errore dell'algoritmo di estrazione dei parametri a partire dai segnali vocali artificiali prodotti con i metodi trattati nel capitolo precedente.
Nella seconda parte saranno riportati i valori delle stime ottenute ed una valutazione delle performance dei metodi proposti.

Per ogni singola ripetizione della vocale sostenuta /a/ eseguita dal soggetto, sono state generate 10 nuove vocali artificiali a partire da sequenze di periodo e ampiezza noti. Dato che il subset scelto per ogni classe è formato da 10 soggetti e per ogni soggetto sono state registrate 3 vocali sostenute /a/, da un totale di 30 segnali vocali per ogni classe si è arrivati ad un totale di 300 vocali artificiali. Come visto nell'algoritmo in Sezione 3.2 che permette di generare sequenze di periodi, ampiezze e CPPS mediante la simulazione Monte Carlo, queste sequenze possono essere ottenute sia con il metodo del passo indipendente, sia con il metodo del passo dipendente. Quindi gli algoritmi di generazione dei segnali artificiali, teoricamente, dovrebbero permette di ottenere, per ogni classe, un totale di:

- 300 vocali artificiali generati con il metodo dello pseudo-periodo candidato a partire da sequenze di periodo e ampiezza basate sul passo indipendente;
- 300 vocali artificiali generati con il metodo dello pseudo-periodo candidato a partire da sequenze di periodo e ampiezza basate sul passo dipendente;

- 300 vocali artificiali generati con il metodo del "liftering" a partire da sequenze di periodo e ampiezza basate sul passo indipendente;
- 300 vocali artificiali generati con il metodo del "liftering" a partire da sequenze di periodo e ampiezza basate sul passo dipendente;
- 300 vocali artificiali generati con il metodo del ricampionamento del periodo originale a partire da sequenze di periodo e ampiezza basate sul passo indipendente;
- 300 vocali artificiali generati con il metodo del ricampionamento del periodo originale a partire da sequenze di periodo e ampiezza basate sul passo dipendente.

Nel capitolo precedente è stato evidenziato come la simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings non sempre è riuscito a generare le sequenze desiderate. Questo è dovuto al fatto che, in alcuni rari casi, il segnale registrato dai pazienti parkinsoniani e dai pazienti patologici e non affetti dalla malattia di Parkinson è estremamente corto e rumoroso. Questo fatto provoca alcuni problemi soprattutto nella scelta del numero di bin con il criterio di Friedman - Draconis che può produrre distribuzioni poco rappresentative. In questo caso il metodo fallisce in quanto non riesce a trovare una sequenza temporale che sia statisticamente compatibile con la distribuzione originale. In questi casi non è stato possibile produrre un segnale artificiale. Il numero reale delle vocali artificiali prodotte dall'algoritmo sono riassunte nella Tabella 4.1 per i soggetti HE, nella Tabella 4.2 per i soggetti PA, nella Tabella 4.3 per i soggetti PD.

Tabella 4.1: Numero dei segnali vocali artificiali prodotti per la classe HE.

Metodo	Passo indipendente	Passo dipendente
Pseudo-periodo	157	300
"Liftering"	300	300
Ricampionamento del periodo originale	272	300

Tabella 4.2: Numero dei segnali vocali artificiali prodotti per la classe PA.

Metodo	Passo indipendente	Passo dipendente
Pseudo-periodo	212	300
"Liftering"	238	300
Ricampionamento del periodo originale	300	300

Tabella 4.3: Numero dei segnali vocali artificiali prodotti per la classe PD.

Metodo	Passo indipendente	Passo dipendente
Pseudo-periodo	221	298
"Liftering"	238	300
Ricampionamento del periodo originale	300	300

L'errore dell'algoritmo di estrazione dei parametri temporali è stato stimato come mostrato in Figura 4.1.

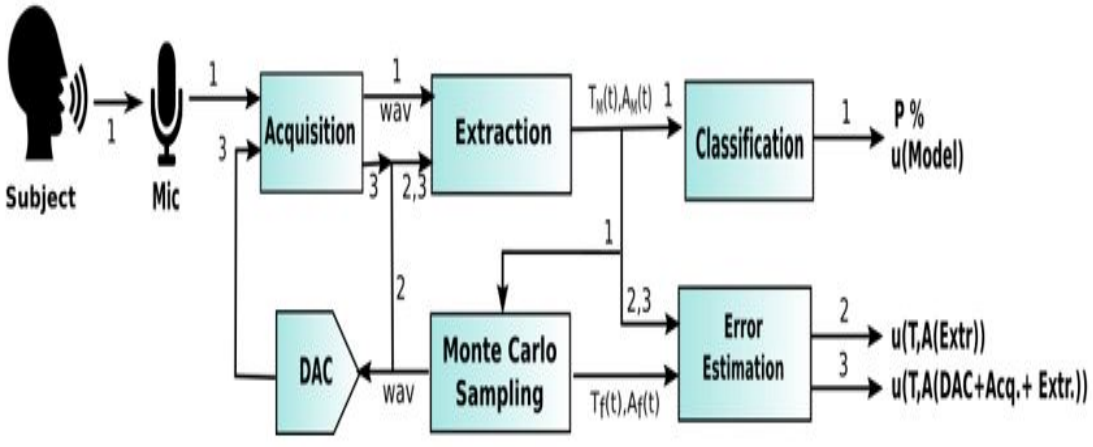

Figura 4.1: Schema per la stima errore dell'algoritmo di estrazione dei parametri temporali.

La numerazione dei passi seguenti fanno riferimento ai numeri in Figura 4.1:

1. la vocale sostenuta /a/ eseguita da un soggetto è stata registrata mediante un microfono in aria. Dal segnale vocale sono state estratte le sequenze di periodo T_M e di ampiezza A_M;

2. le sequenze T_M e A_M sono state utilizzate per generare sequenze di periodi noti T_f e di ampiezze note A_f tramite una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings. In seguito, nuovi segnali artificiali sono stati prodotti a partire dalle sequenze T_f e A_f. Infine, dalle vocali artificiali sono stati estratti i parametri temporali, che insieme alle sequenze note T_f e A_f, hanno permesso di stimare l'errore come:
$$E_r = \left(\frac{V_m - V_e}{V_e}\right) \cdot 100 \tag{4.1}$$
dove V_m è il parametro estratto dal segnale vocale artificiale, V_e è il parametro calcolato a partire dalle sequenze di periodi e ampiezza noti.

   In Tabella 4.4 e in Tabella 4.5 sono state riportate le medie degli errori ottenuti dai soggetti PD e relativi ai parametri temporali e sono stati evidenziati in giallo i valori minori per ogni parametro.

Per quanto riguarda i parametri HNR, F_0, RMS e CPPS, l'errore è stato calcolato come:
$$E_r = \left(\frac{V_a - V_n}{V_n}\right) \cdot 100 \tag{4.2}$$
dove V_a è il parametro dal segnale vocale artificiale, V_a è il parametro estratto dal segnale vocale reale.

Tabella 4.4: Media degli errori per la classe PD e relativi ai parametri per valutare la stabilità in frequenza estratti dai segnali artificiali prodotti con tutti metodi. In giallo i valori minori per ogni parametro.

Metodo	Jitt	Jita	RAP	PPQ	vF_0
Pseudo-periodo e passo indipendente	20.5	20.7	33.9	20.4	2.5
Pseudo-periodo e passo dipendente	31.1	31.2	40.6	27.4	8.1
Liftering e passo indipendente	18.9	19.1	30.8	19.4	3.5
Liftering e passo dipendente	31.6	31.2	40.6	27.2	7.8
Ricampionamento periodo originale e passo indipendente	28.9	28.8	37.2	17.4	6.5
Ricampionamento periodo originale e passo dipendente	12.2	12.2	15.6	11.2	12.2

Tabella 4.5: Media degli errori per la classe PD e relativi ai parametri per valutare la stabilità in ampiezza dai segnali artificiali prodotti con tutti metodi. In giallo i valori minori per ogni parametro.

Metodo	Shim	ShdB	APQ	vAm
Pseudo-periodo e passo indipendente	8.6	8.7	4.3	0.4
Pseudo-periodo e passo dipendente	13.9	14.3	8.4	3.1
Liftering e passo indipendente	6.7	6.8	3.4	0.3
Liftering e passo dipendente	12.5	13.1	7.9	3.3
Ricampionamento periodo originale e passo indipendente	7.6	6.9	2.7	0.4
Ricampionamento periodo originale e passo dipendente	6.8	6.9	4.6	2.1

4.2 Risultati e discussione

Gli errori dell'algoritmo di estrazione dei parametri temporali, calcolati con l'espressione (4.1), sono stati utilizzati per determinare il miglior metodo di generazione dei segnali vocali. In particolare per ogni classe sono stati confrontate le medie di questi errori su tutti i soggetti, per ogni metodo, in modo da determinare quale di questi producesse l'errore medio minore. I risultati ottenuti sono riportati in Figura 4.2 per i soggetti HE, in Figura 4.3 per i soggetti PA e in Figura 4.4 per i soggetti PD.

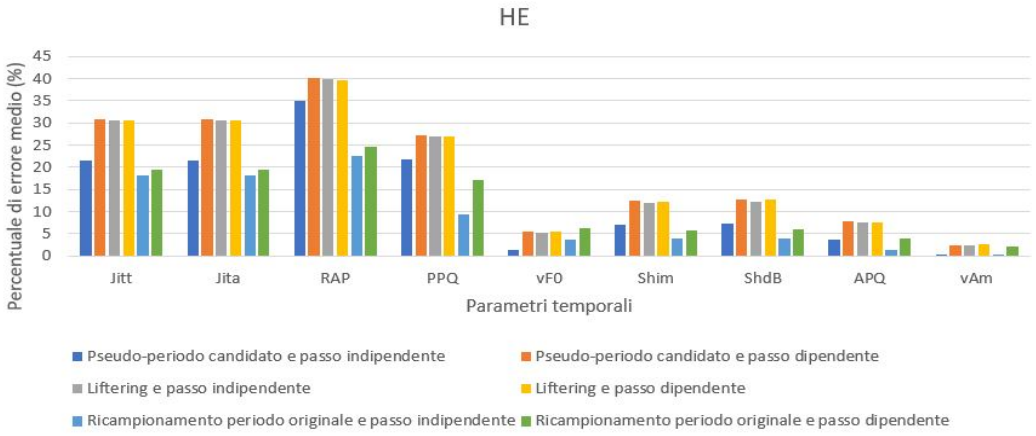

Figura 4.2: Errore medio percentuale per i soggetti HE.

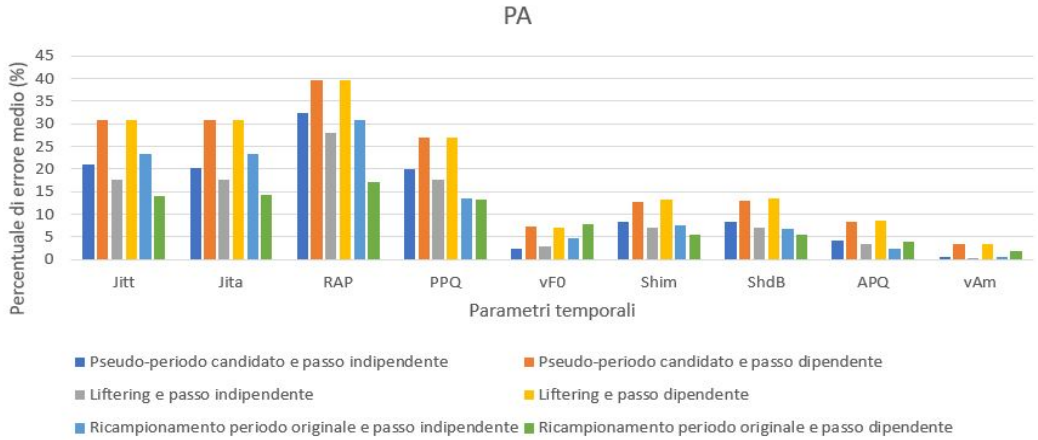

Figura 4.3: Errore medio percentuale per i soggetti PA.

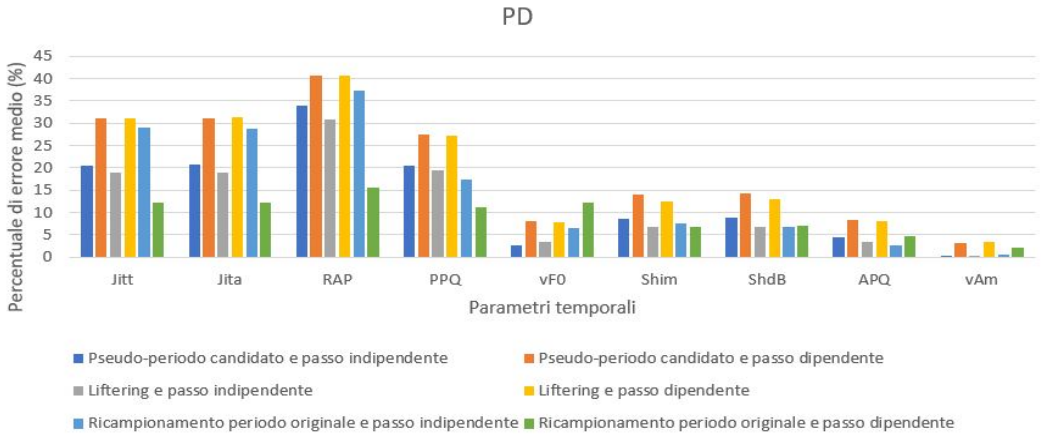

Figura 4.4: Errore medio percentuale per i soggetti PD.

Per la classe PA e per la classe PD e considerando i parametri jitter percentuale, jitter assoluto, RAP, PPQ, shimmer percentuale, shimmer assoluto, è stato osservato che gli errori minori sono legati ai segnali artificiali prodotti con il metodo di ricampionamento del periodo originale, utilizzando i periodi e le ampiezze estratte dalle registrazioni reali per definire il passo precedente nella simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings, mentre gli errori maggiori sono per i segnali artificiali prodotti con il metodo del liftering, generando sequenze di periodi e ampiezze sia con il passo indipendente sia con il passo dipendente.
Per quanto riguarda la classe HE e considerando tutti i parametri temporali, è stato osservato che gli errori minori sono legati ai segnali artificiali prodotti con il metodo di ricampionamento del periodo originale, utilizzando i periodi e le ampiezze generate casualmente ai passi precedenti per definire il passo nella simulazione Monte Carlo con l'algoritmo di Metropolis - Hastings. Queste percentuali di errori differiscono di qualche percento con quelle dovute ai segnali artificiali generati con lo stesso metodo dove però le sequenze prodotte dalla simulazione Monte Carlo con l'algoritmo di Metropolis -Hastings sono basate sul passo dipendente, rispetto agli altri metodi dai quali differiscono di decine di percento. Quindi, sulla base di queste considerazioni e considerando il fatto che l'algoritmo ibrido di Metropolis-Hastings che ha utilizzato il metodo dei passi dipendenti è riuscito a generare tutte le sequenze di periodi e ampiezze noti rispetto a quello con il metodo del passo indipendente, è stato scelto il metodo di ricampionamento dei periodi originali basato sul metodo del passo dipendente come il metodo migliore di generazione dei segnali artificiali. La numerosità del subset è stata aumentata utilizzando i parametri estratti dai segnali artificiali generati con questo metodo, fornendo

contemporaneamente una stima dei loro errori di misura.

Nel capitolo precedente è stato descritto come cambiano le distribuzione di CPPS dei segnali artificiali rispetto a quelle dei segnali registrati dai pazienti. Il passo successivo è stato quello di andare ad analizzare gli errori relativi al parametro cepstrale calcolati con l'espessione (4.2). I risultati ottenuti sono stati riportati in Figura 4.5 per i soggetti HE, in Figura 4.6 per i soggetti PA e in Figura 4.7 per i soggetti PD.

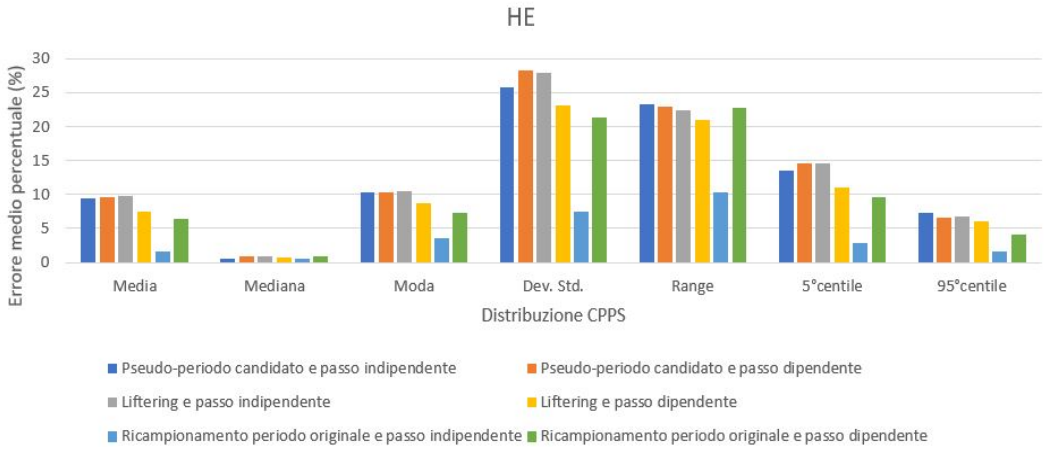

Figura 4.5: Errore medio percentuale per i soggetti HE.

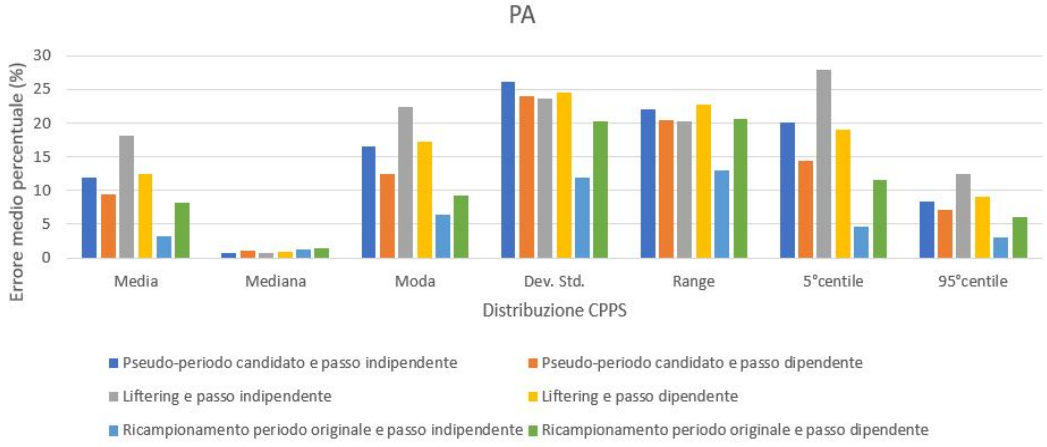

Figura 4.6: Errore medio percentuale per i soggetti PA.

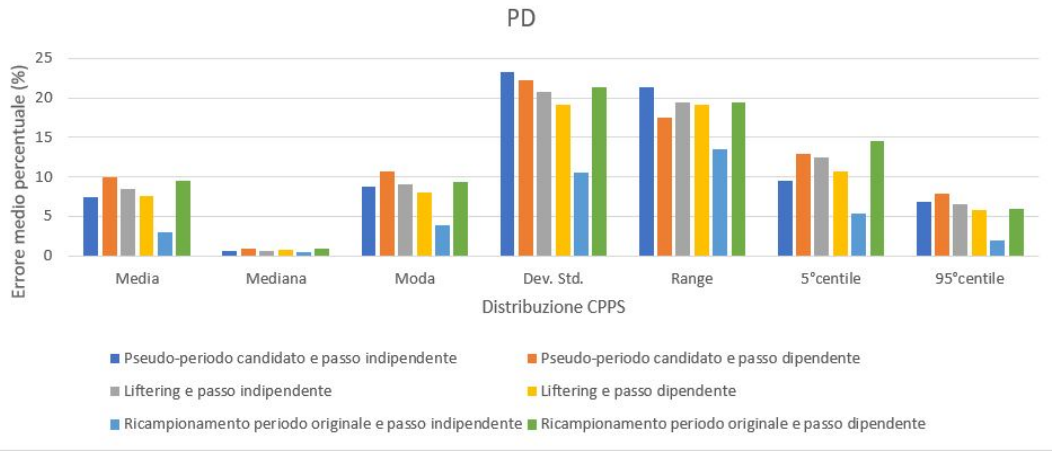

Figura 4.7: Errore medio percentuale per i soggetti PD.

L'idea è stata quella di applicare il metodo di trasformazione del cepstrum al segnale artificiale generato con il metodo di ricampionamento dei periodi originali basato sul metodo del passo dipendente per ottenere stime dell'errore di estrazione dei parametri cepstrali minori rispetto a quelle ottenute fino a questo punto.
Un confronto tra i due segnali artificiali generati, nel dominio del tempo, è mostrato in Figura 4.8 ed una loro rappresentazione tempo-frequenza è stato riportata in Figura 4.9.

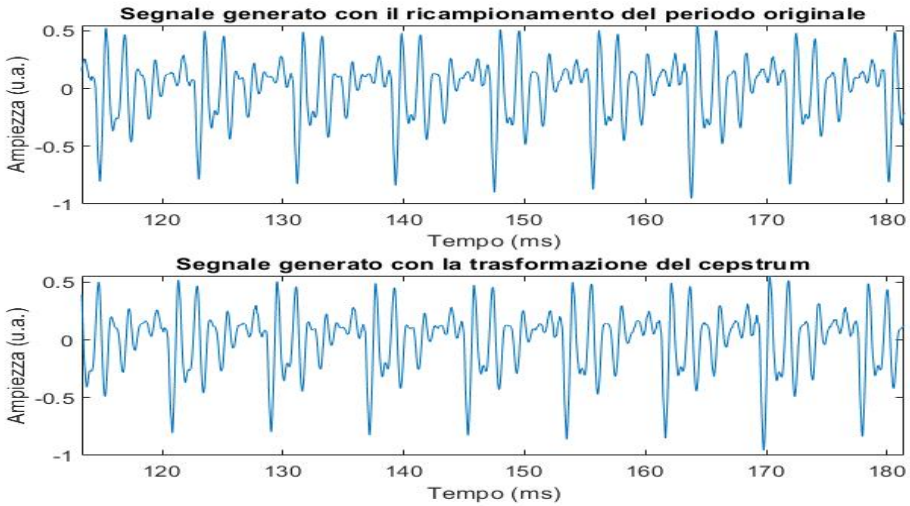

Figura 4.8: Segnale vocale artificiale generato con il metodo del ricampionamento del periodo originale in alto e segnale vocale artificiale generato con il metodo di trasformazione del cepstrum in basso.

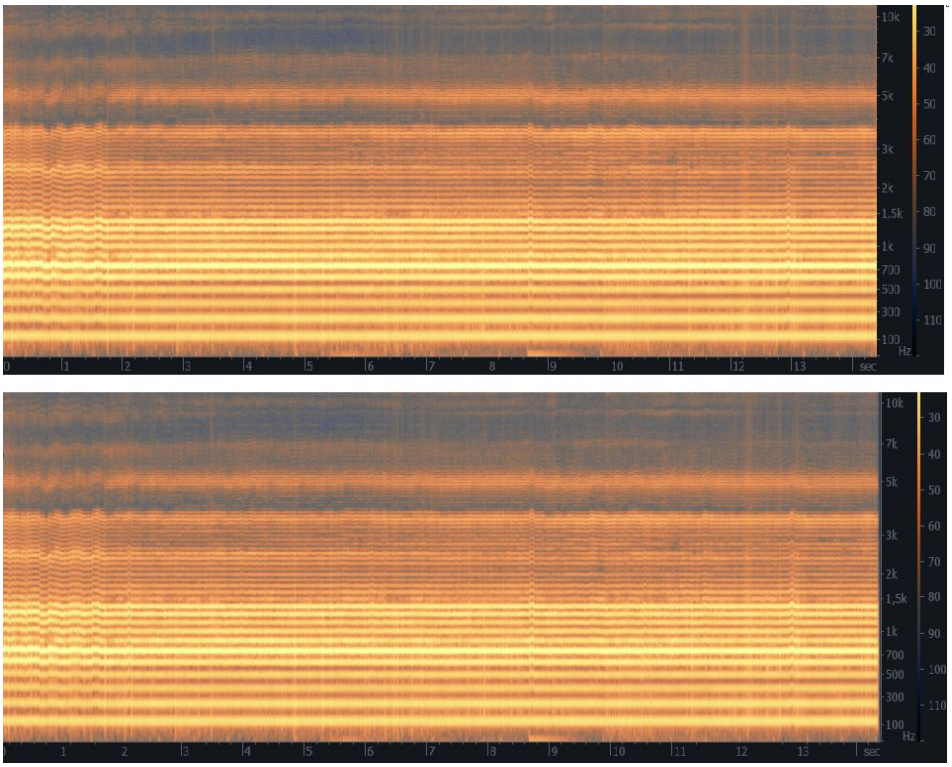

Figura 4.9: Rappresentazione tempo-frequenza del segnale vocale artificiale generato con il metodo del ricampionamento del periodo originale in alto e rappresentazione tempo-frequenza del segnale vocale artificiale generato con il metodo di trasformazione del cepstrum in basso.

Per ogni singolo segnale artificiale delle classi HE, PA e PD generato con il metodo di ricampionamento del periodo originale, sono stati prodotti 10 vocali artificiali con il metodo di trasformazione del cepstrum, generando le sequenze di CPPS note con il passo dipendente, ritenuto migliore rispetto al metodo del passo indipendente in base ai risultati dell'analisi descritta in precedenza . Per quanto riguarda il passo dipendente, il metodo è riuscito a generare un totale di 300 segnali artificiali per tutte le classi. Successivamente sono stati estratti tutti i parametri di interesse dai segnali artificiali ottenuti con il metodo di trasformazione del cepstrum che, insieme alle sequenze di periodi, ampiezze e CPPS note, sono stati utili per stimare l'errore dell'algoritmo di estrazione dei parametri.

Per quanto riguarda i parametri temporali e cepstrali, l'errore è stato calcolato seguendo lo schema in Figura 4.10.

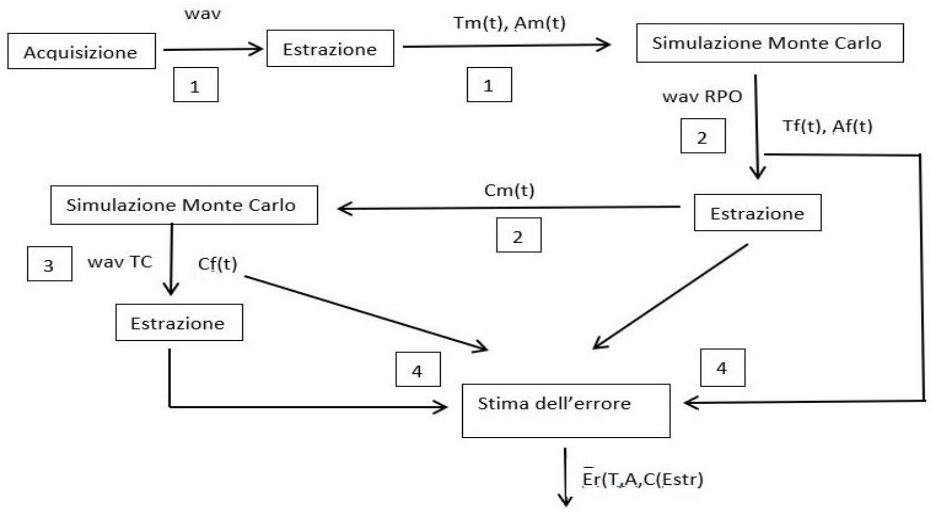

Figura 4.10: Schema per la stima errore dell'algoritmo di estrazione dei parametri temporali e CPPS.

La numerazione dei passi seguenti fanno riferimento ai numeri in Figura 4.10:

1. la vocale sostenuta /a/ eseguita da un soggetto è stata registrata mediante un microfono in aria. Dal segnale vocale sono state estratte le sequenze di periodo T_m e di ampiezza A_m.

2. Le sequenze T_m e A_m sono state utilizzate per generare sequenze di periodi noti T_f e di ampiezze note A_f tramite una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings. In seguito, nuovi segnali artificiali sono stati prodotti con il metodo di ricampionamento del periodo candidato e a partire dalle sequenze T_f e A_f. Infine da questi segnali sono stati estratti i CPPS C_m.

3. Le sequenze C_m sono state utilizzate per generare sequenze di CPPS noti C_f tramite una simulazione Monte Carlo con l'algoritmo di Metropolis-Hastings. In seguito, nuovi segnali artificiali sono stati prodotti con il metodo di trasformazione del cepstrum e a partire dalle sequenze C_f. Infine da questi segnali sono stati estratti i parametri temporali ed i CPPS.

4. I parametri temporali e i CPPS estratti al passo 3, insieme alle sequenze di periodo T_f, A_f e C_f noti, sono stati utilizzati per stimare gli errori con l'espressione (4.1), dove per i parametri temporali, v_e è il parametro calcolato a partire dalle sequenze di periodi e ampiezza che sono serviti per generare il segnale vocale artificiale con il metodo di ricampionamento del periodo originale mentre per i parametri cepstrali, v_e è il parametro calcolato a partire dalle sequenze di CPPS.

Invece per tutti gli altri parametri, l'errore è stato stimato con l'espessione (4.2).
Dall'analisi gli errori medi percentuali relativi ai parametri temporali, è emerso che questi aumentano rispetto al caso del ricampionamento del periodo originale. La differenza tra gli errori stimati per i due metodi è stata riscontrata essere maggiore per i parametri che valutano la stabilità in ampiezza. In Figura 4.11 è mostrato il caso per i soggetti PD.

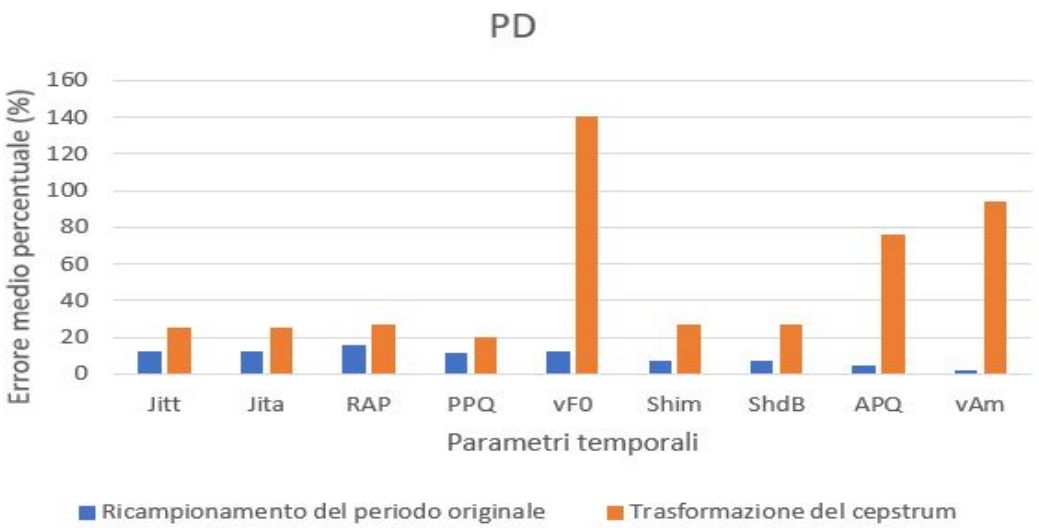

Figura 4.11: Errore medio percentuale per i soggetti PD.

Per quanto riguarda gli errori relativi dei parametri cepstrali, questi sono diminuiti rispetto il caso del ricampionamento del periodo originale per la classe HE (Figura 4.12), per la classe PA (Figura 4.13), per la classe PD (Figura 4.14).

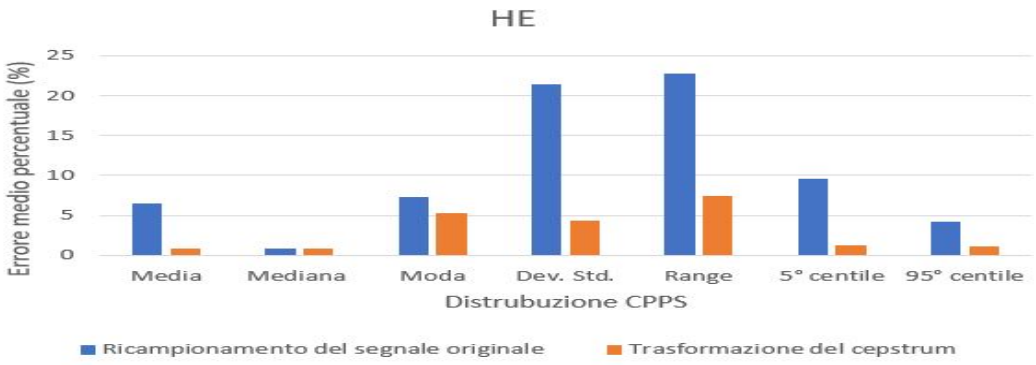

Figura 4.12: Errore medio percentuale per i soggetti HE.

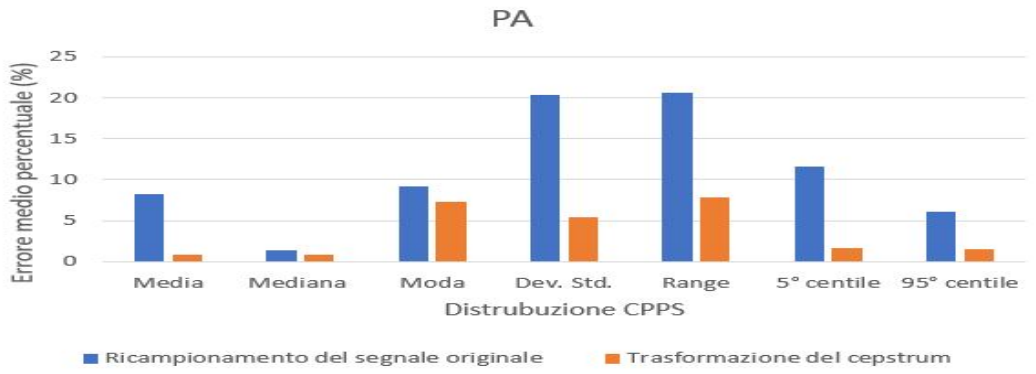

Figura 4.13: Errore medio percentuale per i soggetti PA.

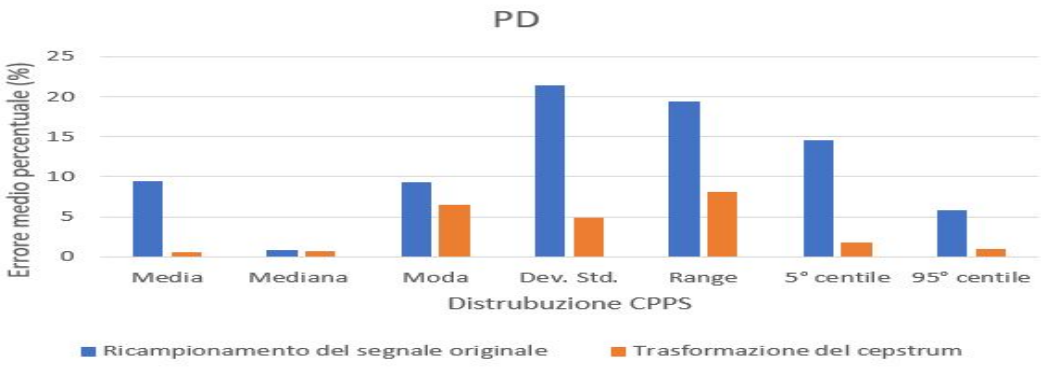

Figura 4.14: Errore medio percentuale per i soggetti PD.

Capitolo 5

Classificazione

5.1 Introduzione

Il capitolo si aprirà con la descrizione dell'algoritmo di selezione delle caratteristiche (*feature selection*) che ha permesso di andare a valutare la bontà delle feature che costituiscono il dataset. Il capitolo proseguirà trattando nel dettaglio l'algoritmo di classificazione pesato basato sulla stima dell'errore dell'algoritmo di estrazione delle feature che ha permesso di discriminare i pazienti affetti dalla malattia di Parkinson dai soggetti sani e i pazienti affetti dal morbo di Parkinson dai pazienti patologici e non parkinsoniani.

5.2 Feature selection

In generale, un algoritmo di feature selection permette di ridurre la dimensione del dataset iniziale, fornendo un subset costituito solo dalle feature rilevanti e non correlate [22]. Il metodo scelto per rimuovere le feature correlate è stato il metodo Correlation-Based Feature Selection (CBFS). Questo algoritmo utilizza la correlazione tra le variabili come misura di bontà delle feature: se la correlazione è superiore ad una certa soglia significa che la combinazione non avrà un alto potere predittivo [23]. L'indice di correlazione assume valori compresi tra -1 ed 1, dove una valore pari ad 1 indica variabili totalmente correlate, un valore pari a 0 indica variabili totalmente scorrelate ed un valore pari a -1 indica variabili totalmente correlate inversamente. Dato che l'informazione utile per valutare l'affidabilità delle feature non è se la correlazione è positiva o negativa, ma se queste sono scorrelate, allora si considera il quadrato della correlazione e di conseguenza la soglia deve avere un valore compreso tra 0 e 1.
Per i dataset delle classi HE, PA e PD, è stata calcolata la matrice di cor-

relazione r_{ij} e i corrispondenti p-value p_{ij} , utilizzando la funzione Matlab "corr". Il tipo di correlazione scelto è stato l'indice di correlazione di Pearson, definito come la covarianza di due variabili X e Y divisa per il prodotto delle deviazioni tipo delle due variabili [24]:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \tag{5.1}$$

dove σ_{XY} è la covarianza tra X e Y, σ_X è la deviazione tipo di X e σ_Y è la deviazione tipo di Y. Le matrici r_{ij} e p_{ij} sono state utilizzate per calcolare la matrice candidata come [20]:

$$candidates_{ij} = [r_{ij}^2 \cdot LM_{ij} + \overline{LM_{ij}}] \tag{5.2}$$

dove

$$LM_{ij} = \begin{cases} 1 & \text{se } p_{ij} < 0.05 \\ 0 & altrimenti \end{cases} \tag{5.3}$$

e $\overline{LM_{ij}}$ è il complemento logico di LM_{ij}.

Il valore $candidate_{ij}$ è uguale ad 1 se il p-value è superiore a 0.05, il che significa che la correlazione (o non correlazione) non è significativa. Il valore 1 corrisponde alle celle bianche nell'esempio in Figura 5.1. I valori più bassi della matrice candidata (rosso scuro nell'esempio in Figura 5.1) sono relativi a feature scorrelate e quindi non ridondanti. Le feature che hanno dato valori della matrice candidata superiori al valore di soglia, scelto pari a 0.5, non sono state considerate mentre le altre sono servite per costruire il subset che poi è stato usato per la classificazione.

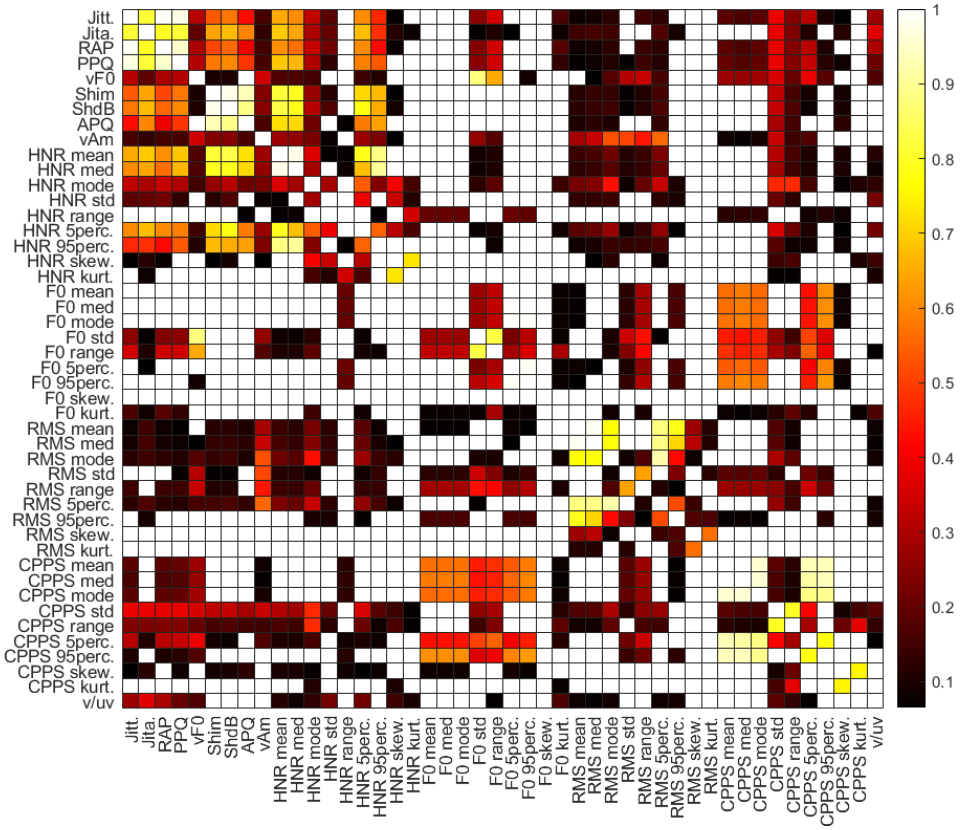

Figura 5.1: Esempio di matrice candidata.

5.3 La regressione logistica

La regressione logistica è un modello statistico che esamina la relazione tra variabili indipendenti e una variabile dipendente dicotomica (assume valori 0 o 1), fornendo una probabilità tramite una funzione logistica [25]. La funzione logistica è una curva a forma di "S" (Figura 5.2) ed è definita dall'equazione:

$$f(x) = \frac{1}{1 + e^x} \tag{5.4}$$

dove x è una combinazione lineare di feature e coefficienti.

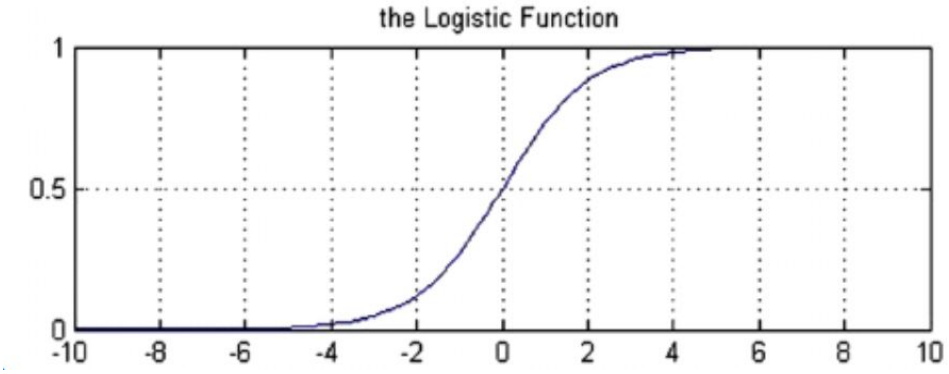

Figura 5.2: Esempio di funzione logistica ad una feature [26].

La regressione logistica è stata utilizzata come metodo di classificazione e fornisce la probabilità di appartenere alla classe dei sani secondo l'espressione:

$$P\{Healthy\} = \frac{1}{1 + e^{-(\beta_0 + \sum_i \beta_i \cdot x_i)}} \tag{5.5}$$

dove $P\{Healthy\}$ è la probabilità di appartenere alla classe dei sani e β_i sono i coefficienti della combinazione lineare con le feature x_i.
La soglia per la classificazione è stata scelta pari a $P\{Healthy\} = 0.5$.
Lo strumento che ha consentito la visualizzazione delle prestazione del classificatore è la matrice di confusione (*confusion matrix*). È una tabella dove le colonne rappresentano la classe reale e le righe rappresentano la classe predetta dal classificatore. Un esempio è riportato in Tabella 7.

		Vera Classificazione	
Risultato Classificatore		**P**	**N**
	P	TP	FP
	N	FN	TN

Tabella 7: Confusion Matrix.

I valori presenti nelle celle sono:

- True Positive (TP): numero di elementi correttamente classificati come positivi (P);
- True Negative (TN): numero di elementi correttamente classificati come negativi (N);
- False Positive (FP): numero di elementi classificati come positivi ma che in realtà sono negativi;
- False Negative (FN): numero di elementi classificati come negativi ma che in realtà sono positivi.

Quindi le celle che si trovano sulla diagonale di questa tabella indicano la numerosità degli elementi che sono stati classificati correttamente classe per classe. Una metrica basata sulla matrice di confusione che può essere scelta per determinare le performance del modello è l'accuratezza. Si calcola come il numero degli elementi che sono stati classificati correttamente diviso il numero totale dei classificati:

$$Acc. = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100 \tag{5.6}$$

La migliore accuratezza che è possibile ottenere è pari al 100%.

5.3.1 L'algoritmo combinatorio

La classificazione è stata fatta per gruppi che vanno da 2 a 5 features, selezionate tramite il seguente algoritmo:

1. Si determina il numero di combinazioni k di n feature in p combinazioni con la seguente formula:

$$k = \frac{n!}{p! \cdot (n-p)!} \tag{5.7}$$

   Nel caso in cui il numero di feature è pari a 45 ($n = 45$) e se $p = 2$, allora il numero di combinazioni è 990 ($k = 990$). Se $p = 3$ e con lo stesso numero di features, il numero di combinazioni è 14190.

2. Per ogni possibile combinazione si calcolano tutti gli accoppiamenti, non ripetuti, tra le feature.

3. Per ogni possibile accoppiamento tra le feature, generato al passo 2, si confronta il corrispondente valore nella matrice candidata con un valore soglia scelto pari a 0.5:

   (a) se $candidates_{ij} > 0.5$, allora la combinazione non è considerata valida;

   (b) altrimenti si va al passo 4;

4. Si allena un modello di regressione logistica;

5. Si valutano i beta stimati dal training:

   - se tutti i beta stimati dal training hanno p-value < 0.05, allora il modello è valido;
   - altrimenti, si torna al passo 3.

5.3.2 Pesi stimati

La regressione logistica pesata è stata utilizzata come algoritmo di classificazione per discriminare le voci dei pazienti PD dalle voci dei soggetti HE e le voci dei pazienti PD dalle voci dei pazienti PA. I pesi sono stati definiti a partire dalla stima dell'errore dell'algoritmo di estrazione delle feature.
L'idea è stata quella di far tenere maggiormente in considerazione al classificatore la combinazione di feature che presentano il minore errore relativo medio. Per fare questo sono stati definiti i pesi in modo che ad errori minori

corrispondessero pesi maggiori e viceversa.
I pesi utilizzati sono stati ricavati a partire dagli errori dell'algoritmo di estrazione, i quali sono stati prima divisi per 100 e poi sono state definite le relazioni che legano i pesi agli errori come:

- il reciproco: il valore del peso w è stato preso pari al reciproco dell'errore err come:
$$w = 1/err; \tag{5.8}$$
- complementare dell'errore rispetto ad 1. Per evitare di avere pesi negativi che si hanno nei casi in cui l'errore percentuale è maggiore del 100%, questo è stato normalizzato per il massimo errore della feature err_{max} associato agli elementi:
$$w = 1 - \frac{err}{err_{max}} \tag{5.9}$$
- funzione parabolica: i pesi possono assumere valori che vanno da 0 a 100, ed in particolare al massimo errore corrisponde il peso uguale a 0 mentre al minimo errore corrisponde il peso pari a 100. La relazione che lega gli altri errori con i pesi è una funzione parabolica con concavità verso il basso. Un esempio è riportato in Figura 5.3.

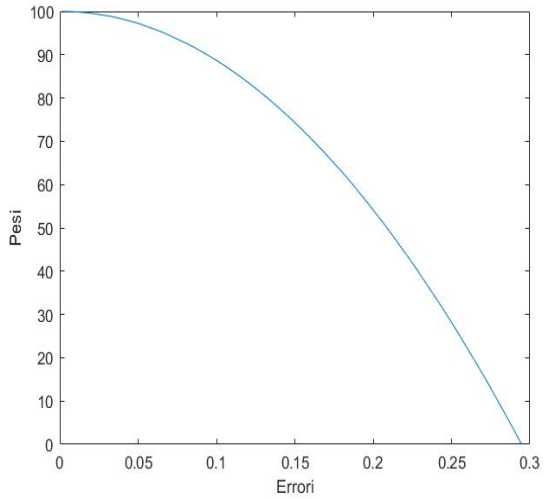

Figura 5.3: Errori e pesi legati da una funzione parabolica.

Il modello di regressione logistica pesata è stato creato utilizzando la funzione Matlab:

mdl = fitglm(PAR,RESP,'Distribution','binomial','link','logit','weights',w);

dove PAR è la matrice contenente le feature, RESP è la classe di appartenenza di ogni elemento da classificare ed assume valore 1 per la classe HE e 0 per la classe PD. Il vettore w è il vettore costituito dai pesi, che possono assumere solo valori positivi.

5.3.3 L'intervallo di fiducia

Sia y il parametro in misura legato alle feature x_i da una generica funzione f come [27]:

$$y = f(x_1, ..., x_N). \tag{5.10}$$

L'incertezza associata alla formula della probabilità (5.5) è stimata con la formula [28]:

$$u(y) = \sqrt{\sum_{i=1}^{N} \left(\frac{\partial y}{\partial x_i}\right)^2 u^2(x_i)} \tag{5.11}$$

dove $\frac{\partial y}{\partial x_i}$ è la derivata parziale rispetto alla variabile i-esima e $u(x_i)$ è l'errore assoluto della variabile i-esima.
Il modello di propagazione dell'incertezza espresso dalla (5.11) è stato applicato per calcolare l'incertezza u_p di $P\{Healthy\}$ (5.5) che è data dalla seguente espressione:

$$u(P) = \frac{e^{\beta_0+\sum_{i=1}^{N}\beta_i f_i}}{(1+e^{\beta_0+\sum_{i=1}^{N}\beta_i f_i})^2} \sqrt{u^2(\beta_0) + \sum_{i=1}^{N} f_i^2 \cdot u^2(\beta_i) + \beta_i^2 \cdot u^2(f_i)} \tag{5.12}$$

dove $u(\beta_i)$ è l'incertezza dei coefficienti e $u(f_i)$ è l'incertezza delle feature.
Quindi, utilizzando la regressione logistica, propagando gli errori stimati ed estendendo l'incertezza con un fattore di copertura vicino a due, il modello ha permesso di ottenere una stima della classificazione con un intervallo di fiducia del 95%.
L'esempio in Figura 5.4 è relativo alla classificazione fatta per discriminare i pazientii PD dai soggetti HE, utilizzando come dataset le feature estratte dai segnali vocali registrati dai soggetti e mediate per le 3 ripetizione della vocale sostenuta /a/ eseguite dal singolo soggetto. I pesi invece sono stati calcolati a partire dalla deviazione tipo delle 3 ripetizioni e la relazione tra errori e pesi è il reciproco.

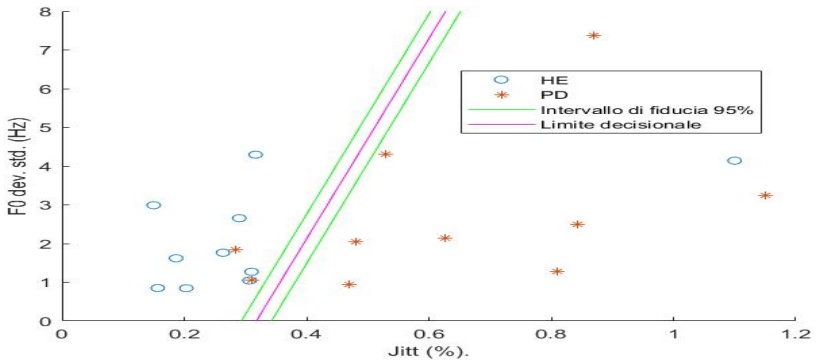

Figura 5.4: Limite decisionale (*decision boundary*) per la media delle 3 vocali reali associate ad ogni soggetto per discriminare i pazienti PD dai soggetti HE con pesi uguali all'inverso della deviazione tipo delle 3 ripetizioni naturali con relativo intervallo di fiducia del 95%.

L'intevallo di fiducia è stato utilizzato per definire altre due metriche per valutare le performance del classificatore: percentuale di classificati e accuratezza per i classificati.

Per elemento classificato si intende quell'elemento la cui forchetta di probabilità data dall'intervallo di fiducia non interseca la soglia di classificazione (soglia: $P\{Healthy\} = 0.5$). Quindi la percentuale di classificati $Class.$ è stata calcolata come:

$$Class. = \frac{NC}{N_{TOT}} \cdot 100 \tag{5.13}$$

dove NC è il numero degli elementi considerati classificati e N_{TOT} è il numero totale degli elementi.

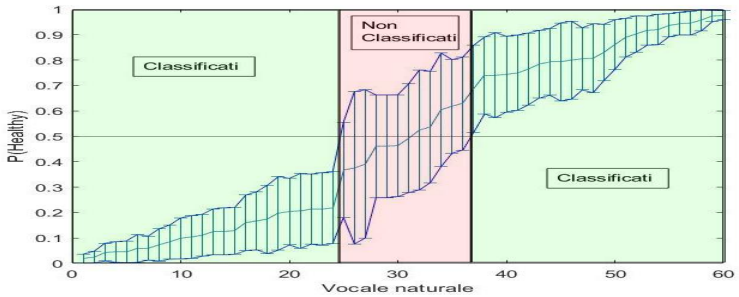

Figura 5.5: Stima della classificazione con relativo intervallo di fiducia del 95%. Gli elementi che si trovano nella regione verde sono considerati classificati mentre quelli che si trovano nella regione rossa sono considerati dubbi.

L'accuratezza dei classificati $Acc.Class.$ si calcola come la somma dei veri positivi e dei veri negativi degli elementi ritenuti classificati N_{TC} diviso il numero totale di elementi N_{TOT}:

$$Acc.Class. = \frac{N_{TC}}{N_{TOT}} \cdot 100 \tag{5.14}$$

Il miglior numero di classificati è uguale al 100% e la miglior accuratezza dei classificati è sempre uguale al 100%.

Capitolo 6

Risultati

6.1 Introduzione

Nella prima parte è stata fatta una classificazione per discriminare i pazienti PD dai soggetti HE utilizzando un modello di regressione logistica e un modello di regressione logistica pesata. Per entrambe i modelli il dataset era costituito dalla media delle feature estratte dalle 3 vocali reali eseguite dal soggetto. I pesi utilizzati per la regressione logistica pesata sono stati definiti come il reciproco della deviazione tipo delle 3 ripetizioni. Poi è stato fatto un confronto tra i due modelli sia in termini di accuratezza, sia in termini di incertezza dei coefficienti che sono dati dal modello creato con la funzione Matlab "fitglm" per mezzo degli errori tipo. Un esempio di risultati ottenuti utilizzando 2 feature sono riportati in Tabella 6.1.

Tabella 6.1: Accuratezza ed incertezza dei coefficienti per la regressione logistica e per la regressione logistica pesata.

Modello	**Feature**	**Acc.(%)**	$\beta 0$	$\beta 1$	$\beta 2$	**SE0**	**SE1**	**SE2**
Regressione logistica	2, 46	85	67.2	-0.26	-0.61	33	0.12	0.31
Regressione logistica pesata	8, 40	95	1.6	-3.7	7.1	0.18	0.07	0.22

I risultati ottenuti hanno mostrato che la regressione logistica pesata fornisce accuratezze maggiori ed incertezze dei coefficienti minori rispetto la re-

gressione logistica. Per questo motivo è scelta la regressione logistica pesata per le classificazioni successive.

Le feature estratte dal segnale reale e dal segnale artificiale, insieme alle stime dell'errore di misura, sono state utilizzate per costruire diversi subset per la classificazione come segue:

- gli elementi da classificare sono la media delle 3 vocali reali associate ad ogni soggetto e gli errori sono la deviazione tipo delle 3 vocali naturali;
- gli elementi da classificare sono la media delle 33 vocali (3 reali + 30 artificiali) associate ad ogni soggetto e gli errori sono la deviazione tipo delle 33 vocali;
- gli elementi da classificare sono le 30 vocali artificiali e gli errori sono i 30 errori relativi alle vocali artificiali;
- gli elementi da classificare sono le 3 vocali reali eseguite dal soggetto e gli errori sono la media dei 10 errori relativi alle vocali artificiali.

Tabella 6.2: I diversi subset usati per la classificazione.

Riferimento Classificazione	**Elementi per un soggetto**	**Errori**
Classificazione 1	$\overline{3\,Vocali_R}$	$\frac{\sigma_R}{\sqrt{3}}$
Classificazione 2	$\overline{33\,Vocali_{R+A}}$	$\frac{\sigma_{R+A}}{\sqrt{33}}$
Classificazione 3	$30\,Vocali_A$	$30\,err_A$
Classificazione 4	$3\,Vocali_R$	$\overline{10\,err_A}$

Tabella 6.3: Pro e contro dei diversi subset usati per la classificazione.

Riferimento Classificazione	**Pro**	**Contro**
Classificazione 1	Si usano le feature reali	L'errore è stimato con poche ripetizioni e non è possibile sapere se è presente un effetto sistematico
Classificazione 2	Massimo oversampling	Riduzione del dataset
Classificazione 3	Oversampling dei campioni	L'errore è stimato con una singola misura
Classificazione 4	Si usano le feature reali	Stima errore fatta su singola vocale, dipende dalle performance del soggetto

Di seguito saranno riportate le prestazioni della classificazione per PD vs. HE e per PD vs. PA.

6.2 Confronto tra voci nella malattia di Parkinson e voci sane

Inizialmente sono state valutate le prestazioni del classificatore per discriminare i pazienti PD dai soggetti HE, dove il segnale artificiale al quale ci si riferisce è quello ottenuto con il metodo di ricampionamento del periodo originale.

L'accuratezza dei classificati, in funzione del numero di feature, per la classificazione 1 (utilizzando tutte le diverse relazioni tra errori e pesi) è riportata in Figura 6.1.

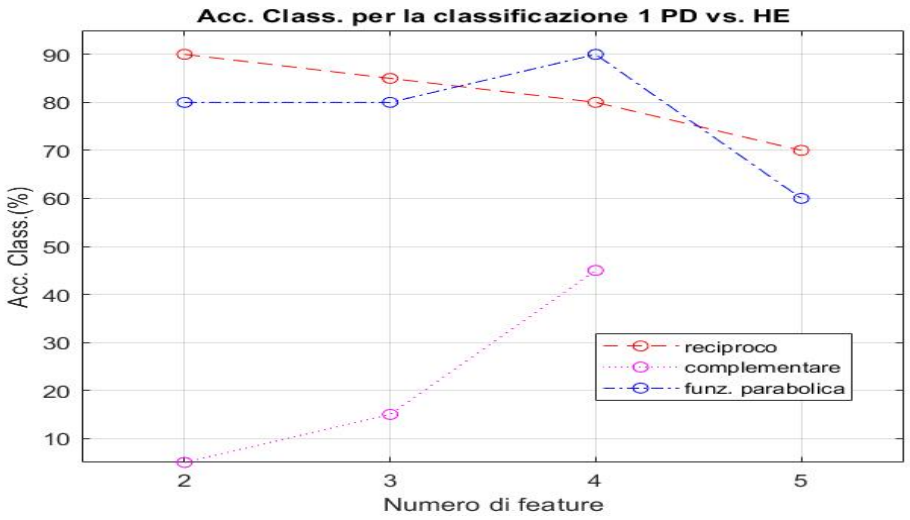

Figura 6.1: Accuratezza dei classificati in funzione del numero di feature per la classificazione 1 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Come si può osservare dal grafico in Figura 6.1 le prestazioni migliori in termini di accuratezza per la classificazione 1 si sono ottenute sia fornendo i pesi legati agli errori per mezzo di una funzione parabolica per un numero di feature pari a 4, sia utilizzando i pesi come reciproco degli errori e per un numero di feature pari a 2. Un confronto tra questi 2 casi è stato fatto valutando anche le altre metriche e i risultati sono riportati in Tabella 6.4. Le prestazioni peggiori sono per il caso in cui i pesi sono i complementari dell'errore rispetto ad 1. Per questo motivo per le classificazioni successive non è stata più presa in considerazione questa relazione.

Tabella 6.4: Prestazioni della classificazione 1 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	1, 8	90	100	90
Funz. parabolica	4, 12, 27,42	95	95	90

Prima di allenare un modello per la classificazione 2 e prima che fossero mediate, sono state confrontate le feature estratte dal segnale vocale reale e dal segnale vocale artificiale per i pazienti PD e i soggetti HE. Un esempio per 2 feature è mostrato in Figura 6.2. Si può osservare come le feature estratte dal segnale vocale artificiale sono plausibili rispetto quelle estratte dal segnale reale.

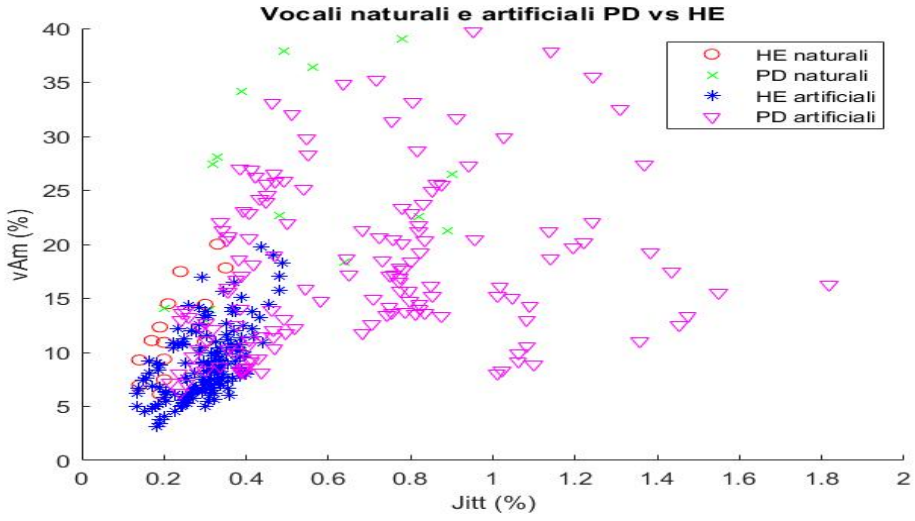

Figura 6.2: Confronto tra 2 feature estratte dal segnale naturale e dal segnale artificiale per i pazienti PD e i soggetti HE.

Successivamente è stato allenato un modello per la classificazione 2 e le prestazioni sono riportate in Figura 6.3.

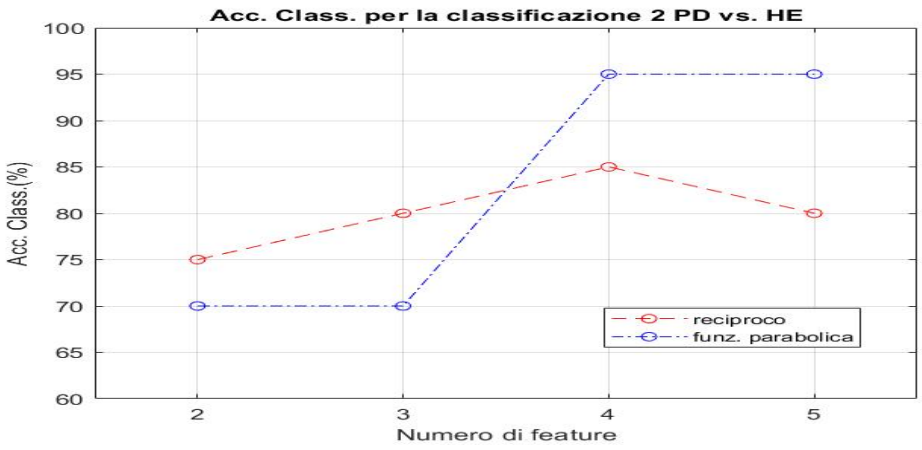

Figura 6.3: Accuratezza dei classificati in funzione del numero di feature per la classificazione 2 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Le prestazioni migliori sono state riscontrate per i pesi e gli errori legati da una funzione parabolica e per 4 e 5 feature. I valori dell'accuratezza dei classificati insieme alle altre metriche sono riportati in Tabella 6.5.

Tabella 6.5: Prestazioni della classificazione 2 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Funz. parabolica	4, 12, 23, 31	95	100	95
Funz. parabolica	1, 5, 12, 23, 32	95	100	95
Reciproco	4, 12, 23, 31	90	90	85

In Figura 6.4 sono riportati i risultati per la classificazione 3.

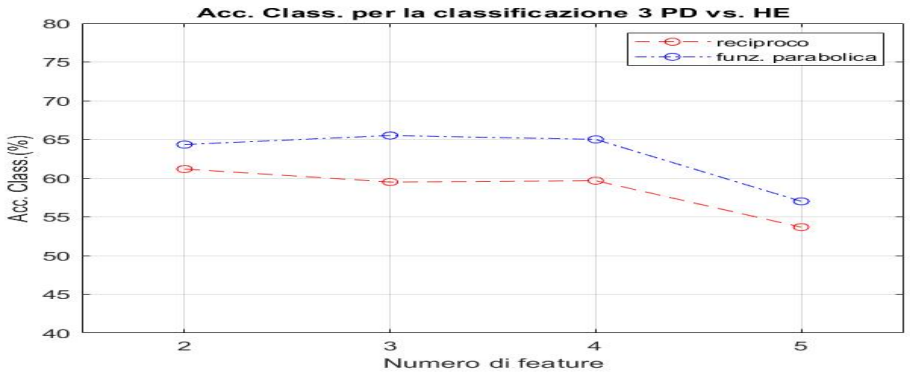

Figura 6.4: Accuratezza dei classificati in funzione del numero di feature per la classificazione 3 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

I risultati ottenuti mostrano basse accuratezze dei classificati ripetto le due classificazioni precedenti.

L'accuratezza dei classificati, in funzione del numero di feature, per la classificazione 4 è riportata in Figura 6.5.

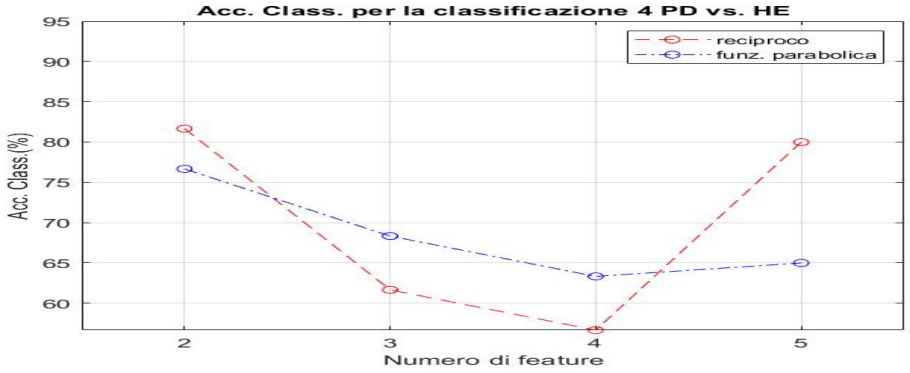

Figura 6.5: Accuratezza dei classificati in funzione del numero di feature per la classificazione 4 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Fornendo i pesi come il reciproco degli errori si sono ottenute le prestazioni migliori in termini di accuratezza dei classificati per un numero di feature

pari a 2 e a 5. Per questo caso sono state riportate anche le altre metriche in Tabella 6.6.

Tabella 6.6: Prestazioni della classificazione 4 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	6, 9	82	95	82
Reciproco	4, 12, 27, 40, 46	95	85	80
Funz. parabolica	6, 32	83	85	77

Successivamente sono state fatte le classificazioni utilizzando come segnali artificiali quelli ottenuti con la trasformazione del cepstrum. I risultati per le classificazioni 2, 3, 4 per discriminare i pazienti PD dai soggetti HE saranno riportate di seguito.

Come si può osservare dalla Figura 6.6, le prestazioni migliori per la classificazione 2 sono state ottenute fornendo i pesi come il reciproco dell'errore e per 2 feature. In Tabella 6.7 sono mostrate le metriche per le combinazioni che hanno dato le prestazioni migliori.

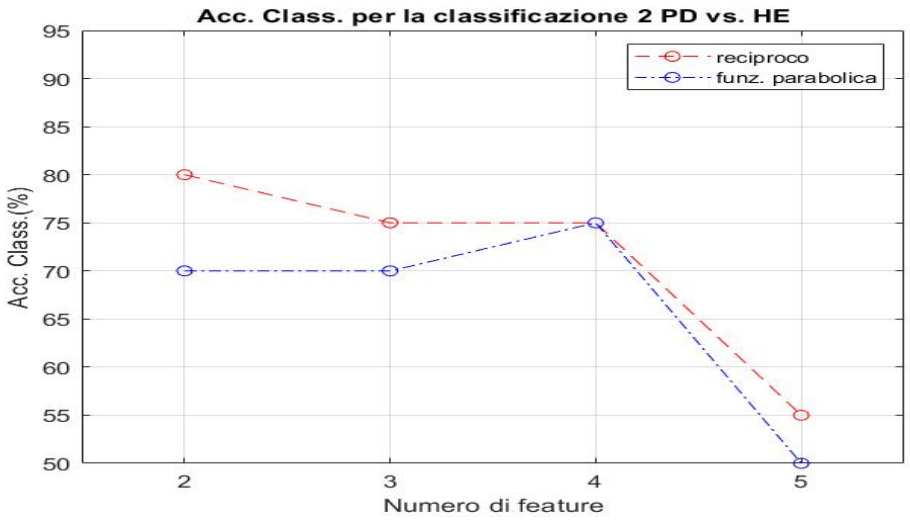

Figura 6.6: Accuratezza dei classificati in funzione del numero di feature per la classificazione 2 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Tabella 6.7: Prestazioni della classificazione 2 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	1, 32	80	100	80
Reciproco	3, 31, 40, 43	90	75	75
Funz. parabolica	4, 16, 31, 37	95	80	85

L'accuratezza dei classificati in funzione del numero di feature per la classificazione 3 è mostrata in Figura 6.7.

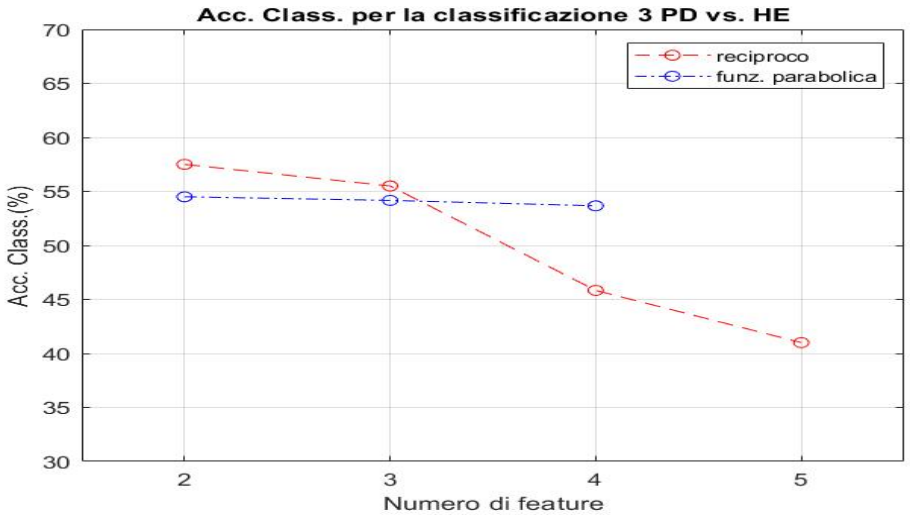

Figura 6.7: Accuratezza dei classificati in funzione del numero di feature per la classificazione 3 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

In questo le prestazioni della classificazione sono diminuite rispetto alla classificazione 3 per tutte le metriche. Nella Tabella 6.8 sono state riportate le combinazioni di feature che hanno dato le accuratezze maggiori.

Tabella 6.8: Prestazioni della classificazione 3 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	1, 12, 15, 44	76	54	46
Funz. parabolica	2, 12, 23, 42	71	66	54

Le prestazioni migliori della classificazione 4 in termini di accuratezza dei classificati sono mostrate in Figura 6.8. Nella Tabella 6.9 sono state riportate le combinazioni di feature che hanno dato le accuratezze maggiori.

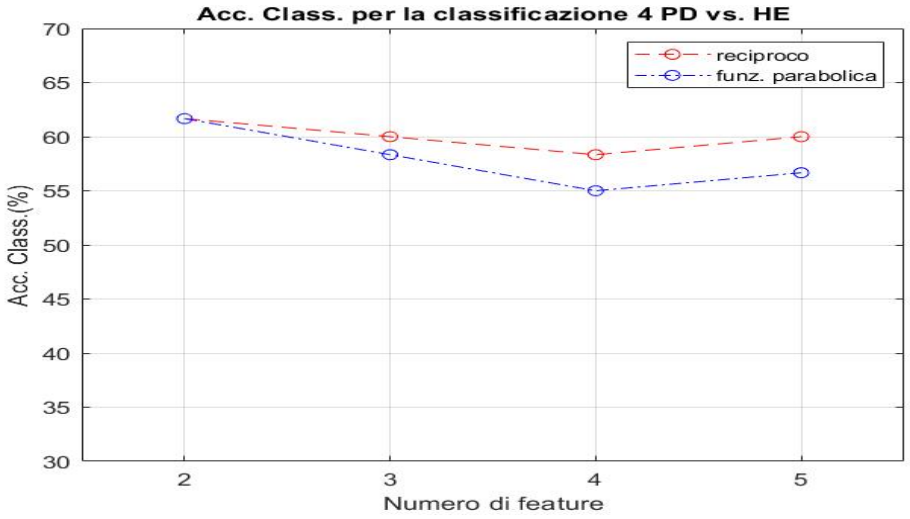

Figura 6.8: Accuratezza dei classificati in funzione del numero di feature per la classificazione 4 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Tabella 6.9: Prestazioni della classificazione 4 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	4, 5, 12, 31, 42	92	60	60
Funz. parabolica	4, 5, 12, 31, 42	90	56	56

Si osservi come l'accuratezza ha raggiunto valori fino al 92% con 5 feature e fornendo i pesi come reciproco degli errori.

I risultati ottenuti mostrano come per il caso in cui i segnali artificiali sono generati con il metodo di trasformazione del cepstrum si ha una diminuzione del numero di classificati e dell'accuratezza dei classificati rispetto il caso in cui i segnali artificiali sono generati con il metodo di ricampionamento del periodo originale. Al contrario l'accuratezza non è diminuita per la classificazione 2 dove entrambi i casi hanno raggiunto valori pari al 95% (Tabella 6.5 e Tabella 6.7) e per la classificazione 4 è passata da un massimo di 95% per il caso di segnali ottenuti con il ricampionamento del periodo originale (Tabella 6.6) ad un massimo di 92% per il caso di segnali ottenuti con la trasformazione del cepstrum (Tabella 6.9).
Sulla base di questi risultati potrebbe essere ragionevole dedurre che l'incertezza associata alla probabilità sia maggiore per il caso dei segnali ottenuti con la trasformazione del cepstrum rispetto ai segnali ottenuti con il ricampionamento del periodo originale. Quindi per valutare i contributi dell'incertezza dei coefficienti e dell'incertezza delle feature è stata presa ad esempio la classificazione 4 con una combinazione di feature pari a 2 per entrambi i casi. In Figura 6.9 e in Figura 6.10 sono riportate la combinazione delle 2 feature con cui sono state ottenute le prestazioni migliori in termini di accuratezza dei classificati per la classificazione 4, rispettivamente, per il caso del ricampionamento del periodo originale e di trasformazione del cepstrum.

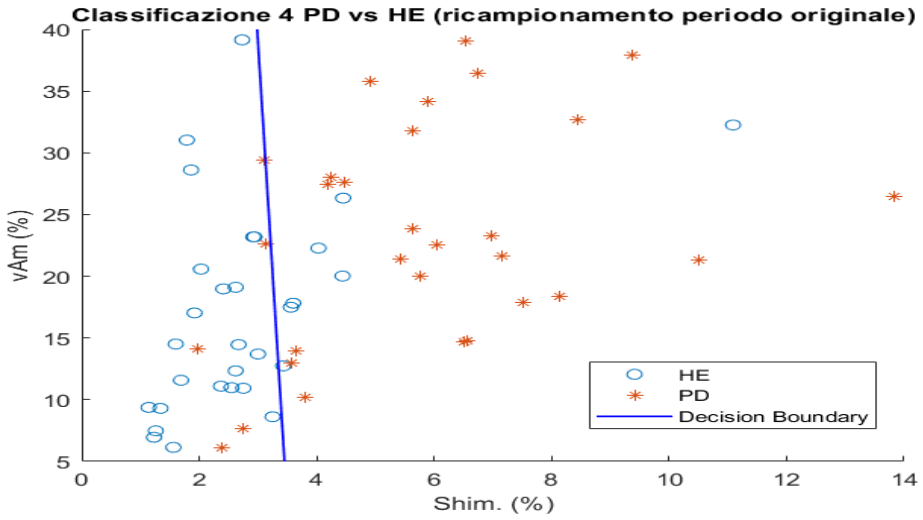

Figura 6.9: Limite decisionale (*decision boundary*) per PD vs. HE per la classificazione 4 (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

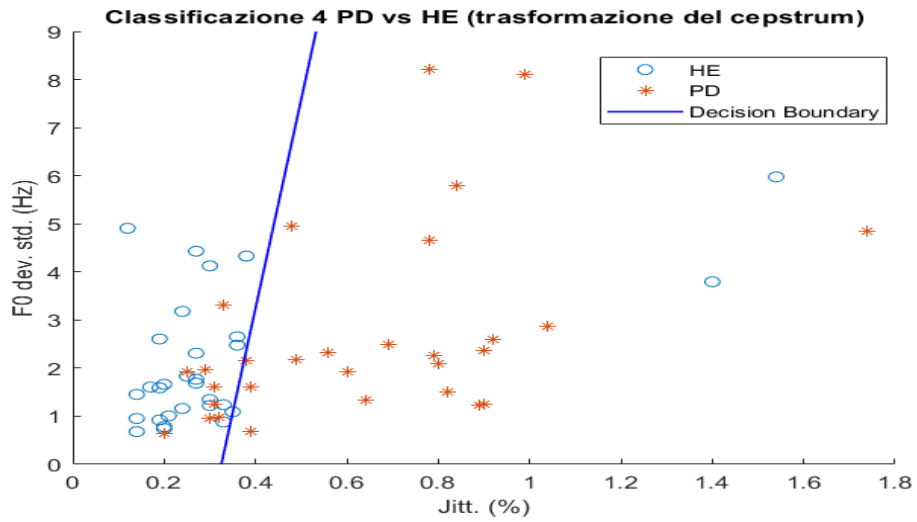

Figura 6.10: Limite decisionale (*decision boundary*) per PD vs. HE per la classificazione 4 (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

In Tabella 6.10 sono state riportate le prestazioni migliori in termini di accuratezza dei classificati, insieme alle altre due metriche, ottenute nella classificazione 4 con una combinazione di 2 feature e fornendo i pesi come reciproco degli errori.

Tabella 6.10: Prestazioni della classificazione 4 PD vs. HE (i segnali artificiali utilizzati per costruire il dataset sono stati generati sia con il metodo di ricampionamento del periodo originale, sia con il metodo di trasformazione del cepstrum).

Metodo	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Ric. periodo originale	6, 9	82	95	82
Trasf. cepstrum	1, 22	83	61	58

I contributi dell'incertezza dei coefficienti sono in Tabella 6.11 e dell'incertezza delle feature sono in Tabella 6.12.

Tabella 6.11: Incertezza dei coefficienti per la classificazione 4 PD vs. HE.

Modello	**Feat.**	$\beta 0$	$\beta 1$	$\beta 2$	**SE0**	**SE1**	**SE2**	**Err. medio(%)**
Ric. periodo originale	6, 9	102.7	-29.2	-0.4	0.43	0.12	0.004	0.6
Trasf. cepstrum	1, 22	2.6	-8.1	0.19	0.23	0.65	0.09	19

Tabella 6.12: Incertezza delle feature per la classificazione 4 PD vs. HE.

Modello	**Feat.**	**Err. medio Feat1 (%)**	**Err. medio Feat2 (%)**
Ric. periodo originale	6, 9	6.2	2.1
Trasf. cepstrum	1, 22	19.4	16.6

Per la classificazione fatta utilizzando i segnali artificiali ottenuti con il metodo di trasformazione del cepstrum aumentano i contributi dell'incertezza dei coefficienti e delle feature rispetto al caso in cui la classificazione è fatta utilizzando i segnali artificiali ottenuti con il metodo di ricampionamento del segnale originale.

6.3 Confronto tra voci nella malattia di Parkinson e voci patologiche

Di seguito verranno riportati i risultati per le tutte le classificazione fatte per discriminare i pazienti PD dai pazienti PA.

Per la classificazione 1, l'accuratezza dei classificati in funzione dei numero del numero di feature è riportato in Figura 6.11.

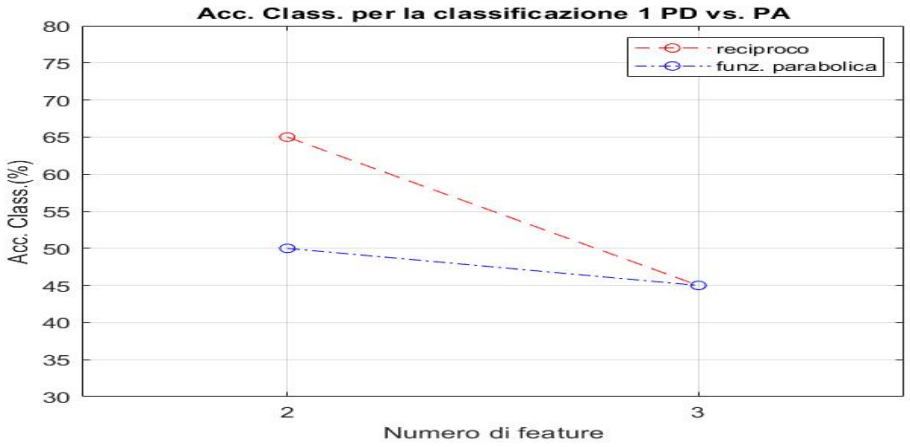

Figura 6.11: Accuratezza dei classificati in funzione del numero di feature per la classificazione 1 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Le combinazione di feature che hanno portato alle prestazioni migliori sono riportate in Tabella 6.13.

Tabella 6.13: Prestazioni della classificazione 1 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	22, 32	75	80	65
Funz. parabolica	1, 37	70	55	50

Dato che all'aumentare del numero di feature le prestazioni della classificazione diminuiscono, è stato scelto di utilizzare solo un numero di feature pari a 2 e a 3.

Per la classificazione 2, l'accuratezza dei classificati in funzione dei numero del numero di feature è riportato in Figura 6.12.

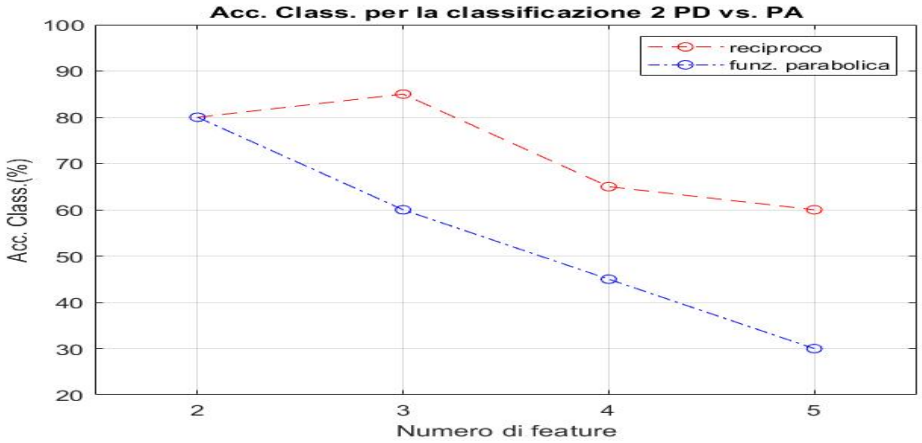

Figura 6.12: Accuratezza dei classificati in funzione del numero di feature per la classificazione 2 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Le combinazione di feature che hanno portato alle prestazioni migliori sono riportate in Tabella 6.14.

Tabella 6.14: Prestazioni della classificazione 2 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	12, 44	85	95	80
Reciproco	12, 42, 44	85	100	85
Funz. parabolica	12, 44	80	90	80

Per la classificazione 3, l'accuratezza dei classificati in funzione dei numero del numero di feature è riportato in Figura 6.13.

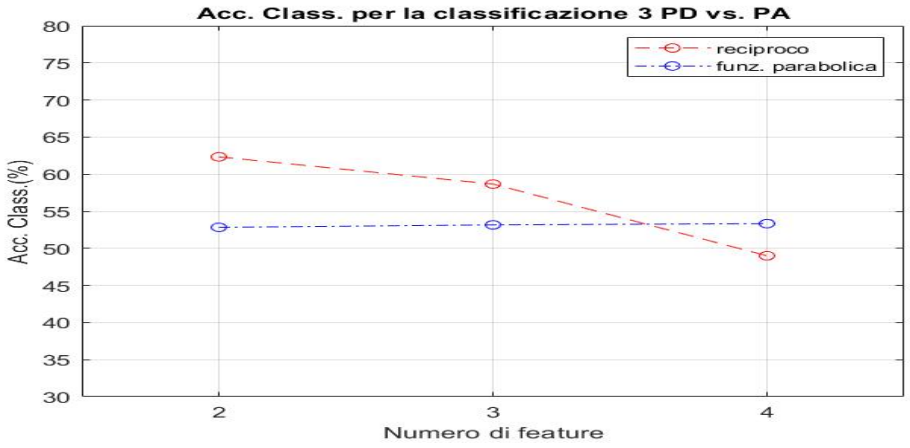

Figura 6.13: Accuratezza dei classificati in funzione del numero di feature per la classificazione 3 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Le combinazione di feature che hanno portato alle prestazioni migliori sono riportate in Tabella 6.15.

Tabella 6.15: Prestazioni della classificazione 3 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	8, 38	67	93	62
Funz. parabolica	2, 5, 8, 38	64	79	53

Per la classificazione 4, l'accuratezza dei classificati in funzione dei numero del numero di feature è riportato in Figura 6.14.

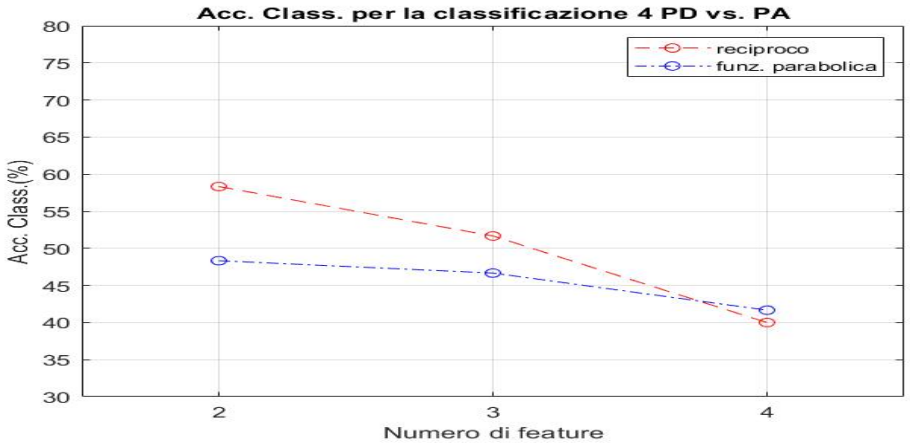

Figura 6.14: Accuratezza dei classificati in funzione del numero di feature per la classificazione 4 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Le combinazione di feature che hanno portato alle prestazioni migliori sono riportate in Tabella 6.16.

Tabella 6.16: Prestazioni della classificazione 4 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di ricampionamento del periodo originale).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	9, 38	67	80	58
Funz. parabolica	32, 38	75	53	48

Successivamente sono state fatte le classificazioni utilizzando come segnali artificiali quelli ottenuti con la trasformazione del cepstrum. I risultati per le classificazioni 2, 3, 4 per discriminare i soggetti PD dai soggetti PA saranno riportate di seguito.

La Figura 6.15 mostra l'accuratezza dei classificati in funzione del numero di feature per la classificazione 2.

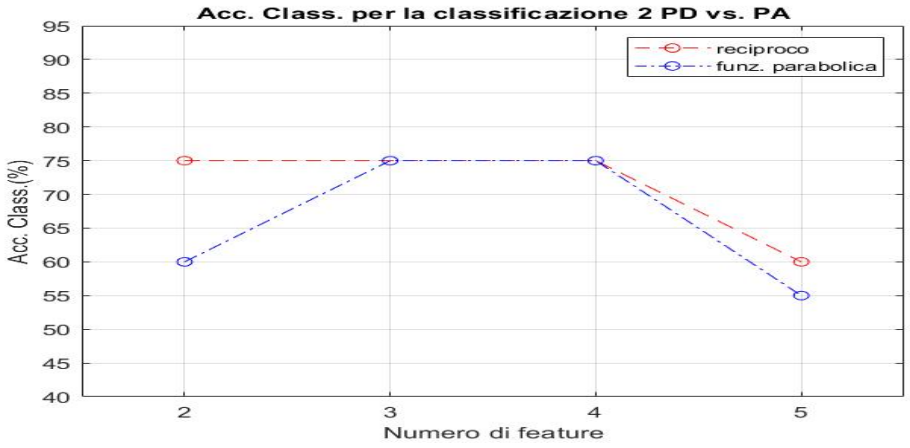

Figura 6.15: Accuratezza dei classificati in funzione del numero di feature per la classificazione 2 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Come si può osservare dalla Figura 6.15, l'accuratezza migliore è stata ottenuta sia utilizzando i pesi come reciproco dell'errore per 2, 3, 4 feature, sia nel caso in cui i pesi e gli errori sono legati da una funzione parabolica per 3 e 4 feature. I risultati per questi casi sono riportati in Tabella 6.17.

Tabella 6.17: Prestazioni della classificazione 2 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	31, 44	75	90	75
Reciproco	31, 38, 40	90	80	75
Reciproco	9, 31, 37, 40	95	75	75
Funz. parabolica	8, 37, 40	95	80	75
Funz. parabolica	8, 15, 38, 40	95	80	75

L'accuratezza dei classificati in funzione del numero di feature è mostrato in Figura 6.16 per la classificazione 3.

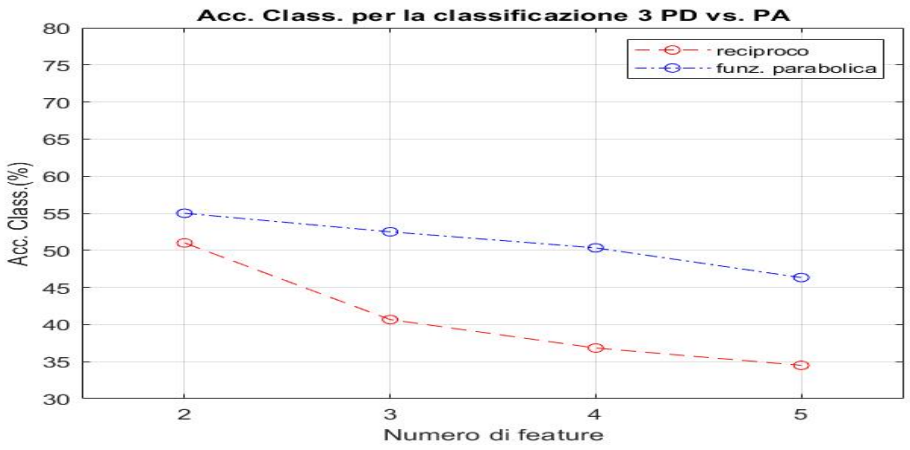

Figura 6.16: Accuratezza dei classificati in funzione del numero di feature per la classificazione 3 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

In questo caso si ha una diminuzione delle prestazioni rispetto la classificazione 2. In Tabella 6.18 sono riportati i risultati relativi alle migliori accuratezze tra le combinazioni riportate in Figura 6.16.

Tabella 6.18: Prestazioni della classificazione 2 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	9, 11, 31, 38, 40	75	36	34
Funz. parabolica	38, 40	72	71	55

La Figura 6.17 mostra l'accuratezza dei classificati in funzione del numero di feature per la classificazione 4.

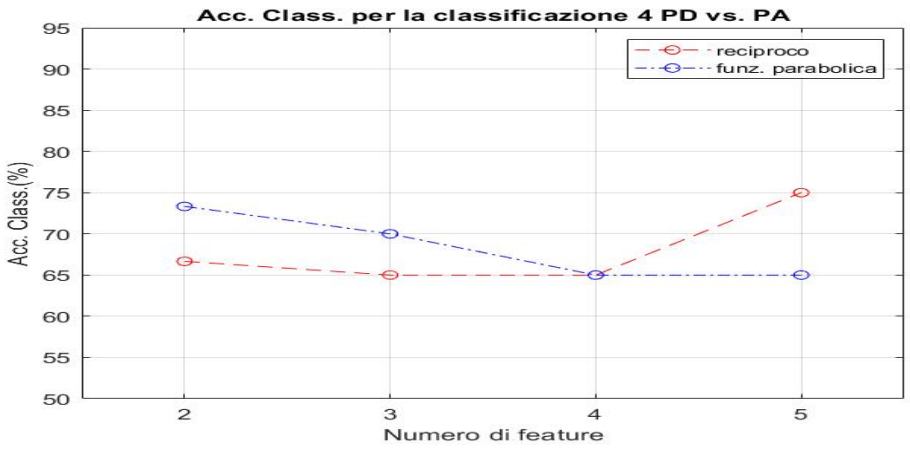

Figura 6.17: Accuratezza dei classificati in funzione del numero di feature per la classificazione 4 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

I risultati in riferimento alle migliori accuratezze dei classificati sono riportati in Tabella 6.19.

Tabella 6.19: Prestazioni della classificazione 4 PD vs. PA (i segnali artificiali utilizzati per costruire il dataset sono stati generati con il metodo di trasformazione del cepstrum).

Pesi-Errori	**Feature**	**Acc.(%)**	**Class.(%)**	**Class.Acc.(%)**
Reciproco	4, 12, 27, 41, 48	93	80	75
Funz. parabolica	6, 32	80	86	73

Capitolo 7

Conclusioni

In questo lavoro di tesi è stato stato proposto un metodo di classificazione pesato basato sulla stima dell'errore di estrazione delle feature per discriminare i soggetti sani (HE) dai pazienti affetti dalla malattia di Parkinson (PD) e i pazienti patologici non parkinsoniani (PA) dai soggetti affetti dalla malattia di Parkinson (PD).
La numerosità del dataset composto da 3 ripetizioni della vocale sostenuta /a/ registrati con un microfono in aria per i 10 soggetti che appartengono alla classe HE, PA e PD, è stata aumentata generando 30 nuove vocali artificiali per ogni soggetto con il metodo di ricampionamento del periodo originale, ritenuto il migliore tra i metodi proposti dato che presenta la media degli errori minore. A partire da questi segnali artificiali sono state generate nuove vocali artificiali al fine di ridurre l'errore relativo ai CPPS.
Le feature estratte dai segnali reali e artificiali, insieme alla stime degli errori, hanno permesso di allenare un algoritmo di regressione logistica pesata. Il metodo di classificazione ha fornito la probabilità di appartenere alla classe dei sani con un intervallo di fiducia del 95%. Per la classificazione sono stati provati 4 diversi subset e i pesi sono stati forniti in modo da far tenere maggiormente in considerazione al classificatore la combinazione di feature che presentano il minor errore relativo medio: ad errori minori corrispondono pesi maggiori e viceversa. Per questo motivo i pesi sono stati definiti come reciproco, complementare rispetto ad 1 e funzione parabolica dell'errore. Date le scarse prestazioni iniziali raggiunte dalla classificazione in cui i pesi sono stati definiti come complementare rispetto ad 1 dell'errore, questa relazione non à stata presa in considerazione per le successive classificazioni.

Nel discriminare i pazienti PD dai soggetti HE, il miglior risultato è stato ottenuto per il caso in cui gli elementi da classificare sono la media delle 33 vocali (3 reali + 30 artificiali) associate ad ogni soggetto e gli errori sono la deviazione tipo delle 33 vocali, dove per vocale artificiale si intende quella

generata dal metodo di ricampionamento del periodo originale. In questo caso, fornendo i pesi in funzione degli errori tramite una parabola, l'accuratezza dei classificati raggiunta è pari al 95%, il numero di classificati è pari al 100% e l'accuratezza al 95% per una combinazione di feature sia pari a 4 (PPQ, HNR media, F_0 range, RMS dev. std.), sia pari a 5 (Jitter, vF_0, HNR media, F_0 range, RMS range). Le peggiori accuratezze sono state trovate per il caso in cui gli elementi da classificare sono le 30 vocali artificiali e gli errori sono i 30 errori relativi alle vocali artificiali, mostrando accuratezze dei classificati che raggiungono al massimo il 66% (errori e pesi legati da una funzione parabolica).
Per le classificazioni in cui i segnali artificiali considerati erano quelli ottenuti con il metodo di trasformazione del cepstrum, l'accuratezza dei classificati maggiore (85%) è stata raggiunta sempre per il caso in gli elementi da classificare sono la media delle 33 vocali (3 reali + 30 artificiali) associate ad ogni soggetto e gli errori sono la deviazione tipo delle 33 vocali, i pesi e gli errori legati da una funzione parabolica e per una combinazione di 4 feature (PPQ,F_0 skewness, RMS dev. std. e CPPS media). Le prestazioni della classificazione utilizzando i segnali artificiali ottenuti con il metodo di trasformazione del cepstrum sono diminuite in termini sia di classificati, sia di accuratezza dei classificati rispetto il caso in cui segnali artificiali sono generati con il metodo di ricampionamento del periodo originale. Invece, l'accuratezza migliore per entrambi i metodi di generazione dei segnali vocali è pari al 95%.

Nel discriminare i pazienti PD dai pazienti PA e considerando come segnali artificiali quelli ottenuti con il metodo di ricampionamento del periodo originale, la migliore accuratezza è stata raggiunta per il caso in cui sono state classificate le medie delle 33 vocali (3 reali + 30 artificiali) associate ad ogni soggetto e gli errori sono la deviazione tipo delle 33 vocali. Il valore raggiunto è pari all'85% per una combinazione di feature pari a 3 (HNR moda, CPPS 5° percentile e CPPS skewness) e defininendo l'errore come reciproco dell'errore. L'accuratezza e il numero di classificati hanno raggiunto, rispettivamente, l'85% e il 100%.
Per le classificazioni in cui i segnali artificiali considerati erano quelli ottenuti con il metodo di trasformazione del cepstrum, l'accuratezza dei classificati maggiore è stata ottenuta fornendo il subset sia come medie delle 33 vocali (3 reali + 30 artificiali) associate ad ogni soggetto e gli errori sono la deviazione tipo delle 33 vocali, sia come le 3 feature reali eseguite dal soggetto e gli errori sono la media dei 10 errori relativi alle vocali artificiali. L'accuratezza dei classificati raggiunta è pari al 75%.
Per discriminare i soggetti HE dai pazienti PD, per il caso in cui i segnali artificiali sono stati generati con il metodo di trasformazione del cepstrum

è stata osservata una diminuzione di prestazioni in riferimento all'accuratezza dei classificati, rispetto al caso in cui i segnali artificiali sono stati generati con il metodo di ricampionamento del periodo originale. Questo è dovuto all'aumento sia dell'incertezza dei coefficienti, sia dell'incertezza delle feature. Per discriminare i soggetti sani HE dai pazienti PD, sembra che l'algoritmo utilizzi frequentemente le feature legate alla valutazione di stabilità in frequenza ed in ampiezza ma è stato osservato come la trasformazione del cepstrum aumenti l'errore delle misure temporali. La trasformazione del cepstrum varia la durata dei periodi perchè il CPPS è misurato con finestre lunghe 1024 campioni (46 ms) e sovrapposte di 44 campioni (2 ms) mentre le altre grandezze sono misurate in maniera sincrona agli pseudo-periodi. Quindi l'incertezza della feature aumenta e di conseguenza si ha un aumento dell'incertezza di appartenere alla classe dei sani, provocando una diminuzione delle prestazioni del classificatore.
Per discriminare i pazienti PA dai pazienti PD, ed in riferimento alla classificazione 4, per il caso in cui i segnali artificiali sono stati generati con il metodo di trasformazione del cepstrum è stata osservata un aumento di prestazioni in riferimento all'accuratezza dei classificati rispetto al caso in cui i segnali artificiali sono stati generati con il metodo di ricampionamento del periodo originale. Il classificatore per discriminare la classe PA dalla classe PD sembra che utilizzi maggiormente i parametri CPPS ed avendo diminuito l'errore delle misure cepstrali con il metodo di trasformazione del cepstrum, l'incertezza di queste feature diminuisce. Come risultato c'è stato un aumento delle prestazioni.

Bibliografia

[1] I. McLoughlin, *Applied Speech and Audio Processing: With Matlab Examples*, Cambridge University Press, 2009.

[2] Stan Z. Li, Anil K. Jain, in *Encyclopedia of Biometrics*, Springer, 2009, pp. 1291.

[3] T. F. Quatieri, *Discrete-Time Speech Signal Processing, Principles and Practices*, Prentice Hall PTR, 2001.

[4] A. Castellana, A. Carullo, S. Corbellini, A. Astolfi, *Disciminating Pathological Voice From Healthy Voice Using Cepstral Peak Prominence Smoothed Distribution in Sustained Vowel*, IEEE Transactions on Instrumentation and Measurement, vol.67, pp. 646-654, 2018.

[5] Rangaraj M. Rangayyan, *Biomedical Signal Analysis: A Case-Study Approach*, Wiley-IEEE Press, 2001.

[6] I. Cantürk, F. Karabiber, *A Machine Learning System for the Diagnosis of Parkinson's Disease from Speech Signals and Its Application to Multiple Speech Signal Types*, Arab J Sci Eng (2016) 41:5049-5059.

[7] J. Jankovic, *Parkinson's disease: clinical features and diagnosis*, J. Neurol. Neurosurg. Psychiatry, pp. 368-376, 2008.

[8] J. Gamboa et al., *Acoustic voice analysis in patients with Parkinson's disease treated with dopaminergic drugs*, Journal of Voice Vol. 1, Issue 3, pp. 314-320, September 1997.

[9] I. Daly, Z. Hajaiej, A. Gharsallah, *Detection of Parkinson's Disease through Acoustic Parameters*, Int J of Advanced Computer Science and Applications, vol. 9, no. 7, pp. 311-315, 2018.

[10] Y. Zhang, J. Jiang, *Acoustic Analyses of Sustained and Running Voices From Patients With Laryngeal Pathologies*, Journal of Voice, 2008.

[11] P. Boersma, *Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound*, IFA 17 proceedings, pp. 97-110, 1993.

[12] B. P. Bogert, M. J. R. Healy, and J. W. Tukey, *The Quefrency Alanysis of Time Series for Echos: Cepstrum, Pseudo Autocovariance, Cross Cepstrum and Saphe Cracking*, in Proceedings of Symposium on Time Series Analysis by Rosenblatt, M., (Ed.), Wiley, N.Y. 1963, pp. 209-243.

[13] D. G. Childers, D. P. Skinner, R. C. Kemerait, *The Cepstrum: A Guide to Processing*, proceeding of the IEEE, Vol. 65, No. 10, October, 1997.

[14] J. Hillenbrand, A. Houde *Acoustic correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech*, J. Speech Hearing Res., vol. 39, no. 2, pp. 311-321, 1996.

[15] J. Hillenbrand, R. A. Cleverand and R. L. Erickson, *Acoustic correlates of breathy vocal quality*, J. Speech Hearing Res., vol. 37, no. 4, pp. 769-778, 1994.

[16] M. Asgari, I. Shafran, *Extracting cues from speech for predicting severity of Parkinson's disease*, 2010 IEEE International Workshop on Machine Learning for Signal Processing.

[17] Yu. A. Shreider, *The Monte Carlo Method: The Method of Statistical Trials*, Pergamon Press, 1996.

[18] L. Martino, *A review of multiple try MCMC algorithms for signal processing*, Digital Signal Precessing, pp.134-152, 2018.

[19] L. R. Rabiner, M. Cheng, A. Rosenberg, C. McGonegal, *A comparative performance study of several pitch detection algorithms*, IEEE Trans. Acoustics, Speech Signal Proc., 24(5): 399-418, 1976.

[20] A. Atzori, A. Carullo, A. Vallan, V. Cennamo, A. Astolfi, *Parkinson disease voice features for rehabilitation therapy and screening purposes*, IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, June 26-28, 2019.

[21] A. V. Oppenheim, R. W. Schafer, *From Frequency to Quefrency: A History of the Cepstrum*, IEEE Signal Processing Megazine 21(5):95-106, October, 2004.

[22] E. C. Blessie, E. Karthikeyan, *A feature Selection Algorithm Using Correlation Based Method*, Journal of Algorithms & Computational Technology, 2012.

[23] Lei Yu, Huan Liu, *Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution*, Conference: Machine Learning, Proceedings of the Twentieth International Conference (ICML 2003), Washington, DC, USA, August 21-24, 2003.

[24] A. Ly, M. Marsman, E. Wagenmakers, *Analytic posteriors for Pearson's correlation coefficient*, Statistica Neerlandica, 2015.

[25] A. C. Leon, *Comprehensive Clinical Psychology*, pp. 243-285, 1998.

[26] N. Salari, S. Shohaimi, F. Najafi, M. Nallappan, I. Karishnarajah, *A Novel Hybrid Classification Model of Genetic Algorithms, Modified k-Nearest Neighbor and Developed Backpropagation Neural Network*, PLOS ONE, 2014.

[27] A. Carullo, U. Pisani, A. Vallan, *Fondamenti di misure e strumentazione elettronica*, CULT, 2006.

[28] S. Sediva, M. Havlikova, *Comparison of GUM and Monte Carlo method for evaluation measurement uncertainty of indirect measurements*, Carpathian Control Conference (ICCC), 2013.