Politecnico di Torino

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Biomedica

Tesi di Laurea

Classificazione di pazienti laringectomizzati per fini riabilitativi attraverso l'estrazione e la selezione di parametri vocali

Relatore:

Prof. Alessio Carullo

Correlatore:

Ph.D.C. Alessio Atzori

Candidato:

Lorenzo Midolo

Anno Accademico 2020/2021

Ai miei genitori, a mia sorella e a mio fratello.

Sommario

I trattamenti clinici per le neoplasie che colpiscono la laringe risultano molto spesso debilitanti per i soggetti interessati. Essi si ritrovano ad affrontare elevate difficoltà nell'espletare funzioni primarie, come l'alimentazione, la fonazione e lo svolgimento delle normali interazioni sociali, e solo in seguito ad un'accurata riabilitazione sono in grado di recuperare buona parte di queste. L'analisi dei tracciati vocali è stata introdotta storicamente in tale ambito clinico per tentare di apportarne generali migliorie. Tra le varie è possibile citare la validazione quantitativa dei diversi trattamenti chirurgici adottati, la valutazione quantitativa dei progressi raggiunti nella riabilitazione o i tentativi di costruzione di modelli in grado di diagnosticare la presenza di disturbi più gravi, come la malattia di Parkinson, morbo di Alzheimer e altri. In letteratura sono presenti molti studi sulla valutazione dei diversi trattamenti clinici e sui relativi esiti, mentre solo negli ultimi anni sono stati intrapresi i primi passi per la creazione di un metodo di diagnostica. Per quanto riguarda gli aspetti dell'attività riabilitativa, invece, si è sempre fatto molto affidamento sulla figura del logopedista, sia per l'esecuzione che per la valutazione, sottolineando un vuoto circa un qualsiasi contributo ingegneristico a tale ambito.

L'obiettivo proposto in questo lavoro di tesi è quello di individuare, avvalendosi del software Matlab®, una procedura caratteristica in grado di riconoscere le differenti severità dei pazienti laringectomizzati. Il tutto è stato svolto nell'ottica di porsi come un tassello di un progetto più ampio sviluppato all'interno dei laboratori del DET, presso il Politecnico di Torino. Esso si prefigge la realizzazione di un dispositivo riabilitativo che permetta al paziente di effettuare l'attività logopedica in qualsiasi luogo ed in totale autonomia, comunicando gli esiti al logopedista in tempo reale, così che abbia costanti aggiornamenti sui progressi del soggetto. Ciò permetterebbe ad ogni soggetto di evitare qualsiasi difficoltà connessa al raggiungimento della sede dello specialista e, al contempo, di fornire dei risultati quantitativi della procedura effettuata. Per lo scopo della tesi in questione sono stati selezionati 32 brani di eloqui di soggetti sottoposti a due diversi tipi di operazione chirurgica. In seguito alla consueta fase di pre-processing e ad un primo tentativo di elaborazione non andato a buon fine, è stata validata l'attività di estrazione dei parametri dalle sole porzioni di tracciato corrispondenti all'emissione di suoni vocalizzati, dette anche porzioni

armoniche. La separazione è stata condotta mediante l'utilizzo del descrittore spettrale della kurtosis, la cui efficacia è stata precedentemente valutata su soggetti sani per poi essere adoperata sui pazienti disfonici in esame. Una volta ottenute le relative finestre armoniche, da ognuna di esse sono stati estratti alcuni parametri appartenenti a domini differenti (alcuni descrittori spettrali, alcuni parametri spettrali e alcuni parametri cepstrali), ricavandone così l'andamento temporale e, successivamente, anche le relative statistiche descrittive (media, mediana, moda, ecc). In seguito, attraverso l'uso della scala di valutazione INFVo (*Intellegibility, Noise, Fluency and Voice*), i soggetti sono stati suddivisi in due classi a seconda del grado di intellegibilità da loro dimostrato. Con l'uso del test statistico di Kolmogorov-Smirnov sono state effettuate delle operazioni di confronto tra tutti gli andamenti dei parametri citati, risalenti ad entrambe le classi generate. Ciò è stato intrapreso allo scopo di evidenziare un qualche tipo di potere discriminatorio da parte delle features estratte e, alla conclusione delle relative iterazioni, sono stati ottenuti dei risultati molto promettenti. In conferma di ciò, sono state condotte due differenti prove di classificazione: una mediante Coarse Decision Tree e l'altra con regressione logistica; entrambe hanno mostrato un'accuratezza opportunamente validata del 81.2% che ha così confermato la bontà dei parametri e dei metodi adoperati per l'analisi.

Indice

Elenco delle figure

Elenco delle tabelle

Acronimi

APQ – Amplitude Perturbation Quotient.

AUC – Area Under Curve.

CPPS – Cepstral Peak Prominence Smoothed.

FFT – Fast Fourier Transform.

FN – False Negative.

Fo – Fundamental frequency.

FP – False Positive.

HNR – Harmonic-to-Noise Ratio.

INFVo – Intelligibility, Noise, Fluency and Voice scale.

Jita – Jitter assoluto.

Jitt – Jitter percentuale.

MFCC – Mel-Frequency Cepstral Coefficients.

OPHL – Open Partial Horizontal Laryngectomy.

PPQ – Pitch period Perturbation Quotient.

RAP – Relative Average Perturbation.

ROC – Receiver Operating Characteristic curve.

SE – Standard Error.

ShdB – Shimmer in dB.

Shim – Shimmer percentuale.

SPI – Soft Phonation Index.

TN – True Negative.

TP – True Positive.

vAm – coefficient of Amplitude Variation.

vFo – coefficient of Fundamental frequency Variation.

Capitolo 1

Introduzione

In questo capitolo si presenta una descrizione generale dell'anatomia del sistema coinvolto nella generazione e modulazione della voce, le principali patologie a carico di tale apparato, i diversi trattamenti clinici adottabili ed i relativi percorsi riabilitativi così da avere le nozioni necessarie per comprendere il contesto in cui si inserisce il lavoro di questa tesi. In conclusione, si mostra una panoramica delle metodologie di valutazione degli interventi chirurgici presentate in studi passati e dei recenti tentativi di costruzione di modelli di diagnostica precoce in soggetti potenzialmente a rischio. In particolare, alcune delle nozioni presenti in queste ricerche hanno fatto da base per chiarire le scelte adottate nel prosieguo di questo scritto, con un accento deciso in merito al cepstrum, di cui è presentata la teoria e le principali caratteristiche ricavabili.

1.1 Anatomia, funzionamento e condizioni mediche

Apparato fonatorio

L'apparato fonatorio è l'insieme di tutte le strutture coinvolte nella creazione e modulazione di fonemi. Esso si compone di alcune porzioni a livello della bocca, delle cavità nasali e del collo ed espleta il suo compito durante la fase di espirazione, usufruendo solo di una piccola quantità dell'aria emessa. Ogni sezione appartenente all'apparato è specializzata ad un compito preciso, il quale è sempre riconducibile alla generazione o alla modulazione del suono. Fanno parte del sistema:

- *La laringe*, l'organo situato nella parte antero-superiore del collo ed in cui risiedono le corde vocali;
- *La faringe*, il condotto situato nella porzione finale della bocca che mette in comunicazione la cavità orale e nasale con la laringe;

- *La cavità orale* che comprende la lingua, il palato, il velo palatino, le labbra ed i denti;
- *La cavità nasale*, sita al di sopra della cavità orale.

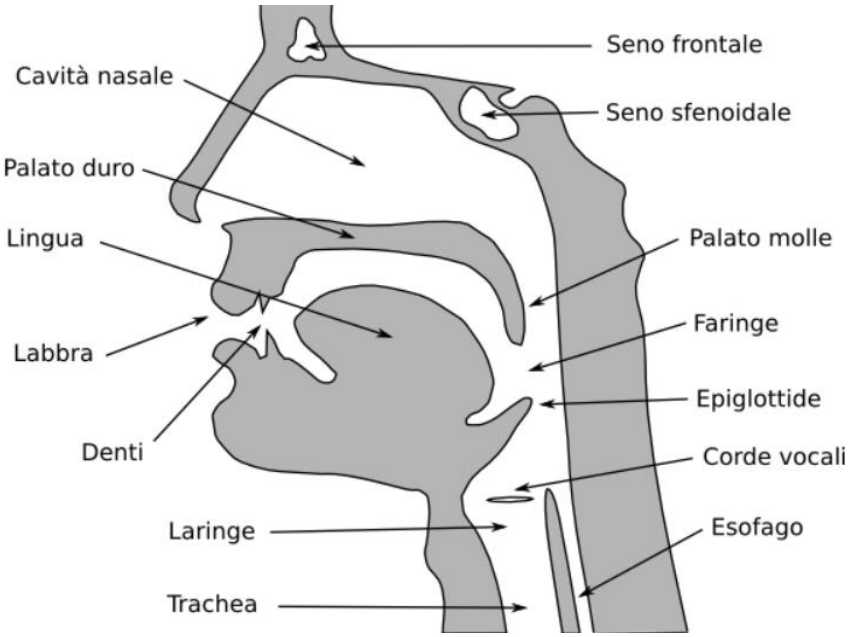

***Figura 1**: Anatomia dell'apparato fonatorio [1].*

La cavità nasale, oltre alla filtrazione, al riscaldamento e all'umidificazione dell'aria inspirata, ha la funzione di regolare le vibrazioni dei suoni. La faringe, detta comunemente gola, è una via di transito sia per l'aria che per il cibo, ospita le tonsille ed è una camera di risonanza per i suoni prodotti. La laringe è composta da cartilagine rivestita da mucosa ed ha una forma piramidale triangolare. È comunemente suddivisa in tre zone: la *zona sopraglottidea*, ossia la sezione superiore che si estende dalla cartilagine laringea, detta epiglottide, fino al paio superiore delle pieghe vestibolari, dette anche corde vocali false. Esse sono delle pieghe formate dalle mucose e sono responsabili della produzione dei fonemi con toni più profondi; la *zona glottidea*, la sezione intermedia, è la sede delle corde vocali vere (il paio inferiore delle pieghe prima citate), appaiono come una coppia di lembi formata da un complesso di legamenti, muscoli ed epitelio squamoso che delimitano uno spazio variabile, detto rima della

glottide (o soltanto glottide); la *zona sottoglottidea*, ossia la sezione inferiore, si estende dalla porzione subito successiva alla glottide fino alla cartilagine laringea, nota come cartilagine cricoide, dove risulta in diretto collegamento con la trachea.

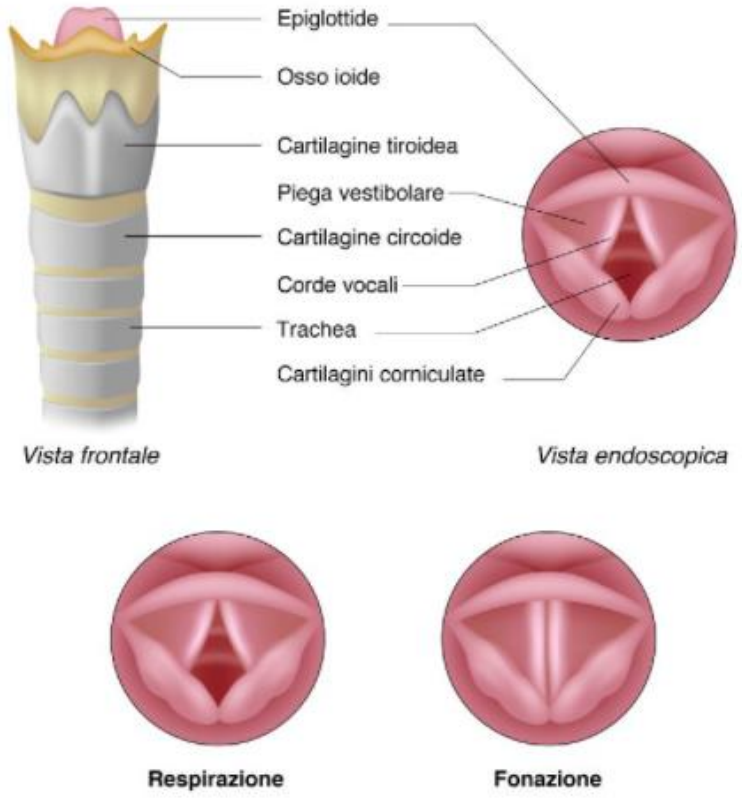

***Figura 2**: Vista frontale ed endoscopica della laringe [2].*

Fonazione

La generazione del suono ha origine dai polmoni, i quali contraendosi, a seguito dell'attivazione dei muscoli intercostali e del diaframma, generano un flusso di aria continuo. Questo, risalendo lungo la trachea, attraversa la laringe, s'infrange contro le corde vocali, le quali avranno una forma ed un'apertura specifica regolata da un articolato gruppo di muscoli, e subisce un repentino aumento di pressione. Quando il valore raggiunto da tale pressione è superiore al valore di tensione delle corde vocali, si ha il passaggio del flusso attraverso la glottide ed una conseguente rapida diminuzione della pressione. La corrente d'aria che oltrepassa la fessura avrà acquisito una certa oscillazione diventando a tutti gli effetti un'onda sonora, la quale verrà poi modulata

dalla restante parte dell'apparato, che ha la funzione di "ambiente risonante", ossia le diverse cavità fanno sì di mitigare le componenti fuori risonanza e accentuare quelle in prossimità alla risonanza (il cosiddetto fenomeno della risonanza) [1].
Il segnale vocale emesso è un segnale complesso, costituito generalmente da una vibrazione quasi-periodica delle corde vocali e da un rumore turbolento generato dal passaggio del flusso d'aria nei vari ambienti risonanti. A seconda del tipo di sorgente scatenante il fonema, è possibile produrre due diversi tipi di suoni:

- *suoni vocalizzati*, ossia quelli che prevedono il passaggio di aria attraverso la glottide e la conseguente vibrazione delle corde vocali. Le principali caratteristiche di questi sono la frequenza fondamentale, definita come la frequenza di apertura e chiusura delle corde vocali, e le formanti, definite come le frequenze caratteristiche attorno alle quali si presentano dei picchi d'ampiezza nel tracciato del segnale e che risultano generate dagli effetti delle cavità risonanti;
- *suoni non vocalizzati*, in cui non vengono coinvolte le corde vocali nella produzione del fonema e l'aria viene fatta passare in modo forzato attraverso una costrizione del tratto risonante, così che generi una turbolenza [3]. Un esempio è la consonante /s/ come pronunciata in "silenzio".

Il campo di frequenze fondamentali adoperato nei suoni vocalizzati è diverso per i soggetti di sesso maschile e femminile. Per i primi si parte da 80 Hz fino a 160-180 Hz, mentre per i secondi si va da 160 Hz fino ad arrivare a 260-300 Hz [4].
In generale, un piccolo cambiamento di tensione alle corde vocali, una minima modifica a carico della forma, della conformazione o dell'elasticità dalla diverse cavità risonanti genera dei segnali vocali a frequenze diverse; quindi, in considerazione di ciò, risulta fondamentale studiare il segnale in modo adeguato procedendo alla suddivisione dei diversi eventi fonatori, distinguendo tra quelli con alto contenuto armonico e quelli con maggior presenza di disturbi, i quali andrebbero a corrompere le informazioni utili.

Patologie

Sono diverse le patologie che provocano disturbi alla voce, dette anche disfonie: invecchiamento, traumi, allergie, malformazioni, laringiti, polipi alle corde vocali,

neoplasie, e tante altre. Per quanto riguarda i tumori, la loro incidenza a carico di zone come la laringe e la faringe rappresenta il 10% circa di tutte le neoplasie nel mondo per i soli uomini, mentre per le donne si approssimano attorno al 4%. In Italia, considerando la totalità dei tumori alla zona testa-collo, il numero di nuovi casi registrati all'anno arriva circa a 12000, rendendolo così la quinta neoplasia più diffusa. I tumori alla faringe risultano avere un tasso di mortalità più elevato rispetto a quelli alla laringe, nonostante abbiano, statisticamente, un'uguale incidenza. Le principali condizioni associate alla comparsa di tali neoplasie sono il fumo ed il consumo di alcol (circa il 90% dei pazienti ricade in almeno una delle due categorie), ma negli ultimi anni è emerso un incremento di casi collegati all'infezione da Papilloma Virus. Per i tumori alla faringe, l'età media al momento della diagnosi si aggira attorno ai 64 anni, mentre per le neoplasie a carico della laringe è superiore ai 55 anni. Entrambe le condizioni si mostrano inizialmente con sintomi blandi che molto spesso possono essere sottovalutati (cambio del tono della voce, difficoltà o dolore alla deglutizione, dolore all'orecchio, ecc.) ed il percorso di remissione dipende dalla zona d'insorgenza e dall'estensione del tumore. In generale, la percentuale di sopravvivenza a distanza di cinque anni dall'intervento è del 60%, fino a raggiungere picchi positivi del 90-95%, per soggetti con tumori limitati, e picchi negativi del 19% per soggetti aventi neoplasie diffuse [5].

1.2 Laringectomie e riabilitazione

A seguito della diagnosi di una neoplasia, il trattamento scelto prescinde dalla gravità della situazione che si presenta. In caso di tumori a stadio avanzato, il principale trattamento è l'intervento chirurgico, detto laringectomia, che consta nell'asportazione di una sezione della laringe o, nei casi più gravi, della sua totalità. A secondo dell'area asportata, l'intervento viene distinto in:

- *Laringectomia parziale*, consiste nell'asportazione di una porzione limitata dell'organo. Include diverse tipologie, tra cui la laringectomia sopraglottica (la quale prevede la resezione di una porzione o di tutta la regione sovrastante la glottide) e la laringectomia subtotale ricostruttiva (caratterizzata dalla ricostruzione delle regioni rimaste integre mediante la tecnica di ancoraggio);

- *Laringectomia totale*, consta nella rimozione in toto della laringe, andando a coinvolgere, nei casi più gravi, anche le strutture limitrofe. È, quindi, un tipo di trattamento applicato alle situazioni più critiche, viene anche adoperato in quei casi in cui terapie alternative non hanno dato l'esito sperato [6].

Negli ultimi anni, in alternativa al trattamento chirurgico, sono state proposte delle terapie basate sulla preservazione dell'organo. Queste prevedono un trattamento radioterapico affiancato ad uno chemioterapico ed hanno mostrato un'efficacia pari a quella ottenuta per via chirurgica. Tuttavia, l'uso di sostanze tossiche possono innescare diverse complicazioni, portando anche ad una non completa riacquisizione delle funzionalità o, addirittura, ad un peggioramento complessivo del paziente [7].
Le laringectomie parziali necessitano di un'apertura temporanea della trachea da cui poter asportare i vari tessuti ed hanno il vantaggio di ripristinare la fonazione e l'alimentazione per vie naturali. Inizialmente, per i pazienti si ha un'inevitabile difficoltà a adempiere entrambe le funzioni, con particolare disagio per la deglutizione che può comportare tosse o innescare ulteriori complicazioni come bronchiti e broncopolmoniti. Per ottenere un apprezzabile recupero di queste funzioni, risulta fondamentale la riabilitazione durante il decorso post-operatorio, generalmente eseguita tramite sedute con un logopedista che predispone esercizi mirati per il soggetto [8].
Nelle laringectomie totali, invece, si ha una separazione fissa delle vie aeree da quelle digestive. Questo rende obbligatorio un'apertura permanente della trachea e porta il paziente a scontrarsi con una serie di difficoltà dovute alla sua nuova condizione: l'aria arriva direttamente ai polmoni senza essere stata precedentemente filtrata, umidificata e riscaldata, rendendolo più incline a problematiche respiratorie; è necessario impedire in tutti i modi che l'acqua entri in tale fessura, accrescendo le difficoltà del soggetto nelle operazioni di lavaggio. Oltre a queste complicazioni, il paziente mostra una perdita nella coordinazione respiratoria, un'iniziale mancanza della percezione olfattiva ed una deglutizione parzialmente compromessa, nonostante questa avvenga senza il rischio di inalazione della saliva. In aggiunta, è presente anche un'elevata disfonia che costringe ad una laboriosa ricerca dei metodi di recupero, solitamente ricondotta a metodologie come: la voce esofagea (aria immagazzinata nell'esofago che viene poi eruttata portando alla vibrazione dello sfintere esofageo e delle strutture limitrofe) o la valvola fonatoria (una valvola che se occlusa con un dito permette all'aria espirata di passare attraverso le strutture sovrastanti) [6]. Tutto ciò si traduce, in generale, in un attento

lavoro di formazione del soggetto riguardo l'uso degli strumenti sostitutivi ed un'intensa attività riabilitativa che gli permetterà il recupero della gestione polmonare, del senso dell'olfatto e della deglutizione [8].

1.3 Stato dell'arte - Metodi e parametri

Gli studi condotti nell'ambito dell'analisi dei segnali vocali di soggetti disfonici sono stati rivolti al raggiungimento di diversi scopi. Alcuni valutano l'efficacia di un determinato intervento chirurgico mediante la valutazione delle capacità riacquisite dai pazienti (Di Nicola et al. [9], Markeieff et al. [10], Kosztyła-Hojna et al. [11]). Alcuni confrontano l'efficacia di due diversi tipi di trattamento tramite i risultati ottenuti dai relativi pazienti (Krengli et al. [12]). Altri ancora confrontano diversi percorsi riabilitativi per stabilirne quale sia il migliore o, più in generale, per valutarne l'efficacia (Van Sluis et al. [13], Chhestri et al. [14]). Tutti questi studi hanno in comune alcuni passi fondamentali: eseguono delle estrazioni, mediante software dedicati, di alcuni parametri acustici scelti tra la grande varietà disponibile; a questi sono affiancati dati percettivi estrapolati da apposite scale, i quali hanno lo scopo di dare maggior consistenza ai valori assunti dai parametri estratti; infine, si eseguono delle operazioni di classificazione, di varia tipologia a seconda dello studio, per validare l'accuratezza delle informazioni estratte dai parametri e comprenderne l'efficacia. Tra i diversi parametri estraibili si menzionano, in tabella 1, alcuni tra i più comunemente usati:

***Tabella 1**: Elenco di alcuni dei più diffusi parametri acustici.*

Parametri	Descrizione
Fo	*Frequenza fondamentale media* (Hz)
vFo	*Variazioni della frequenza fondamentale* (%)
Jitt	*Jitter percentuale* (%) – stabilità d'ampiezza.
Shim	*Shimmer percentuale* (%) – stabilità del periodo.
vAm	*Variazione del picco d'ampiezza* (%) – stabilità d'ampiezza.

SPI	*Soft Phonation Index* – misura la struttura armonica dello spettro [9].
HNR	*Harmonic-to-Noise Ratio* – valuta la potenza della componente armonica del segnale [15].
DUV	*Degree of Voiceless* (%) – è una stima relativa sulla valutazione delle aree non armoniche (dove non è possibile riscontrare la presenza di Fo) all'interno del campione vocale [16].

Le scale percettive sono molto usate in ambito clinico per valutare le competenze ed i miglioramenti fonatori riscontrati nei pazienti durante le varie sedute logopediche. Esse constano, in linea di massima, in un elenco di caratteristiche vocali a cui uno o più esperti del settore associa un voto. In genere, ad un voto basso corrisponde una bassa qualità della voce o della caratteristica in esame, mentre ad un voto alto si ha la situazione contraria. Esistono delle scale compilate dagli stessi pazienti, definiti questionari sulla qualità della vita, da cui è possibile estrapolare dati di natura percettiva dal diretto interessato. Vi sono anche scale che determinano lo stato di una determinata patologia e al contempo riescono a dare informazioni sulle competenze fonatorie, come accade per i soggetti affetti dalla malattia di Parkinson. Tra le principali scale percettive si menzionano quelle riassunte in tabella 2:

***Tabella 2**: Elenco di alcune delle più diffuse scale percettive.*

Scale	Descrizione
GRBAS	*Global Roughness Breathiness Asthenia Strain scale* – valuta il grado globale della disfonia, il grado di raucedine, il grado di fievolezza della voce, l'astenia vocale e lo sforzo vocale.
V-RQOL	*Voice – Related Quality of Life measure* – questionario compilato direttamente dal paziente.
VHI	*Voice Handicap Index* – questionario compilato direttamente dal paziente.

INFVo	*Intellegibility, Noise, Fluency and Voice* – restituisce un punteggio da 0 (molto buono) a 10 (molto scarso) ad ognuna delle caratteristiche vocali da cui la scala prende il nome.
UPDRS	*Unifed Parkinson Disease Rating Scale* – questionario che identifica lo stato di un generico paziente parkinsoniano e dedica alcune sue sezioni alla valutazione della capacità fonatoria.

Alcune ricerche, invece, si sono discostate da questi scopi valutativi per tentare di mettere a punto un metodo di diagnostica di patologie più gravi su soggetti che mostrano sintomi di disfonia. Una delle principali applicazioni è stata quella sui pazienti affetti dalla malattia di Parkinson [17][18][19]. Solo all'alba di tali studi, quindi solo a partire dagli anni Dieci del secondo millennio, è stato intensificato lo studio di parametri appartenenti al dominio spettrale e cepstrale nell'analisi dei segnali vocali. Inizialmente relegati al solo uso nella pratica del riconoscimento vocale, a parte rare eccezioni (César et al. [20]), hanno poi trovato spazio in vari ambiti del campo biomedicale nel corso dell'ultimo decennio, mostrando anche qualche risultate promettente.

1.3.1 Cepstrum

La prima volta che si sentì parlare di cepstrum era il 1963 in uno studio di Bogert e colleghi [21], in cui veniva presentato come una tecnica euristica in grado di trovare i tempi di arrivo dell'eco di un segnale composto [22]. La parola cepstrum deriva dall'anagramma della parola "spectrum" e viene definito come lo spettro di potenza del logaritmo dello spettro di potenza del segnale, ovvero: applicando la trasformata di Fourier al segnale di partenza si ottiene il suo spettro di potenza, si calcola il logaritmo di quest'ultimo ed al risultato ricavato è nuovamente applicata la trasformata di Fourier, come se si volesse ottenere il relativo spettro di potenza [23]. Quindi, in formula:

$$C_p(\tau) = \mathcal{F}\{\log|\mathcal{F}[x(t)]|^2\}^2 \qquad (1.1)$$

Dove *x(t)* è il segnale temporale, $\mathcal{F}$ indica la trasformata di Fourier, $|\mathcal{F}[x(t)]|^2$ indica lo spettro di potenza del sagnale e τ è la variabile che definisce il dominio del cepstrum,

che non essendo né il dominio temporale né quello frequenziale, viene chiamato quefrency (anagramma di frequency).

Prendendo l'esempio specifico della produzione della voce, il segnale generato *y(t)* è la composizione di una componente derivata dalla sorgente vocale, *f(t)*, e da una componente del tratto risonante, *h(t)*. Il segnale *y(t)* viene quindi descritto dalla convoluzione delle due componenti:

$$y(t) = f(t) * h(t) \tag{1.2}$$

Passando dal dominio temporale al dominio cepstrale, quindi grazie alle proprietà del logaritmo, secondo la quale il logaritmo del prodotto di due numeri è uguale alla somma dei logaritmi dei due, la relazione muta in:

$$\log Y(f) = \log F(f) + \log H(f) \tag{1.3}$$

e di conseguenza

$$\mathcal{F}^{-1}\{\log[Y(f)]\} = \mathcal{F}^{-1}\{\log[F(f)]\} + \mathcal{F}^{-1}\{\log[H(f)]\} \tag{1.4}$$

La quale risulta molto utile per la separazione analitica della sorgente del segnale dalla componente dovuta all'ambiente risonante [23]. Rappresentando graficamente il cepstrum è possibile evidenziare la presenza di una forte componente, ossia un picco, che corrisponde alla regolarità del picco armonico. Per segnali con una struttura armonica ben definita, quindi, il relativo grafico cepstrale mostrerà un picco molto prominente [24], così come è possibile in figura 3.

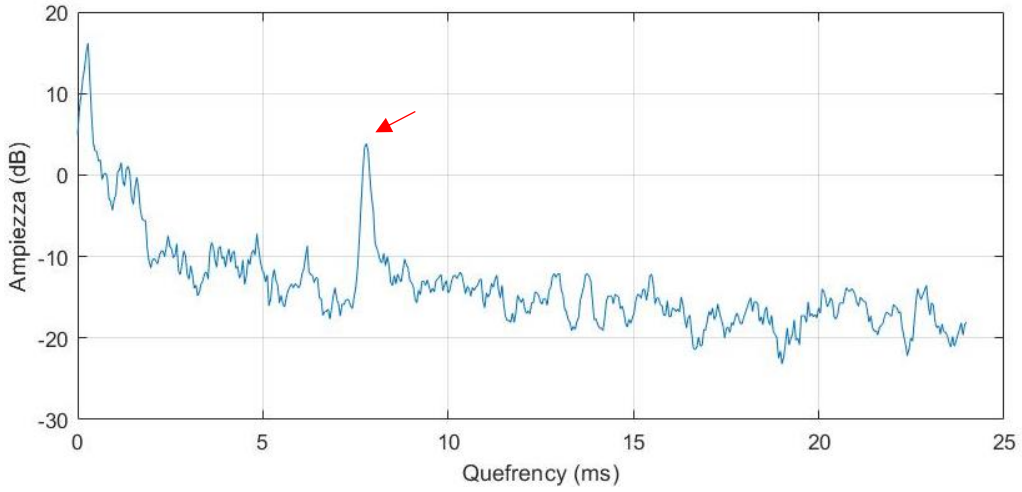

***Figura 3**: Rappresentazione del cepstrum di un soggetto sano registrato durante l'eloquio. La quefrency in cui ricade il picco indicato dalla freccia corrisponde al periodo fondamentale del segnale [24].*

I principali parametri ricavabili dal cepstrum sono riassunti in tabella 3:

***Tabella 3**: Elenco di alcuni dei principali parametri cepstrali.*

Parametri	Descrizione
CPP	*Cepstral Peak Prominence* (dB) – è una misura dell'ampiezza del picco cepstrale, normalizzata su tutta l'ampiezza del segnale attraverso una linea di regressione lineare che correla la quefrency con l'ampiezza cepstrale [4]. Di fatto, il CPP viene ottenuto come la differenza di ampiezza (in dB) tra il picco cepstrale ed il corrispondente valore della linea di regressione che sta sotto al picco [24].
CPPS	*Cepstral Peak Prominence Smoothed* – è un CPP a cui viene apportata una piccola modifica, ossia è eseguito uno smoothing del cepstrum del segnale prima di estrarne il picco e di calcolarne la relativa ampiezza. Negli algoritmi in cui è stato adoperato, tale modifica del CPP ha mostrato un netto miglioramento dei risultati di predizione [24].

MFCC	*Mel-Frequency Cepstral Coefficients* – sono coefficienti ottenuti dalla separazione dello spettro in diverse finestre (dette anche *frame*) e scalate secondo una scala detta di Mel. In pratica, lo spettro è suddiviso mediante un certo numero di finestre triangolari, è eseguita una trasformata di Fourier del logaritmo del modulo su ciascuno di essi e l'energia media ricavata da ogni banda è detta coefficiente di Mel [19].

Negli ultimi anni, tali parametri, hanno trovato ampio spazio nell'ambiente biomedicale, venendo adoperati per diversi scopi. Ad esempio, per l'analisi di pazienti affetti da disfonia (Hillenbrand et al. [24], Castellana et al. [4], Pietruch et al. [25]) o, come precedentemente citato, per l'implementazione di metodi di diagnostica (Atzori et al. [17], Soumaya et al. [18], Karan et al. [19]).

In generale, tali studi non hanno mai dato un apporto decisivo alle terapie riabilitative, almeno per quanto riguarda l'affiancamento del logopedista durante le sedute o un possibile generale aiuto da offrire al paziente. Alcuni ricercatori hanno intrapreso i primi passi in tale direzione, presentando qualche risultato incoraggiante, ma il percorso è ben lungi dall'essere completato. Sono necessari ulteriori studi che possano mettere in luce metodologie e dati sempre più coerenti ed efficaci, in modo da realizzare gli ambiziosi obiettivi auspicati.

Capitolo 2

Materiali e Metodi

Alla luce delle promettenti qualità riscontrate nei vari studi menzionati nel paragrafo 1.3, in questo lavoro di tesi sono stati esaminati i parametri del dominio cepstrale e spettrale, insieme ad altri introdotti per l'occasione, per estrarre informazioni di valenza riabilitativa dall'analisi degli eloqui di pazienti sottoposti a laringectomia. Nel seguente capitolo, in particolare, sono state espresse tutte le fasi del lavoro svolto per giungere a tale scopo. Sono stati presentati i dati a disposizione, i metodi inizialmente provati, quelli introdotti per sopperire alle criticità riscontrate e tutta la relativa trattazione sui parametri che sono stati definiti adeguati allo studio. Infine, sono state illustrate le procedure per la validazione di quest'ultimi mediante confronti statistici ed operazioni di classificazione.

2.1 Dati

Per l'esecuzione di questo studio sono stati selezionati 32 tracciati vocali facenti capo ad altrettanti pazienti laringectomizzati, i quali si differenziano per il tipo di laringectomia parziale a cui sono stati sottoposti. Queste operazioni fanno parte della famiglia delle laringectomie parziali orizzontali e sono:

- *Laringectomia sopracricoidea* (detta anche Open Partial Horizontal Laryngectomy di tipo II, OPHL II) [7], di cui ne fanno parte 22 soggetti dei 32 totali;
- *Laringectomia sopratracheale* (detta anche Open Partial Horizontal Laryngectomy di tipo III, OPHL III) [7], di cui ne fanno parte i restanti 10 soggetti.

La maggior parte dei pazienti è di sesso maschile, ma considerata la bassa qualità della fonazione e del relativo segnale audio acquisito da ognuno di essi, è stato scelto un range di frequenza fondamentale generale che si estende da 80 Hz fino a 400 Hz. I pazienti presentano un'età media di 66 anni e tutti sono stati sottoposti a terapie riabilitative nelle settimane o nei mesi precedenti alla data di registrazione dei suddetti tracciati.

Questi sono stati incisi in un ambiente il più possibile privo di disturbi rumorosi, mediante un microfono in aria e con una frequenza di campionamento di 50000 Hz. La mansione richiesta è stata la lettura di brano foneticamente scelto, il quale è presentato qui di seguito:

> *"Notturno.*
> *Vi è un profondo silenzio nel buio della notte. Vicino al pozzo, nella cui acqua si specchiano la luna ed una scia di stelle, la magnolia stende i suoi rami, cespugli di rose olezzano nell'aria. Il temporale è cessato e la pioggia, ormai, non cade più. Solo le rane gracidano nei fossi oltre quel prato."*

Si tratta di un testo in cui la frequenza di comparsa di ogni fonema è uguale per ciascuno di essi. In particolare, per questo brano si hanno, oltre alle inevitabili porzioni di registrazione silenziose dovute alle pause occorse tra una parola e l'altra, sezioni armoniche e sezioni non armoniche equamente presenti.
A fianco di ogni soggetto è stata fornita la relativa scala INFVo, precedentemente presentata in tabella 2. Essa è stata redatta da un'equipe di esperti del settore, i quali hanno assegnato un punteggio da 0 (molto buono) a 10 (molto scarso) ad ognuna delle seguenti caratteristiche vocali: intellegibilità (I), voce eccessivamente rumorosa (N), fluidità nel parlato (F) ed il grado di vocalizzazione (Vo), o meglio la capacità di articolare diverse intonazioni durante l'emissione di suoni. Nel prosieguo del lavoro è stato impiegato solo il parametro I, i cui valori sono stati mostrati qui di seguito.

***Tabella 4**: Elenco dei pazienti trattati con OPHL III.*

Soggetto n°	Sesso	Età (anni)	I (INFVo)
1	M	55	6,2
2	M	77	1,8
3	M	65	4,3
4	M	61	6,7
5	M	41	3,2
6	M	69	6,8

7	M	65	9,7
8	M	63	5,9
9	M	68	6,6
10	M	69	2,7

***Tabella 5**: Elenco dei pazienti trattati con OPHL II.*

Soggetto n°	Sesso	Età (anni)	I (INFVo)
1	M	78	7,6
2	M	74	3,7
3	M	69	0,9
4	M	66	1,3
5	F	65	3,6
6	M	67	1,3
7	M	79	2,5
8	M	55	0,8
9	M	67	3,5
10	M	74	2,2
11	F	76	8,3
12	M	74	6,9
13	M	65	2,7
14	M	63	0,8
15	M	71	3,6
16	M	60	2,3
17	M	55	0,5
18	M	85	6,3
19	M	62	8,8
20	M	51	0
21	M	47	1,2
22	M	75	1,5

Infine, sono stati introdotti dieci campioni vocali di soggetti considerati sani privi di qualsiasi forma di disfonia, di cui ne fanno parte cinque maschi e cinque femmine, i quali hanno riprodotto lo stesso brano foneticamente bilanciato dei pazienti laringectomizzati. Questi tracciati sono stati acquisiti mediante l'applicazione di registrazione vocale presente nel relativo cellulare di ciascun soggetto, in un ambiente quanto più possibile privo di disturbi. Nonostante la lunga serie di problematiche correlate a questo metodo di acquisizione, non è stato possibile fare altrimenti a causa della situazione pandemica che ha limitato di molto le attività di ricerca. In ragion di ciò, è stata adottata tale procedura, la quale può essere vista anche nell'ottica di rappresentare, in una certa misura, lo scenario di un futuro soggetto disfonico che esegua la terapia riabilitativa in autonomia, usufruendo dei soli strumenti a sua disposizione.

2.2 Pre-processing

La pre-elaborazione dei segnali vocali è stata svolta allo scopo di ottenere i relativi tracciati privati dei principali disturbi e di avere, di conseguenza, un'elevata fiducia nei risultati conclusivi del lavoro. Inizialmente è stato necessario ripulire, mediante l'ausilio del software *Audacity*, ciascun brano vocale dalla presenza di alcuni suoni indesiderati, come voci esterne o rumori ambientali, che sono rimasti incisi nel tracciato al momento della registrazione. In questo modo è stato possibile ricavare dei segnali in cui sia presente solo il parlato del paziente.

I segnali ottenuti sono stati caricati all'interno dell'ambiente Matlab (R2021a) e rappresentati sotto forma di vettori contenenti le ampiezze campionate lungo tutta la durata del tracciato. È stato eseguito il sottocampionamento di ciascuno vettore ad una frequenza di 44100 Hz, in modo da uniformarsi allo standard presente in tutti gli studi bibliografici del settore. Infine, in ogni campione vocale è stato rimosso l'offset ed è stata eseguita la normalizzazione rispetto al valore assoluto del rispettivo massimo. Ciò è stato realizzato allo scopo di ricavare una dinamica del segnale uniforme per ciascun paziente. Successivamente, è stata attuata l'esclusione delle porzioni di segnale silenziose dai rispettivi vettori. Il procedimento adottato è stato quello di suddividere ogni tracciato in finestre di lunghezza fissa, di esaminare ciascuna di esse ricavandone il valore efficace (indicato con rms_f nello schema a blocchi presentato in figura 4) e,

infine, di confrontare tali valori con una soglia di riferimento presa pari alla metà del valore efficace dell'intero segnale (indicato nello schema come rms_th).

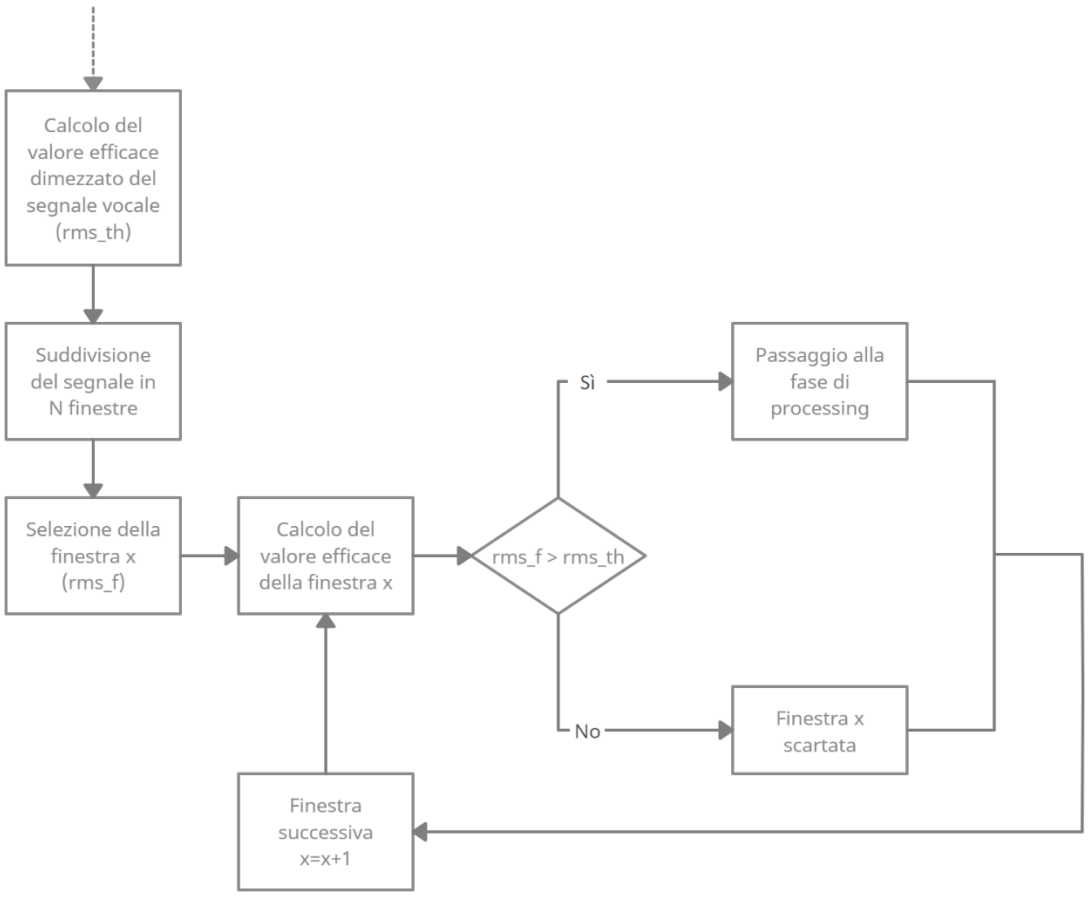

***Figura 4**: Schema a blocchi della rimozione dei silenzi.*

La finestra esaminata in un'iterazione generica è stata giudicata come contenente segnale valido e, quindi, idonea alla prosecuzione delle operazioni, solo se il suo valore efficace risulta maggiore della soglia. In caso contrario, essa è stata valutata come porzione di segnale silenzioso, quindi conseguentemente scartata, e l'analisi è stata trasferita al frame successivo.

2.3 Prima versione dell'elaborazione

In un primo approccio a questo lavoro di tesi è stato eseguito un metodo che non ha portato a risultati soddisfacenti. Esso è stato incentrato sul tentativo di estrarre i parametri d'interesse solo dalle porzioni "valide" di segnale, in modo da ottenere informazioni efficaci e meno corrette da disturbi. Ciò è stato eseguito mediante una condizione che permette di distinguere le finestre armoniche di segnale da quelle non armoniche. Essendo il rapporto armoniche-rumore (HNR) in grado di quantificare il livello di armoniosità all'interno del segnale, esso è stato selezionato come parametro per l'esecuzione del confronto sopra menzionato ed è stato calcolato a partire dall'autocorrelazione del segnale. In aggiunta, per rimuovere i frame silenziosi è stata eseguita la procedura descritta al paragrafo 2.2, con l'unica differenza che la misura della lunghezza della finestra è stata impostata in funzione del massimo valore assunto dalla funzione di autocorrelazione e, quindi, risulta variabile da iterazione a iterazione.

La teoria dei segnali afferma che, nel caso di segnali nel tempo stazionari (ossia un segnale le cui statistiche rimangono costanti per tutto il tempo), l'autocorrelazione è una funzione del ritardo τ così definita:

$$r_x(\tau) = \int x(t)\, x(t+\tau)\, dt \qquad (2.1)$$

Tale funzione risulta avere un massimo assoluto per $\tau=0$ che corrisponde anche alla potenza del segnale. Prendendo l'esempio di un segnale periodico $h(t)$ avente un periodo T_0 (ossia il fattore di ritardo che indica la posizione del primo massimo locale della funzione) a cui viene aggiunto del rumore $n(t)$; se queste due parti risultano indipendenti, allora il calcolo dell'autocorrelazione del segnale totale sarà pari alla somma delle singole autocorrelazioni. In particolare, per $\tau=0$ si ottiene $r_x(0) = r_h(0) + r_n(0)$ e se il rumore aggiunto è bianco (ossia ha uno spettro costante), allora è presente un massimo locale al ritardo $\tau_{max}=T_0$ con un'altezza pari a $r_x(t_{max})=r_h(T_0)=r_h(0)$. In conclusione, poiché il valore del massimo assoluto al ritardo zero è pari alla potenza del segnale, allora l'autocorrelazione normalizzata al ritardo τ_{max} rappresenta la potenza relativa della componente periodica (detta anche armonica) del segnale ed il suo complemento la potenza relativa alla componente rumorosa:

$$a)\ r'_x(\tau_{max}) = \frac{r_h(0)}{r_x(0)}; \qquad b)\ 1 - r'_x(\tau_{max}) = \frac{r_n(0)}{r_x(0)} \qquad (2.2)$$

Da qui, infine, è possibile definire il rapporto armoniche-rumore in forma logaritmica [15]:

$$HNR(in\ dB) = 10\log_{10}\frac{r'_x(\tau_{max})}{1 - r'_x(\tau_{max})} \qquad (2.3)$$

È da notare come a valori elevati di HNR corrisponde un'elevata presenza di componenti armoniche rispetto a quelle rumorose, mentre per valore bassi di HNR si ha la situazione opposta.

Le ultime tre equazioni riportate sono state implementate nell'ambiente Matlab al fine di realizzare la condizione di discriminazione delle finestre. Una volta trovato l'HNR della finestra in esame, questo è stato messo a confronto con una soglia fissa di -40 dB. La scelta, nel caso dell'analisi di soggetti sani, è ricondotta solitamente a 0 dB, ma nel caso dei soggetti disfonici studiati è stato necessario ampliare tale valore a causa della loro scarsa qualità vocale. Tuttavia, la soglia risulta essere eccessivamente tollerante rispetto a quanto auspicato. Ciò è stato tradotto in un'accettazione della quasi totalità dei frame analizzati ed una conseguente compromissione del rendimento di questo procedimento, poiché è stato consentito di estrarre informazioni anche da porzioni di segnale potenzialmente non armoniche. In generale, continuando con la trattazione, se la finestra in questione è stata in grado di soddisfare la condizione dell'HNR, allora da essa sono stati estratti i seguenti parametri:

- *Jitter* (%): rappresenta la variabilità percentuale relativa da periodo a periodo, ossia mostra la valutazione della variabilità del tono del periodo [9]. È stata calcolata come

$$Jitt = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|To^{(i)} - To^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} To^{(i)}} \qquad (2.4)$$

  dove $To^{(i)}$ rappresenta il periodo in esame estratto ed N è il numero totale dei periodi ricavati. Risulta non eccessivamente influenzata dalla frequenza

fondamentale media del campione vocale [16].

- *Jitter assoluto* (µs): rappresenta la variabilità relativa da periodo a periodo ed è stata ottenuta da

$$Jita = \frac{1}{N-1}\sum_{i=1}^{N-1}\left|To^{(i)} - To^{(i+1)}\right| \qquad (2.5)$$

dove $To^{(i)}$ rappresenta il periodo in esame estratto ed *N* è il numero totale di periodi ricavati. È fortemente influenzata dalla frequenza fondamentale media della voce del soggetto [16].

- *RAP* (relative average perturbation, %): è la stima relativa della variabilità del tono da periodo a periodo con un fattore di smoothing pari a tre periodi. È stato calcolato come

$$RAP = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\left|\frac{To^{(i-1)} + To^{(i)} + To^{(i+1)}}{3} - To^{(i)}\right|}{\frac{1}{N}\sum_{i=1}^{N} To^{(i)}} \qquad (2.6)$$

dove $To^{(i)}$ rappresenta il periodo in esame estratto ed *N* è il numero totale di periodi ricavati. Lo smoothing permette di ridurre la sensibilità all'errore di estrazione del tono [16].

- *PPQ* (pitch period perturbation quotient, %): è la valutazione relativa della variabilità del tono da periodo a periodo con un fattore di smoothing pari a cinque periodi. È stato ottenuto come

$$PPQ = \frac{\frac{1}{N-4}\sum_{i=1}^{N-4}\left|\frac{1}{5}\sum_{r=0}^{4} To^{(i+r)} - To^{(i+2)}\right|}{\frac{1}{N}\sum_{i=1}^{N} To^{(i)}} \qquad (2.7)$$

dove $To^{(i)}$ rappresenta il periodo in esame estratto ed *N* è il numero totale di periodi ricavati. Lo smoothing riduce la sensibilità all'errore di estrazione del

tono [16].

- *vFo* (coefficient of Fundamental Frequency Variation, %): è la deviazione tipo relativa della frequenza fondamentale calcolata e ne riflette le variazioni a lungo termine [9]. È stata ottenuta da

$$vFo = \frac{\sigma}{Fo} = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N} Fo^{(j)} - Fo^{(i)}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}} \qquad (2.8)$$

dove $Fo = \frac{1}{N}\sum_{i=1}^{N} Fo^{(i)}$, $Fo^{(i)} = \frac{1}{To^{(i)}}$ è il valore della frequenza fondamentale periodo per periodo ed *N* è il numero totale di periodi estratti [16].

- *Shimmer* (%): rappresenta la variabilità relativa da periodo a periodo dell'ampiezza picco-picco del segnale vocale [9]. È stata calcolata come

$$Shim = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}\left|A^{(i)} - A^{(i+1)}\right|}{\frac{1}{N}\sum_{i=1}^{N} A^{(i)}} \qquad (2.9)$$

dove $A^{(i)}$ sono le varie ampiezze picco-picco ricavate ed *N* è il numero totale di periodi estratti [16].

- *Shimmer assoluto* (dB): rappresenta la variabilità relativa da periodo a periodo in dB dell'ampiezza picco-picco del campione vocale. È stata ottenuta da

$$ShdB = \frac{1}{N}\sum_{i=1}^{N-1}\left|20\log\left(\frac{A^{(i+1)}}{A^{(i)}}\right)\right| \qquad (2.10)$$

dove $A^{(i)}$ sono le varie ampiezze picco-picco ricavate ed *N* è il numero totale di periodi estratti [16].

- *APQ* (Amplitude Perturbation Quotient, %): è la valutazione relativa della variabilità dell'ampiezza picco-picco da periodo a periodo con uno smoothing di undici periodi. È stato calcolato come

$$APQ = \frac{\frac{1}{N-10}\sum_{i=1}^{N-10}\left|\frac{1}{11}\sum_{r=0}^{10}A^{(i+r)} - A^{(i+5)}\right|}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \qquad (2.11)$$

  Dove $A^{(i)}$ sono le varie ampiezze picco-picco ricavate ed *N* è il numero totale di periodi estratti. Lo smoothing riduce la sensibilità all'errore di estrazione [16].

- *vAm* (Coefficient of Amplitude Variation, %): è la deviazione tipo dell'ampiezza picco-picco calcolata da periodo a periodo e rappresenta la variazione molto a lungo termine [9]. È stata ottenuta da

$$vAm = \frac{\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{N}\sum_{j=1}^{N}A^{(j)} - A^{(i)}\right)^2}}{\frac{1}{N}\sum_{i=1}^{N}A^{(i)}} \qquad (2.12)$$

  dove $A^{(i)}$ sono le varie ampiezze picco-picco ricavate ed *N* è il numero totale di periodi estratti [16].

- *Dt_v/uv* (%): rappresenta il rapporto tra suoni vocalizzati e suoni non vocalizzati. È stato ottenuto tenendo conto dei frame di segnale definiti come contenenti porzioni di segnale armonico e di quelli contenenti porzioni non armoniche, per poi svolgere

$$Dt_{v/uv} = 100\frac{Har}{Har + Unhar} \qquad (2.13)$$

  dove *Har* rappresenta la totalità delle finestre armoniche e *Unhar* la totalità delle finestre non armoniche.

Oltre a questi sono anche stati memorizzati gli andamenti dei valori di HNR, del valore efficace (RMS), del CPPS e della frequenza fondamentale (Fo) in tutti i frame considerati armonici. Da ogni evoluzione temporale ottenuta sono state calcolate le relative statistiche descrittive, ossia sono stati ricavati la media, la mediana, la moda, la deviazione tipo, il range, il 5° percentile, il 95° percentile, la skewness e la kurtosis per ognuna di esse. Infine, tutte le distribuzioni di ogni andamento sono state unite ai parametri sopra elencati, in modo da costituire il data set completo da somministrare agli algoritmi di classificazione. La scelta dei metodi per quest'ultimo passo è stata indirizzata al Coarse Decison Tree e alla regressione logistica. Essi sono stati adoperati anche per la versione definita e la spiegazione del loro funzionamento è stata rimandata al relativo paragrafo 2.4.4, ma in questa versione dell'elaborato hanno riportato un modesto risultato in termini di accuratezza. In particolare, per il Coarse Decision Tree è stata rinvenuta un'accuratezza pari al 87.5%, mentre per la regressione logistica è stata del 75%. In quest'ultimo, inoltre, è stata impostata una condizione di attendibilità più tollerante rispetto alla prassi, ossia è stato predisposto un requisito sul limite di fiducia dei risultati rinvenuti pari al 90%, anziché il consueto 95% riportato nella totalità degli studi scientifici. Senza contare, in aggiunta, che entrambi i modelli sono stati implementati con l'uso di tutto il dataset a disposizione, quindi, in assenza di una vera e propria fase di validazione, la quale, oltre a restituire dei risultati più realistici, porterebbe ad un ulteriore calo delle stime di accuratezza.

Questo approccio, appena descritto, è stato particolarmente utile nell'analisi di segnali costituiti dall'emissione di una vocale continua, ma nel caso dello studio riportato, in cui sono stati presi in esame gli eloqui dei soggetti, sono state rilevate evidenti difficoltà che hanno portato a dei risultati statisticamente non accettabili. Senza dubbio la soglia di HNR troppo tollerante non ha permesso di estrarre con buona approssimazione le zone armoniche, ma anche la selezione di parametri non propriamente adeguati al caso ha contribuito alla non funzionalità del metodo descritto.

2.4 Elaborazione "golden standard"

Alla luce dei risultati ottenuti nella versione precedente, è stato necessario operare alcune modifiche al fine di ottenere risultati accettabili. Gli iniziali passi di pre-

processing, riguardanti la rimozione dei disturbi dovuti al volume della traccia audio e la rimozione dei frame silenziosi mediante il controllo sul valore efficace, sono stati riportati anche in questa versione dell'elaborato, con la sola modifica di considerare, per le varie iterazioni, frame di lunghezza fissa pari a 1024 campioni. Le principali problematiche riscontrate nella prima stesura sono state oggetto di tutte le analisi e le ricerche perseguite, in modo da ottenere, in seguito alla loro rettifica, un netto miglioramento dei risultati. Di seguito sono state mostrate le correzioni apportate insieme alla spiegazione di tutti i passi della procedura che ha permesso di ottenere ottimi risultati.

2.4.1 Separazione dei frame

A seguito di un'attenta ricerca bibliografica, in cui sono stati vagliati diversi parametri e procedure, l'attenzione è stata rivolta su di un particolare pacchetto dell'ambiente Matlab denominato Audio Toolbox™. In esso sono contenute una serie di funzioni in grado di fornire istruzioni per il processamento dei segnali audio, per l'analisi del parlato e per l'ottenimento di misure acustiche. Tra le varie funzioni a disposizione, ne sono fornite alcune definite descrittori spettrali, le quali riescono a descrivere la forma delle tracce audio. Tra queste istruzioni si menzionano:

- *spectralCentroid*, è la somma pesata della frequenza normalizzata rispetto alla somma non pesata. In sostanza, rappresenta il baricentro dello spettro [26]

$$\mu_1 = \frac{\sum_{k=b_1}^{b_2} f_k s_k}{\sum_{k=b_1}^{b_2} s_k} \tag{2.14}$$

  dove b_1 e b_2 sono gli estremi del range, dati in termini di frame, in cui calcolare il baricentro spettrale, f_k è la frequenza in Hz del k-esimo frame, s_k è il valore dello spettro (ampiezza o potenza spettrale sono definizioni entrambe valide) al k-esimo frame;

- *spectralSpread*, è la deviazione tipo calcolata attorno al baricentro dello spettro e ne rappresenta anche la larghezza di banda [26]

$$\mu_2 = \sqrt{\frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^2 s_k}{\sum_{k=b_1}^{b_2} s_k}} \tag{2.15}$$

dove b_1 e b_2 sono gli estremi del range, dati in termini di frame, f_k è la frequenza in Hz del k-esimo frame, s_k è il valore dello spettro al k-esimo frame e μ_1 è il baricentro spettrale;

- *spectralKurtosis,* calcolata dal momento del quarto ordine della distribuzione spettrale [26]

$$\mu_4 = \frac{\sum_{k=b_1}^{b_2}(f_k - \mu_1)^4 s_k}{(\mu_2)^4 \sum_{k=b_1}^{b_2} s_k} \tag{2.16}$$

dove b_1 e b_2 sono gli estremi del range, dati in termini di frame, f_k è la frequenza in Hz del k-esimo frame, s_k è il valore dello spettro al k-esimo frame, μ_1 è il baricentro spettrale e μ_2 è la diffusione spettrale. Tale funzione restituisce una misura della planarità dello spettro attorno al suo baricentro, ossia indica quanto l'andamento dello spettro tenda o meno a mostrare dei picchi. Infatti, in casi di rumore bianco maggiormente presente in una determinata porzione di segnale vocale, il valore di *spectralKurtosis* in tale finestra diminuirà rispetto alle precedenti, poiché il relativo spettro in tale porzione avrà una minor tendenza a mostrare dei picchi;

- *spectralEntropy*, misura la prominenza armonica dello spettro del segnale, ossia quanto lo spettro in questione presenti dei picchi di ampiezza

$$entropy = \frac{-\sum_{k=b_1}^{b_2} s_k \log(s_k)}{\log(b_2 - b_1)} \tag{2.17}$$

dove b_1 e b_2 sono gli estremi del range, dati in termini di frame, e s_k è il valore dello spettro al k-esimo frame. Tale funzione è stata ampiamente usata per il riconoscimento automatico del parlato.

Le ultime due istruzioni hanno suscitato particolare interesse e sono stati i principali oggetti delle elaborazioni atte alla ricerca di un criterio per la distinzione dei frame armonici/non armonici. In figura 5 è stato mostrato il confronto tra l'andamento dell'ampiezza, della kurtosis spettrale e dell'entropia spettrale di un campione vocale appartenente ad un soggetto disfonico.

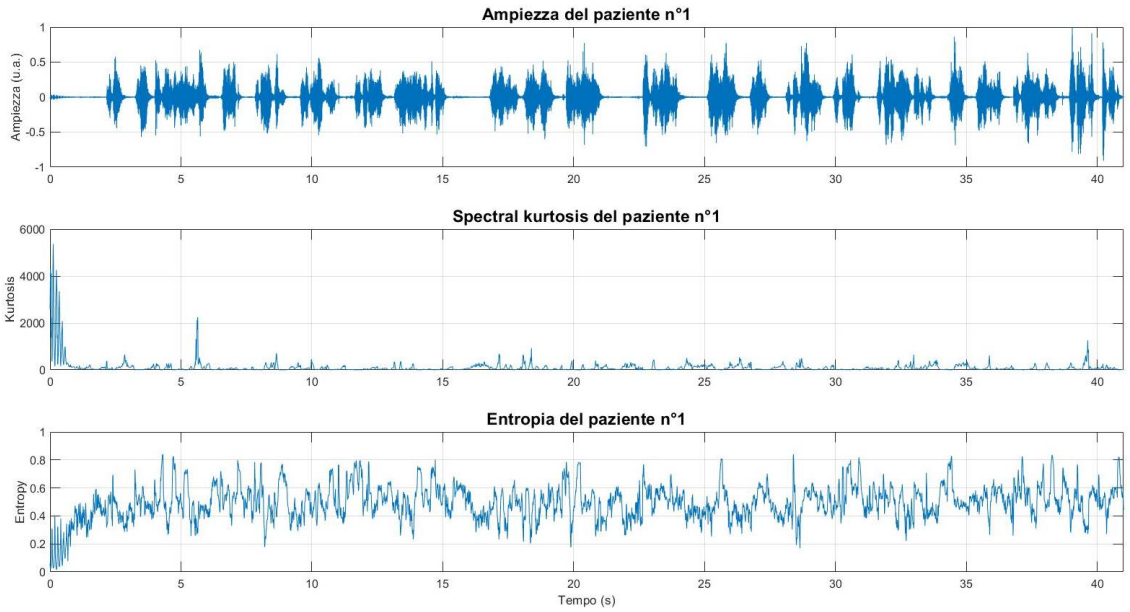

***Figura 5**: Andamento dell'ampiezza, della kurtosis spettrale e dell'entropia spettrale del paziente 1 del gruppo degli OPHL II.*

In seguito ad una meticolosa analisi di tali descrittori spettrali, è stata notata una peculiarità nel comportamento della kurtosis spettrale. Durante l'eloquio di un soggetto, al momento della generazione di un suono non armonico, il quale è ravvisato dalla pronuncia di consonanti non vocalizzate come /s/ o /t/, l'andamento dell'ampiezza del segnale tende ad assomigliare a quello di disturbo ed il relativo valore di kurtosis spettrale tende a diminuire repentinamente. Usufruendo del software Audacity e ascoltando di volta in volta i relativi campioni, è stato possibile localizzare temporalmente le sezioni di segnale che presentano consonanti non vocalizzate. Un esempio di quanto enunciato è stato riportato in figura 6.

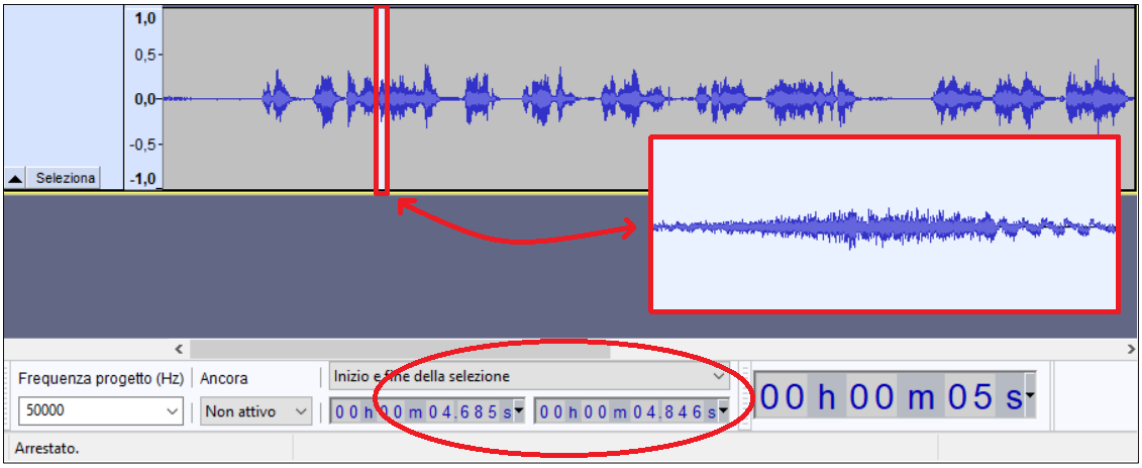

***Figura 6**: Localizzazione del frame corrispondente all'emissione della consonante /s/ tramite Audacity.*

Dalle informazioni di localizzazione estratte è stato possibile individuare i relativi frame anche all'interno dell'ambiente Matlab e studiarne i relativi andamenti, come evidenziato in figura 7. In tali finestre sono stati riscontrati evidenti cali del valore della kurtosis confermando una possibile correlazione tra il suo crollo repentino e la comparsa di frame inarmonici.

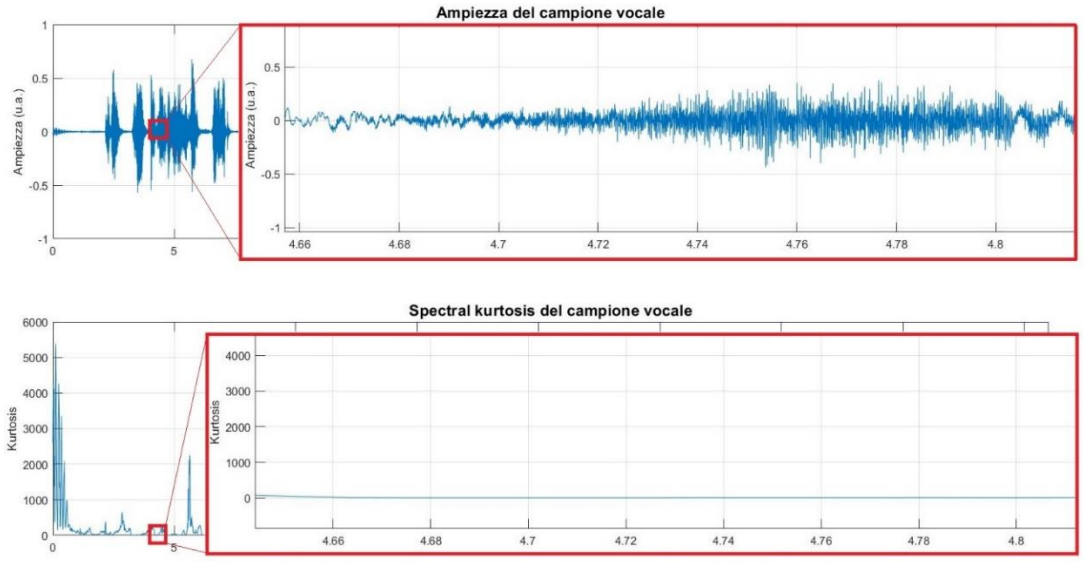

***Figura 7**: Evidenza degli andamenti di ampiezza e kurtosis spettrale del frame selezionato in Figura 6.*

Sulla base di questa intuizione è stato implementato un controllo della kurtosis spettrale sui diversi frame del segnale, in modo da scartare quelle finestre con un valore al di sotto di una certa soglia e, quindi, di catalogarle come inarmoniche. Un primo tentativo è stato portato avanti mediante una soglia arbitrariamente bassa, la quale è stata scelta in seguito ad un'evidenza visiva dei valori di kurtosis assunti nelle finestre di segnale corrispondenti all'emissioni di consonanti non vocalizzate. Successivamente, tramite l'ascolto diretto del vettore del segnale privato dei frame considerati non armonici, è stato possibile verificare che la procedura implementata riesce a soddisfare in buona misura la condizione di discriminazione ricercata.

Per decidere come impostare un valore di soglia rigorosa e che non faccia riferimento all'arbitrarietà di chi analizza i tracciati, è stata effettuata un'elaborazione sui soggetti sani. Questi sono stati trattati mediante la condizione di discriminazione dei frame basata sul valore di HNR, la quale è stata presentata precedentemente al paragrafo 2.3. Ogni qualvolta che tale metodo, perfettamente funzionante su soggetti privi di qualsiasi forma di disfonia, è stato in grado di identificare un frame come non armonico, ovvero quando questo risulta avere un valore di HNR al di sotto di zero dB, da esso è stato calcolato il relativo valore di kurtosis spettrale e memorizzato all'interno di un vettore. Alla conclusione di tutte le iterazioni, eseguite per ciascuno dei dieci soggetti sani presi in esame, sono stati ottenuti altrettanti vettori contenenti gli andamenti della kurtosis dei frame inarmonici. Per ciascun vettore sono state calcolate le statistiche descrittive di media, mediana, moda, range, 5°, 10° e 95° percentile che hanno riportato i risultati riassunti in tabella 6:

***Tabella 6**: Risultati delle distribuzioni statistiche della kurtosis spettrale sui soggetti sani.*

Soggetti	Media	Mediana	Moda	Range	5° $\%_{ile}$	10° $\%_{ile}$	95° $\%_{ile}$
1	217.968	6.427	1.730	$2.42 \cdot 10^3$	2.444	2.986	$1.38 \cdot 10^3$
2	202.085	53.095	1.366	$2.13 \cdot 10^3$	2.173	2.392	960.736
3	44.535	33.716	3.234	167.077	4.556	4.966	129.276
4	101.318	26.865	1.090	995.478	1.691	2.168	432.475
5	105.177	9.05	1.423	$1.58 \cdot 10^3$	1.945	2.363	510.302
6	66.822	27.709	1.318	355.081	1.922	2.645	239.382
7	493.460	297.861	2.124	$2.29 \cdot 10^3$	3.162	4.468	$1.59 \cdot 10^3$

8	771.124	591.709	4.384	$3.62 \cdot 10^3$	9.046	77.174	$2.32 \cdot 10^3$
9	938.655	747.182	1.998	$4.93 \cdot 10^3$	2.443	4.302	$2.76 \cdot 10^3$
10	282.643	198.389	1.571	$2.30 \cdot 10^3$	2.546	3.514	881.246

È stato possibile riscontrare forti fluttuazioni tra i diversi valori mostrati in Tabella 6. Ciò è stato principalmente spiegato a causa dei diversi trattamenti di compressione e modulazione eseguiti dalle diverse app di registrazione adoperate dai soggetti. Altre motivazioni potrebbero essere lo stato e le caratteristiche dei microfoni installati all'interno dei cellulari dei soggetti, come anche le condizioni ambientali al momento della registrazione che potrebbero aver minato parzialmente la qualità dei tracciati. Per sopperire a tale problema, è stato inizialmente ricavato l'errore standard (SE) per ogni statistica adoperata. Esso è calcolato come:

$$SE = \frac{2\sigma}{\sqrt{N}} \tag{2.18}$$

dove σ rappresenta la deviazione tipo di tutte le metriche statistiche menzionate ed N il numero totale dei soggetti sani. Successivamente, è stata ricavata la media di ciascuna statistica descrittiva lungo i dieci soggetti, la quale è stata poi rappresentata insieme al relativo valore di errore standard, come mostrato nella tabella 7 e in figura 8.

***Tabella 7**: Media e SE delle distribuzioni statistiche.*

Statistiche	Media	Errore Standard
Media	320	180
Mediana	200	150
Moda	2.02	0.59
Range	2078	826
5° %ile	3.2	1.3
10° %ile	11	13
95° %ile	1120	508

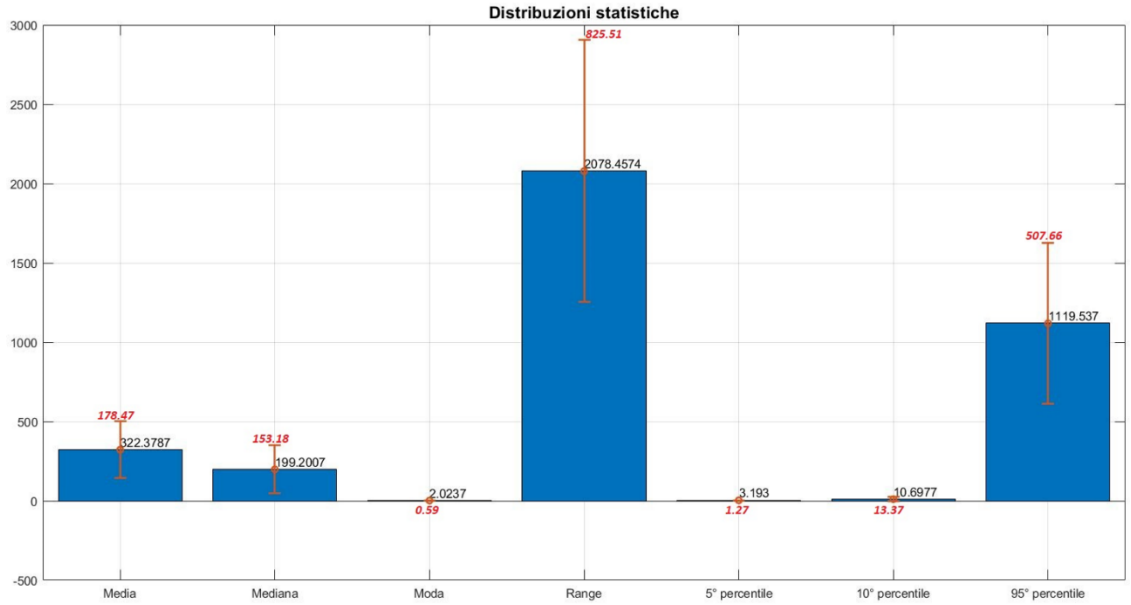

***Figura 8**: Rappresentazione della media delle statistiche descrittive (altezza delle barre blu) e dei relativi valori errori standard (valori in rosso).*

In ragione di quanto è stato espresso dal grafico di figura 8, la moda risulta il parametro con il valore di errore standard più contenuto, quindi, la sua media è stata scelta come valore soglia definitivo e rigoroso, detto anche soglia ***golden standard***, per la condizione di discriminazione dei frame.

2.4.2 Estrazione dei parametri

Una volta risolta la problematica riguardante il criterio di discriminazione dei frame, l'analisi è stata trasferita alla ricerca e all'estrazione di parametri acustici più adeguati al lavoro della tesi in esame.

In conseguenza di una minuziosa indagine bibliografica e un'attenta valutazione dei parametri già in possesso, sono state selezionate diciotto features come possibili buoni candidati da estrarre nelle sole finestre armoniche di segnale. Alcune di queste sono già state presentate in paragrafi precedenti, come l'indice SPI e gli MFCC. Alcune sono anche state adoperate nella prima versione dell'elaborato, come il CPPS. Altre sono state aggiunte in seguito ai risultati rinvenuti dalla ricerca bibliografica, come l'entropia

spettrale, la kurtosis spettrale e l'inclinazione spettrale (detta anche spectral tilt). Di seguito, ciascuna di esse è stata dettagliatamente definita e calcolata.

2.4.2.1 CPPS

Sinteticamente introdotto nella tabella 3 del paragrafo 1.3, è un parametro facente parte del dominio del cepstrum. Rappresenta una misura dell'ampiezza del picco cepstrale normalizzata su tutta l'ampiezza del segnale, ma tale misurazione avviene solo in seguito ad un processo generale di smoothing operato lungo tutto il cepstrum. È stato utilizzato nella prima versione di questo lavoro ed è stato riproposto anche nella versione definitiva visto la promettente qualità nell'accuratezza delle predizioni che dimostra in ogni studio in cui è stato impiegato.
L'algoritmo per il calcolo del CPPS è stato implementato in ambiente Matlab come una funzione separata dagli altri script, la quale prende in input il vettore del segnale vocale relativo al paziente laringectomizzato in esame e la relativa frequenza di campionamento per poi svolgere i seguenti passi:

- Il segnale vocale è stato sottocampionato di un fattore 2, con la frequenza di campionamento che risulta, quindi, pari a 22050 Hz. Il vettore ottenuto è stato sottoposto alla rimozione dell'offset e alla normalizzazione rispetto al suo massimo;
- È stato definito un range di frequenza da 80 Hz a 400 Hz, essendo questo l'intervallo in cui è statisticamente probabile che cada la frequenza fondamentale del soggetto in esame;
- È stata impostata una finestra di misurazione lunga 1024 campioni, la quale viene fatta scorrere lungo tutto il vettore del segnale con un salto di sovrapposizione pari a 44 campioni;
- Il segnale, scansionato frame per frame, è stato ripulito dalle finestre silenziose attraverso il metodo del controllo del valore efficace, descritto nel paragrafo 2.2. Successivamente è stato eseguito il controllo sull'armonicità del frame in esame mediante la valutazione della kurtosis spettrale, come mostrato precedentemente nel paragrafo 2.4.1;

- Per ogni frame armonico è stato calcolato il valore assoluto del prodotto della sua trasformata di Fourier con la finestra di Hanning lunga 1024 campioni (ossia è stato calcolato il primo spettro di potenza). Il risultato rinvenuto, indicato con *yfft*, è stato sottoposto al seguente calcolo

$$a)\ \ logyfft = 20 \cdot log_{10}(yfft)$$

$$b)\ \ yfft2 = \ 20 \cdot log_{10}\left|\frac{FFT(logyfft)}{512}\right| \tag{2.19}$$

- Una volta che tutto il segnale è stato esaminato e che i relativi dati del cepstrum sono stati memorizzati, si passa alla fase di smoothing. Essa è svolta, inizialmente, lungo il dominio del tempo prendendo una finestra di smoothing di lunghezza pari a sette frame. Successivamente, è eseguita anche nel dominio della quefrency mediante una finestra di ugual dimensione. In seguito a ciò, è stata costruita una retta di regressione considerando solo i cepstrum che presentano una quefrency maggiore di un millisecondo, poiché quelli con valore minore di tale soglia risultano essere meno influenzati dalla periodicità spettrale e maggiormente inficiati dall'inviluppo spettrale, il quale muta lentamente [4].
- Infine, dal cepstrum ottenuto all'uscita delle due fasi smoothing è misurata la prominenza come la lunghezza del vettore proiettato sulla retta di regressione e dal risultato ricavato sono stati estratti la posizione ed il valore delle prominenze con valore positivo, nei quali sono stati saturati eventuali andamenti presenti al di sopra dei 40 dB e al di sotto dei -40 dB.

In figura 9 è stato mostrato un esempio dell'andamento del cepstrum di un soggetto sano e di un soggetto disfonico all'uscita da entrambe le fasi di smoothing, così da enfatizzarne le relative differenze.

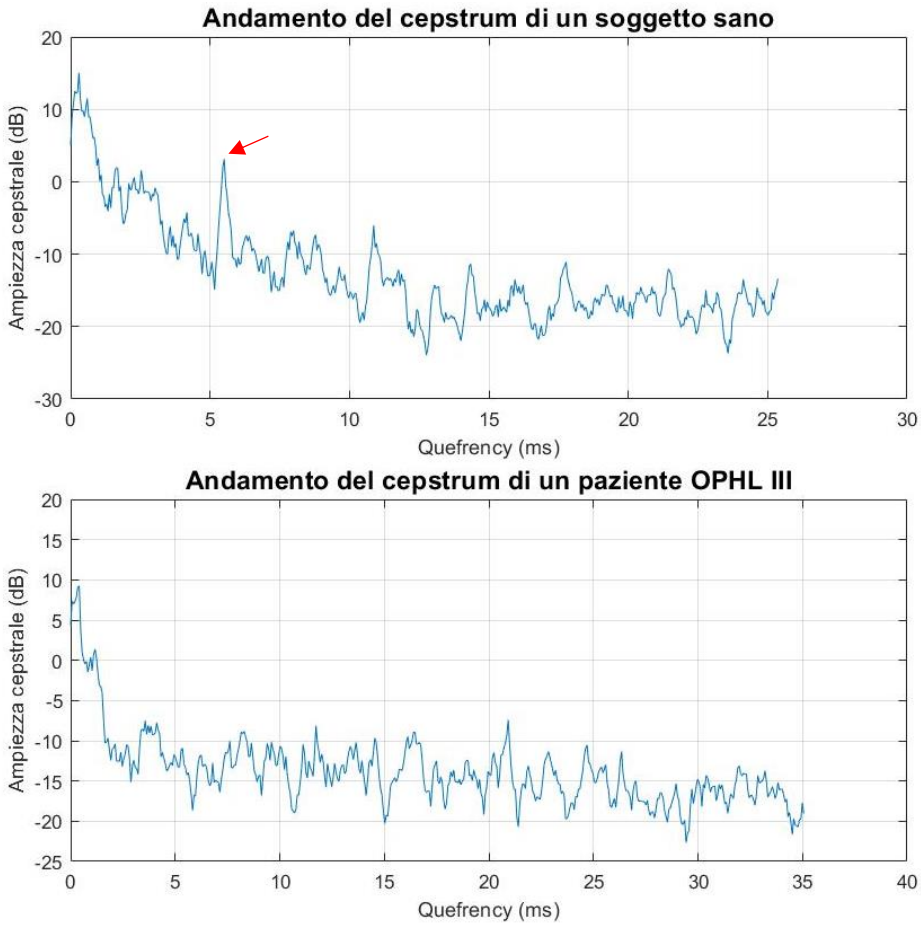

***Figura 9**: Andamento del cepstrum per un sano ed un paziente OPHL III, dove nel primo è chiaramente visibile il picco, mentre nel secondo risulta essere meno enfatizzato.*

Con la procedura sopra citata è stato ottenuto un vettore di CPPS per ognuno dei 32 pazienti disfonici in esame, ognuno di lunghezza diversa. Ciascuno di questi vettori è stato memorizzato all'interno di una matrice tridimensionale. Essa è stata introdotta per essere usata nella procedura dei confronti, la quale contiene nelle colonne tutte le diciotto features selezionate per la fase di estrazione di questo lavoro di tesi, nelle righe i valori degli andamenti di ciascuna di esse e nella terza dimensione i 32 pazienti. Infine, da ciascun vettore di CPPS sono state estratte le metriche statistiche di media, mediana, moda, deviazione tipo, range, 5° percentile, 95° percentile, skewness e kurtosis e questi nove valori ottenuti sono, poi, stati inseriti in una seconda matrice, chiamata *Data*, la quale è stata introdotta per essere adoperata come dataset per i processi di

classificazione. In figura 10 è stato mostrato il valore di ciascuna statistica descrittiva per ognuno dei 32 pazienti in esame.

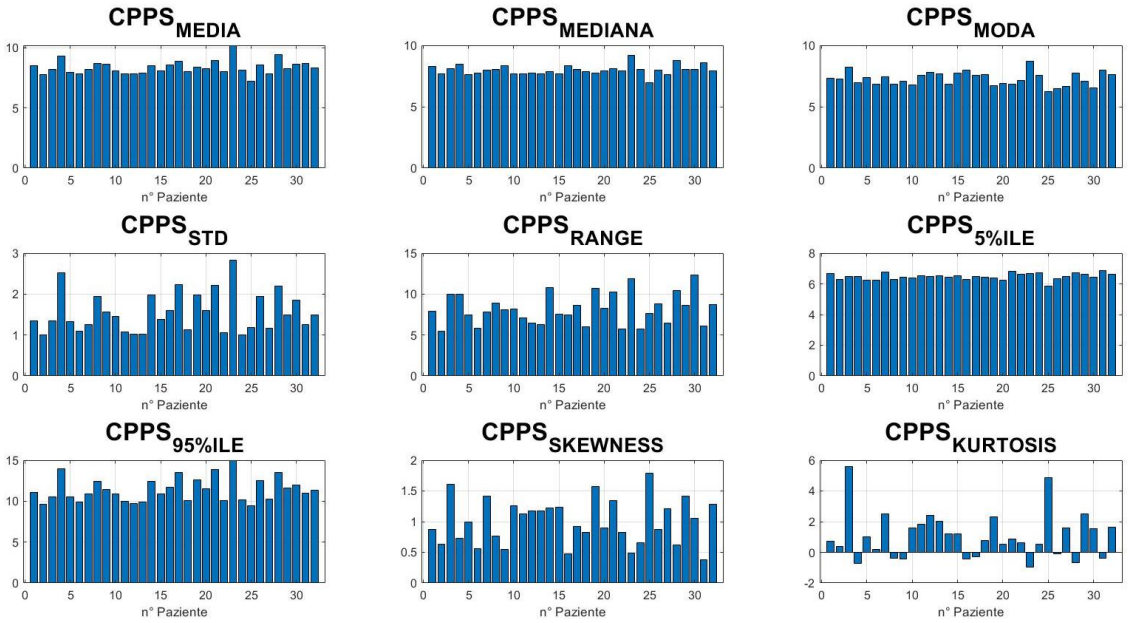

***Figura 10**: Statistiche descrittive del CPPS per i 32 pazienti.*

2.4.2.2 MFCC

Succintamente presentato nella tabella 3 del paragrafo concernente lo stato dell'arte, gli MFCC sono dei parametri estraibili dal cepstrum. È una tecnica basata sulle caratteristiche dell'orecchio umano e sulle differenze di frequenze che questo è in grado di distinguere [27]. I coefficienti MFCC sono ottenuti mediante la suddivisione dello spettro in un certo numero di frame, sui quali è eseguita la trasformata di Fourier (FFT). I coefficienti della trasformata sono filtrati mediante un banco di filtri conformi alla scala di Mel. Essi rappresentano un insieme di filtri passabanda triangolari parzialmente sovrapposti tra loro che moltiplicano le componenti spettrali, così che l'esito del filtraggio si approssimi a quello della scala Mel [27]. Infine, il risultato ottenuto all'uscita del banco di filtri è convertito in energia in scala logaritmica e poi fatto passare attraverso un blocco di trasformazione discreta del coseno [19]. Di seguito è stato

presentato il procedimento analitico che porta al calcolo dei coefficienti ed in figura 11 è stato mostrato lo schema a blocchi dello stesso:

- Indicando con *y(n)* il frame di segnale vocale, esso è convertito nel dominio della frequenza mediante una trasformata di Fourier discreta ad *M* punti e lo spettro di energia risultate è

$$|Y(k)|^2 = \left| \sum_{n=1}^{M} y(n) e^{\frac{(-j2\pi nk)}{M}} \right|^2 \qquad (2.20)$$

  Dove $1 \leq k \leq M$;

- Successivamente, lo spettro ottenuto è sottoposto al banco di filtri ed il risultato è uguale alla somma pesata tra le diverse risposte dei filtri, indicate con $\varphi_i(k)$, e l'energia spettrale $|Y(k)|^2$, ossia

$$e(i) = \sum_{k=1}^{M} |Y(k)|^2 \ \varphi_i(k) \qquad (2.21)$$

  dove con *e(i)* si indica il risultato ottenuto all'uscita del banco di filtri;

- Infine, la trasformata discreta del coseno è eseguita sul logaritmo del risultato precedentemente ottenuto {*log[e(i)]*} ed i coefficienti MFCC sono così restituiti all'uscita di tale operazione.

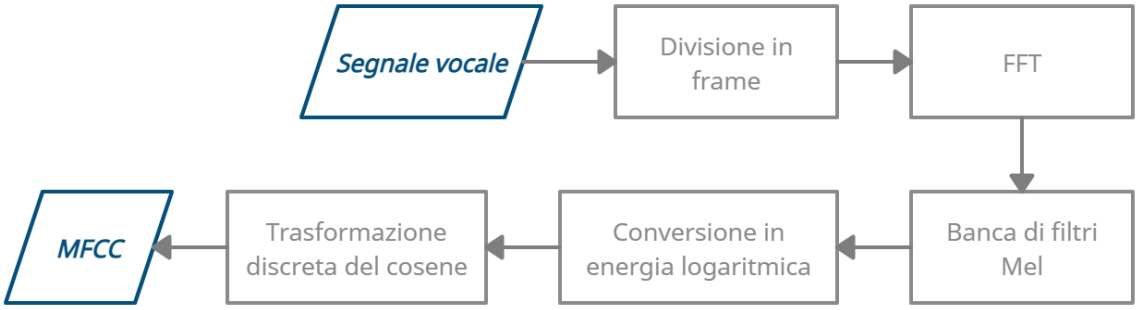

***Figura 11**: Schema a blocchi della procedura di estrazione degli MFCC.*

Gli MFCC sono features ampiamente usate nel riconoscimento e identificazione del parlato [19] e sono state riconosciute come adeguate allo scopo di questa tesi in seguito

ai promettenti risultati riscontrati in bibliografia.

Tale parametro risulta essere già implementato all'interno del pacchetto Matlab Audio Toolbox™, menzionato nel paragrafo 2.4.1 in merito alla condizione di valutazione dei frame, quindi, nel seguito è stata mostra la procedura attuata a partire da tale funzione. La sua scrittura generale è:

Coeffs = mfcc(x, fs);

Dove *x* rappresenta l'intero segnale vocale, *fs* è la relativa frequenza di campionamento e *Coeffs* è una matrice di N righe, pari al numero di finestre di segnale esaminate, ed M colonne, pari al numero di coefficienti MFCC estratti. Quest'ultimo valore risulta modificabile ed impostato per il lavoro in esame a 13. Il banco di filtri Mel è stato adoperato in configurazione standard, di conseguenza ha i relativi filtri triangolari impostati con una sovrapposizione pari alla metà della lunghezza della banda, così che i bordi di ciascuna di queste (tranne per la prima e per l'ultima) corrispondano alle frequenze centrali dei filtri adiacenti. Inoltre, sempre per impostazioni di default, tale banco risulta costituita da 40 filtri triangolari, dove i primi dieci sono spaziati linearmente e i restanti logaritmicamente, e tutti insieme coprono un campo di frequenza compreso tra 133 Hz e 6864 Hz, così come mostrato in figura 12.

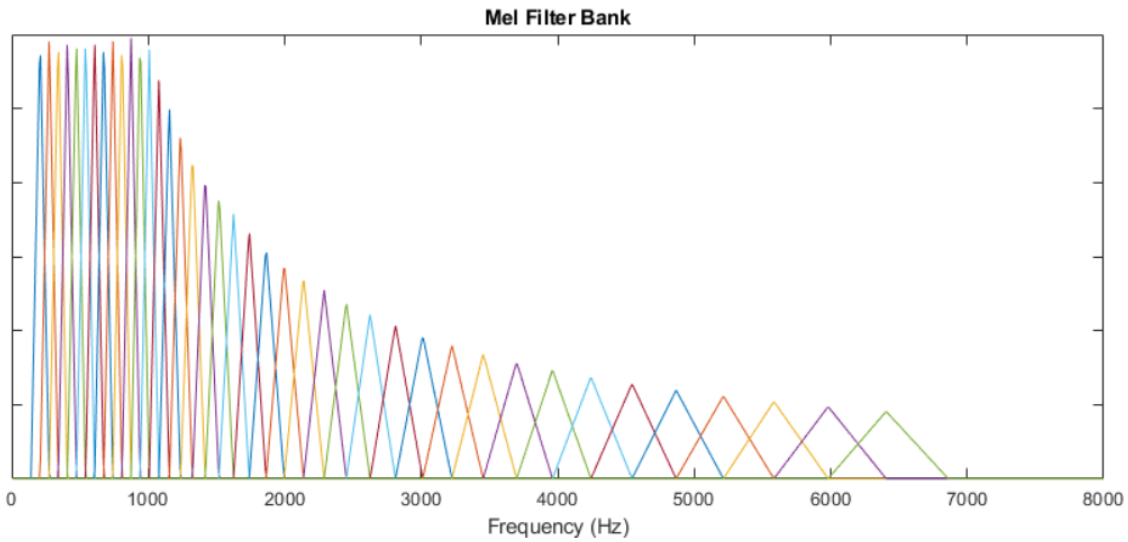

***Figura 12**: Rappresentazione del banco di filtri Mel in configurazione standard.*

Nell'elaborazione di questa tesi, tuttavia, la funzione è stata eseguita solo sui singoli frame riconosciuti come armonici, quindi, ad essa sono state apportate alcune modifiche. In particolare, sono stati dati come input: la finestra armonica in esame, la frequenza di campionamento (rimasta invariata dopo il pre-processing a 44100 Hz) ed una finestra di Hanning lunga quanto il frame di segnale usato, in modo da calcolare la trasformata di Fourier. L'imposizione del tipo e della lunghezza di finestra è stata indotta allo scopo di ottenere un singolo vettore di 13 coefficienti per ogni frame armonico, il quale è stato memorizzato insieme a tutti gli altri vettori estratti dalle relative finestre, così da ricavare l'andamento temporale degli MFCC per il soggetto in questione. Tutti gli andamenti di ciascun paziente sono stati memorizzati all'interno della matrice tridimensionale, citata al paragrafo 2.4.2.1, e per ognuno di essi sono state ricavate le nove metriche statistiche, poi inserite nel corrispettivo dataset chiamato *Data* insieme alle altre già presenti. In figura 13 sono stati mostrati i valori delle nove metriche statistiche di un coefficiente MFCC, per ciascuno dei 32 laringectomizzati.

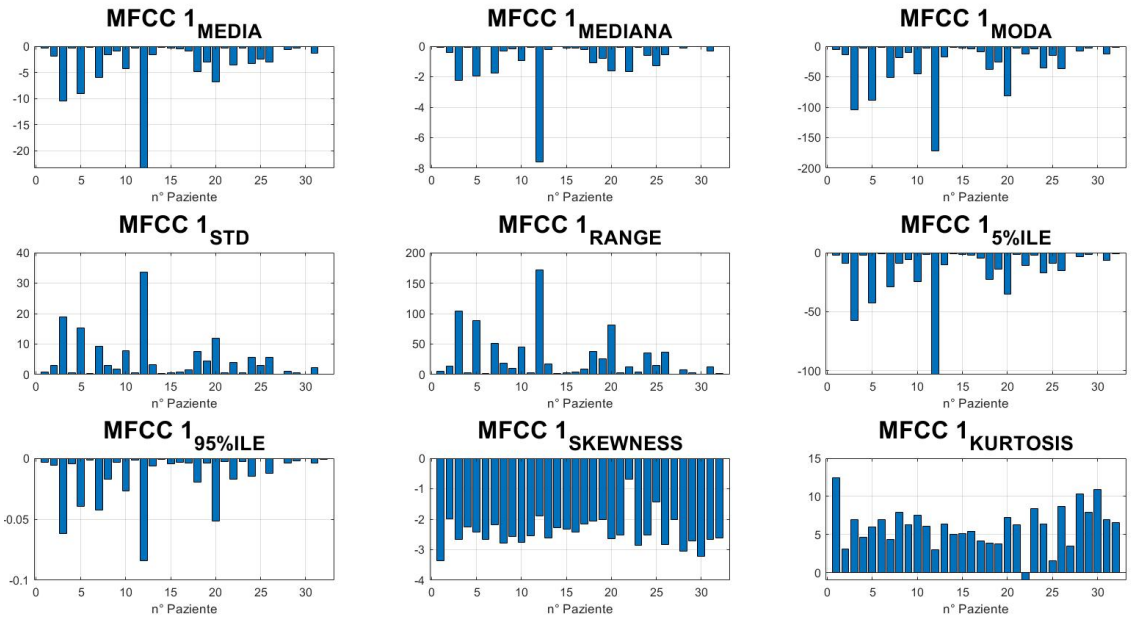

***Figura 13**: Statistiche descrittive di MFCC1 per i 32 pazienti.*

2.4.2.3 Spectral kurtosis e Spectral entropy

Precedentemente presentati nel capitolo 2.4.1, la *kurtosis spettrale* è una misura di quanto lo spettro del segnale si avvicini all'andamento di una gaussiana ed è ottenuta

mediante l'equazione 2.16. L'*entropia spettrale*, invece, è la misura della prominenza armonica dello spettro del segnale ed è ricavata tramite la formula 2.17.
Entrambe, come espresso nel capitolo 2.4.1, fanno parte del pacchetto Matlab Audio Toolbox™, quindi, le relative implementazioni sono già a disposizione.

kurtosis = spectralKurtosis(x,f) *entropy = spectralEntropy(x,f)*

Quelle sopra scritte sono le diciture generali delle rispettive funzioni, dove con *x* si indica il campione vocale e con *f* la relativa frequenza di campionamento in Hz.
Nel caso specifico del lavoro di tesi, l'elaborazione è stata condotta in modo da ottenere, per ogni frame armonico, un singolo valore rappresentante la kurtosis spettrale ed un singolo valore rappresentate l'entropia spettrale. In ragion di ciò, le funzioni prima citate sono state leggermente modificate in modo da ricevere in input: il frame armonico sotto esame, la frequenza di campionamento (44100 Hz) ed una finestra di Hanning lunga quanto il frame. In questo modo, sono stati ottenuti i singoli valori di kurtosis spettrale ed entropia spettrale per ciascun frame, i quali sono stati memorizzati all'interno di due distinti vettori così da ottenere, alla conclusione delle iterazioni, due vettori rappresentanti i rispettivi andamenti temporali. Sono stati ricavati i due vettori per ciascun paziente ed ognuno di questi è stato memorizzato nella matrice tridimensionale. Infine, da tutte le evoluzioni temporali sono state ricavate le nove metriche statistiche e inserite nella matrice *Data*. In figura 14 e 15 sono state raffigurate le statistiche descrittive, rispettivamente, della kurtosis spettrale e dell'entropia spettrale di ciascun paziente disfonico preso in esame per questo lavoro di tesi.

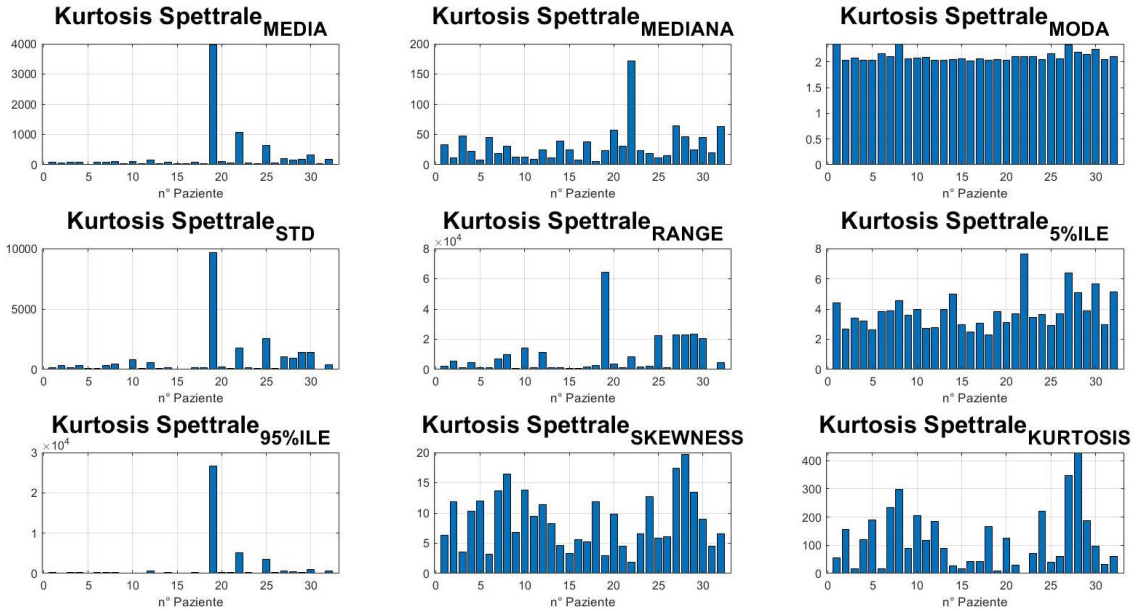

***Figura 14**: Statistiche descrittive della kurtosis spettrale per i 32 pazienti.*

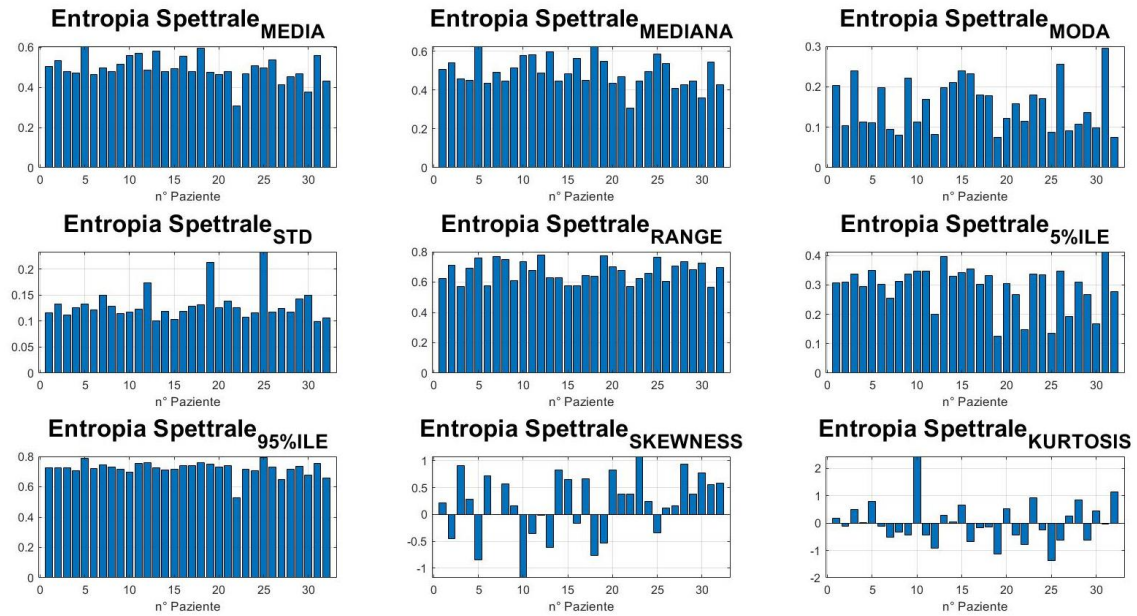

***Figura 15**: Statistiche descrittive dell'entropia spettrale per i 32 pazienti.*

2.4.2.4 Logaritmic Spectral Tilt e SPI

L'inclinazione spettrale, detta comunemente *spectral tilt* (o anche spectral slope), rappresenta la velocità di riduzione dell'ampiezza delle armoniche nel tracciato spettrale [29]. Questo parametro è calcolato mediante il confronto tra l'energia spettrale a bassa frequenza (tra 0 e 1 kHz) e l'energia spettrale ad altra frequenza (tra 1 kHz e 5 kHz). Essa è un parametro già presente in alcuni studi bibliografici, ma, per ammissione degli stessi, i risultati che lo riguardano sono poco significativi. Questo è principalmente dovuto alle esigue ricerche che lo hanno preso in considerazione e, conseguentemente, alla modesta quantità di dati da cui è stato estratto, rendendolo poco consistente per le generali applicazioni dell'ambito sotto esame [13]. Nonostante ciò, è stato ugualmente preso in considerazione per il lavoro di questa tesi in ragione delle caratteristiche che dimostra. La sua estrazione è stata eseguita nei soli frame armonici, dai quali è stato calcolato il modulo della relativa trasformata di Fourier mediante un algoritmo di trasformata di Fourier (FFT). Successivamente, da ciascuno di questi sono state ottenute le energie spettrali a bassa frequenza (E1), ovvero i valori efficaci delle trasformate di Fourier valutate in corrispondenza del range a basse frequenze (da 0 a 1kHz), e le energie spettrali ad alta frequenza (E2), quindi, i valori efficaci delle trasformate valutate in corrispondenza del relativo range (da 1 kHz a 5 kHz). Infine, è stato calcolato

$$StdB = 20 \cdot log_{10} \frac{E1}{E2} \tag{2.22}$$

in modo da ottenere l'inclinazione spettrale logaritmica (*logaritmic spectral tilt*, *StdB*). Sono stati memorizzati all'interno della matrice tridimensionale gli andamenti temporali di ogni inclinazione spettrale ottenuta per ciascuno dei 32 pazienti. Da queste evoluzioni sono poi state estratte le nove metriche statistiche e sono state inserite all'interno della matrice *Data*. In figura 16 è stato mostrato il valore di ogni statistica descrittiva ricavata da ciascuno dei 32 pazienti disfonici in esame.

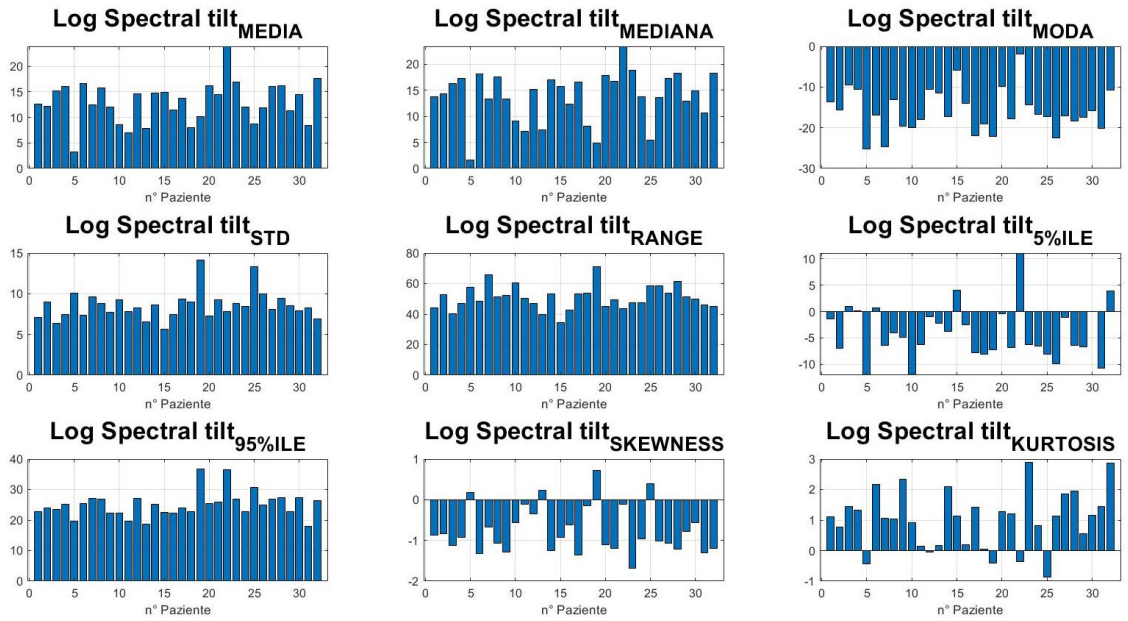

***Figura 16**: Statistiche descrittive dell'inclinazione spettrale logaritmica per i 32 pazienti.*

L'indice di fonazione morbida (*soft phonation index*, *SPI*), è già stato presentato sommariamente nella tabella 1 del paragrafo 1.3. Esso rappresenta una misura della struttura armonica dello spettro [9] ed è calcolato come il rapporto medio tra l'energia armonica a bassa frequenza (tra 70 Hz e 160 Hz) con l'energia armonica ad alta frequenza (tra 160 Hz e 4.5 kHz). Dagli studi presenti in letteratura risulta essere fortemente correlato con lo stato di disfonia del soggetto. Infatti, ad esempio, un aumento del suo valore può essere indice di un'adduzione incompleta o totalmente persa delle corde vocali [9]; risulta essere un parametro significativo per le voci maschili rauche, ma non per quelle femminili, ed un ottimo indicatore della presenza di respiri nella voce [30]. Analogamente al logaritmic spectral tilt, tale indice è stato estratto solamente dai frame armonici, da cui è stato calcolato il modulo della relativa trasformata di Fourier e, successivamente, è stata estratta l'energia armonica a bassa frequenza (ovvero il valore efficace della trasformata di Fourier valutata nel campo di frequenze [70-160] Hz) e l'energia armonica ad alta frequenza (ossia il valore efficace della trasformata valutata nel range [160-4500] Hz). Infine, è stato calcolato il logaritmo in base dieci del rapporto tra le due energie. Ciò è stato eseguito e memorizzato per ogni frame armonico di ciascuno dei 32 pazienti, così da assemblare i 32 vettori rappresentanti i rispettivi andamenti temporali, i quali sono poi stati archiviati nella

matrice tridimensionale. Infine, come da prassi, da tali evoluzioni temporali sono state estratte le nove metriche statistiche, sono state inserite nella matrice *Data* e ne sono stati mostrati i relativi valori, per ciascun paziente laringectomizzato, in figura 17.

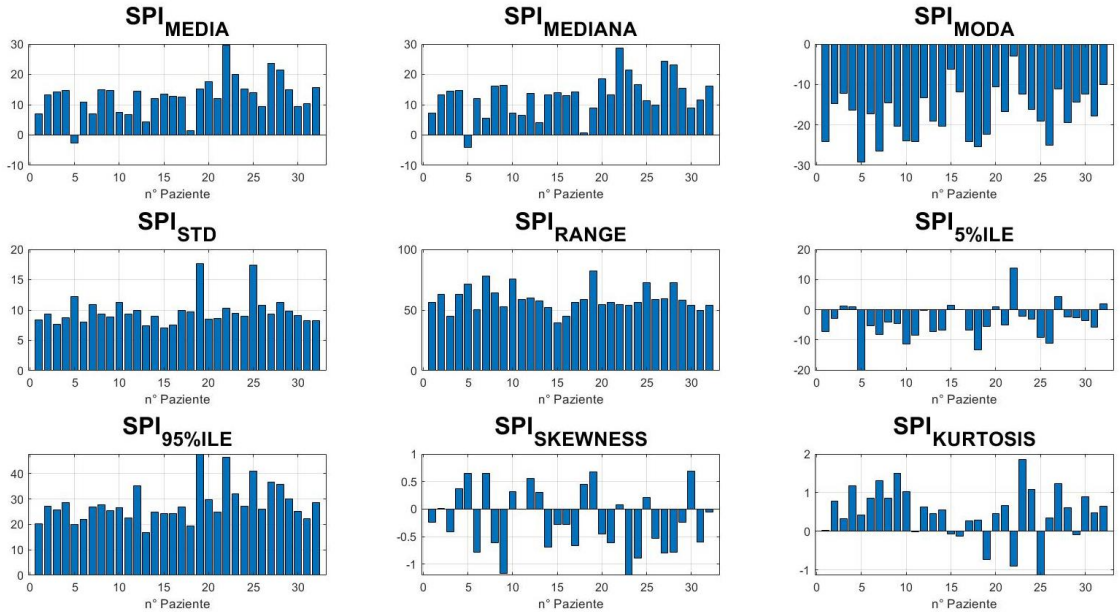

***Figura 17**: Statistiche descrittive di SPI per i 32 pazienti.*

2.4.3 Confronti degli andamenti temporali

Il passo successivo all'estrazione dei parametri è stato quello di identificare un'eventuale capacità di questi di essere, a loro modo, grandezze di misura in grado di riconoscere i pazienti aventi una buona qualità vocale oppure quelli con qualità vocale peggiore. Tali abilità sono state verificate mediante l'esecuzione di confronti tra i singoli andamenti dei parametri, memorizzati nella matrice tridimensionale più volte citata e riconducibili a pazienti con caratteristiche vocali diverse. Per effettuare la procedura, tutti i soggetti disfonici in esame sono stati prima suddivisi in due gruppi a seconda della loro qualità di esecuzione vocale. Ciò è stato ottenuto tramite l'utilizzo dei valori del parametro I (intellegibilità) della scala INFVo, riportati nelle tabelle 4 e 5 e disponibili per tutti i pazienti, in cui è stata fissata una soglia a metà del range di possibili valori, quindi scelta pari a 5, per identificare le due diverse classi di qualità vocale. Così facendo,

i pazienti, i relativi segnali vocali e gli andamenti dei parametri da essi estratti sono stati considerati di buona qualità ed inseriti nella relativa classe 0, se essi posseggono valori di intellegibilità sotto la soglia impostata. Al contrario, i pazienti con un valore di I maggiore di 5 sono stati considerati di scarsa qualità vocale ed inseriti nella classe 1. A causa di uno sbilanciamento nel numero di pazienti riscontrato nel primo dei due gruppi, in cui sono stati collocati 21 dei 32 a disposizione, è stato deciso di ridimensionarlo scartandone alcuni, così da trattare le operazioni di confronto con un dataset bilanciato. In concreto, è stato reclutato un numero di pazienti della classe 0 pari a quelli presenti nella classe 1, i quali risultano pari a 11, con un criterio di cernita basato sulla selezione dei disfonici con la migliore qualità di voce, ovvero sono stati presi gli undici soggetti la cui media tra i rispettivi I risulta essere il più vicino possibile al valore zero della scala INFVo. In tabella 8 è stato presentato il risultato a seguito della cernita ed è stato possibile notare una maggior presenza di pazienti OPHL II, rispetto a quelli OPHL III, nella classe 0.

***Tabella 8**: Dataset ridotto per le operazioni di confronto.*

Classe 0 sottosoglia (<5)		**Classe 1 sopra soglia (>5)**	
Pazienti	**I (INFVo)**	**Pazienti**	**I (INFVo)**
pz.3 OPHL II	0.9	*pz.1 OPHL II*	7.6
pz.4 OPHL II	1.3	*pz.11 OPHL II*	8.3
pz.6 OPHL II	1.3	*pz.12 OPHL II*	6.9
pz.8 OPHL II	0.8	*pz.18 OPHL II*	6.3
pz.10 OPHL II	2.2	*pz.19 OPHL II*	8.8
pz.14 OPHL II	0.8	*pz.1 OPHL III*	6.2
pz.17 OPHL II	0.5	*pz.4 OPHL III*	6.7
pz.20 OPHL II	0	*pz.6 OPHL III*	6.8
pz.21 OPHL II	1.2	*pz.7 OPHL III*	9.7
pz.22 OPHL II	1.5	*pz.8 OPHL III*	5.9
pz.2 OPHL III	1.8	*pz.9 OPHL III*	6.6

Una volta ottenute le due classi, sono stati eseguiti i confronti degli andamenti dei parametri mediante la funzione *kstest2* di Matlab, la quale implementa il test di Kolmogorv-Smirnov a due campioni. Esso è un test d'ipotesi non parametrico ampiamente usato per il confronto tra due campioni, verificando l'appartenenza di questi alla stessa popolazione [31]. In particolare, valuta la differenza tra le funzioni cumulative delle distribuzioni dei due campioni su di un determinato range *x* per ogni dataset a disposizione.
Il test può essere calcolato in due modalità:

- Il *test bilaterale*, il quale usa la massima differenza assoluta tra le funzioni cumulative delle distribuzioni dei campioni

$$D^* = max\left|\widehat{F_1}(x) - \widehat{F_2}(x)\right| \tag{2.23}$$

  Dove $F_1(x)$ è la proporzione dei valori x_1 minori o uguali a x ed $F_2(x)$ è la proporzione dei valori x_2 minori o uguali a x

- Il *test unilaterale*, il quale usa il valore effettivo della differenza tra le funzioni di ripartizione dei due campioni anziché il valore assoluto

$$D^* = max\left(\widehat{F_1}(x) - \widehat{F_2}(x)\right) \tag{2.24}$$

La funzione Matlab, in generale, è scritta come *h = kstest2(x1, x2)* e restituisce una decisione per il test di ipotesi nulla nel caso in cui i vettori *x1* e *x2* appartengono alla stessa distribuzione continua con un livello di fiducia del 5% (ossia il p-value è minore di 0.05). In tal caso, si dice che è rigettata l'ipotesi nulla ed *h* risulta uguale ad 1. Se i due vettori, invece, appartengono a distribuzioni differenti, allora il test d'ipotesi non è rigettato ed *h* risulta uguale a zero. Alla luce di questo, l'algoritmo è stato implementato facendo sì che esegua i confronti di ciascun parametro tra i diversi soggetti appartenenti alla stessa classe, sia per il gruppo sottosoglia che per quello sopra soglia, ed i confronti di ciascun parametro tra i soggetti delle due classi differenti. Sono state contate tutte le volte che il test ha riscontrato un'affinità per entrambi i confronti intraclasse e, in contemporanea, sono state contate tutte le non affinità ottenute dai confronti

interclasse. Il risultato di tale procedimento è stato presentato e ampiamente commentato nel capitolo 3, relativo all'esposizione di tutti i risultati ottenuti da questo lavoro di tesi, al paragrafo 3.1.

2.4.4 Metodi di classificazione

Per validare definitivamente la fondatezza dei buoni risultati ottenuti al passo precedente, sono stati condotti due processi di classificazione. Il primo mediante l'implementazione di un algoritmo di *regressione logistica* ed il secondo tramite il modello di *Coarse Decision Tree* messo a disposizione dall'applicazione "Classification Learner" dell'ambiente Matlab.

La *regressione logistica* è un modello utilizzato nello studio di relazioni causali tra una variabile dipendente dicotomica (ossia che assume valori binari) e una o più variabili indipendenti quantitative. Essa non analizza le probabilità degli eventi di per sé, ma studia la trasformazione di queste in logaritmo naturale [32]. Infatti, la struttura di tale modello può essere così descritta: indicando con Y_i la variabile dipendente all'iterazione riferita al generico dato *i*, essa può assumere valore 0 se l'evento in questione è assente oppure 1 se è presente. Da ciò, si evince come le probabilità che Y_i assuma, rispettivamente, valore 1 e 0 su di un insieme di *n* variabili indipendenti $\{X_{1i}, X_{2i}, \ldots, X_{ni}\}$ è

$$P_r[Y_i = 1 \quad | X_{1i}, X_{2i}, \ldots, X_{ni}] = \pi_i \tag{2.25a}$$

$$P_r[Y_i = 0 \quad | X_{1i}, X_{2i}, \ldots X_{ni}] = 1 - \pi_i \tag{2.25b}$$

e la regressione logistica ha la seguente forma

$$\ln\left(\frac{\pi_i}{1 - \pi_i}\right) = \beta_0 + \beta_1 X_{1i} + \ldots + \beta_n X_{ni} \tag{2.26}$$

in cui β_0 è detto intercetta, la quale rappresenta il valore previsto di P_r quando il generico X è nullo, mentre β_1, β_2, …, β_n sono i coefficienti di regressione parziale [33]. In questo lavoro di tesi, tale regressione è stata implementata mediante la funzione Matlab *fitglm*, la quale crea un modello di regressione lineare e con opportune modifiche è stato

trasformato in un modello di regressione binomiale. La scrittura di quanto enunciato risulta:

mdl=fitglm (PAR, sp, 'Distributional', 'binomial', 'Link', 'logit');

dove *PAR* rappresenta la matrice dei dati da classificare, ossia il dataset *Data* menzionato nel paragrafo 2.4.2, e *sp* il vettore che contiene tutte le variabili di risposta dicotomiche, mentre '*Distributional*' e '*Link*' sono due opzioni aggiunte al fine di adattare la funzione alle esigenze richieste. Esse sono state configurate, rispettivamente, come '*binomial*' e '*logit*', in cui: la prima definisce che la distribuzione della variabile di risposta è una distribuzione binomiale; la seconda configura la funzione della regressione come in formula 2.25, così da impostare il modello di regressione logistica. Le iterazioni sono state svolte in modo che i parametri siano combinati in piccoli gruppi definiti, con una correlazione massima e minima tra loro rispettivamente di 0.9 e 0 e, infine, che sia presente una condizione sul massimo valore di p-value dei coefficienti del modello della regressione pari a 0.05 (ovvero che le predizioni sono state ottenute con un'attendibilità del 95%). In sostanza, è stato messo a punto il seguente procedimento:

a) È stato scelto il numero *p* desiderato di feature da combinare, quindi, è stato determinato il numero di possibili combinazioni *K* non ripetute, avendo a disposizione *N* parametri, nel seguente modo

$$K = \frac{N!}{N! \cdot (N-p)!} \tag{2.27}$$

che rappresenta la formula del coefficiente binomiale. Essendo il numero totale di parametri pari a 162 e scegliendo p=2, allora il numero di combinazioni K è uguale a 13041. Con p=3, K=695520. Con p=4, K=27646920;

b) La generica combinazione è stata analizzata in modo che rispetti la condizione sulla correlazione massima e minima. Se ciò è verificato, allora la combinazione in esame è abilitata al prosieguo della procedura, altrimenti è scartata;

c) È eseguita la regressione logistica sulla combinazione in questione mediante la funzione *fitglm* poco prima presentata;

d) Dal modello ottenuto sono valutati i coefficienti in modo che rispettino la condizione sul massimo p-value. Se questa è verificata, allora il modello è considerato valido ed è poi eseguita la fase di predizione. Se la condizione non è appurata, il modello è scartato e l'iterazione è riportata al passo b) andando a considerare la successiva combinazione di parametri.

Il passo di predizione è stato svolto mediante la funzione Matlab chiamata *predict*:

prob=predict (mdl, PAR, 'Alpha', 0.05);

Essa prende in input: *mdl,* ossia il modello precedentemente allenato dalla regressione logistica; *PAR*, la matrice contenente i dati da classificare che corrisponde a *Data*; l'opzione '*Alpha'*, con la quale è impostata la condizione sul massimo valore di p-value. Da tale funzione è ricavato un vettore '*prob*' contenente i risultati predetti dalla classificazione, il quale è messo a confronto con il vettore contenente le variabili di risposta (*sp*) per dimostrare l'efficacia di quanto ottenuto. La valutazione è stata effettuata attraverso la costruzione della *confusion matrix*, di cui ne viene riportato un esempio in tabella 9. Questa rappresenta lo strumento adatto per visualizzare le prestazioni del classificatore ed è, appunto, una matrice nelle cui righe sono collocati i valori predetti dalla classificazione, mentre nelle colonne sono disposti i veri valori assunti dai dati nel dataset. Le singole celle all'interno della matrice rappresentano le 4 possibili situazioni verificatesi durante una generica classificazione e sono indicate da queste quattro diciture:

- *TP, true positive*, indica l'ammontare delle variabili classificate correttamente nella prima classe, definita come "positiva" e che nel caso di questa tesi corrisponde alla classe 0;
- *TN, true negative*, indica quante variabili sono state classificate correttamente nella seconda classe, definita come "negativa" che corrisponde alla classe 1;
- *FP, false positive*, indica quante variabili sono state classificate come positive quando invece risultano essere "negative";

- *FN, false negative*, indica l'ammontare delle variabili classificate erroneamente come "negative", quando il loro valore ricade invece nella classe "positiva".

***Tabella 9**: Esempio di una confusion matrix.*

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

Inoltre, dai parametri appena definiti è possibile ricavare altre misure di prestazione del classificatore, ovvero:

- *Accuratezza* (%), ottenuta come il rapporto tra la totalità dei corretti classificati ed il numero totale di predizioni

$$Accuratezza = \frac{TP + TN}{TP + TN + FP + FN}; \qquad (2.28)$$

- *Sensitività* (detta anche true positive rate, %), è ottenuta dal rapporto della quantità di predizioni corrette positive con il numero totale di positivi all'interno del dataset

$$Sensitività = \frac{TP}{TP + FN}; \qquad (2.29)$$

- *Specificità* (detta anche true negative rate, %), è ottenuta dal rapporto tra il numero delle predizioni corrette negative ed il numero totale di dati negativi nel dataset

$$Specificità = \frac{TN}{TN + FP}; \tag{2.30}$$

- *Precisione* (%), ottenuta come il rapporto tra il numero di corretti positivi e la totalità delle predizioni positive

$$Precisione = \frac{TP}{TP + FP} \tag{2.31}$$

In una prima implementazione, è stata eseguita una classificazione prendendo tutto il dataset a disposizione sia per il training che per la validazione. Da questa procedura è stata ottenuta la migliore combinazione di parametri, successivamente usata per svolgere la fondamentale fase di validazione del modello. Tutti i vari risultati ottenuti sono stati presentati e interamente commentati all'interno del capitolo 3, al paragrafo 3.2.

L'altro modello di classificazione adottato è stato il *Coarse Decision Tree*, il quale appartiene alla classe degli algoritmi definiti *Decision Tree*. Questi fanno parte di una categoria più ampia che rappresenta una delle più importanti forme di *Machine Learning*, ossia le tecniche di apprendimento supervisionato, le quali devono il loro nome al fatto che il processo di apprendimento è svolto con le variabili di risposta completamente note. In particolare, gli algoritmi di *Decision Tree* sono dei modelli predittivi che utilizzano la rappresentazione ad albero, da qui il loro nome, per prevedere la risposta della classificazione. Tale processo di predizione parte dal nodo iniziale, detto anche nodo radice, segue i diversi cammini in base alla situazione in esame e, infine, raggiunge uno dei possibili nodi finali, detti anche nodi foglia, i quali contengono le variabili di risposta. La figura 18 ne mostra un esempio.

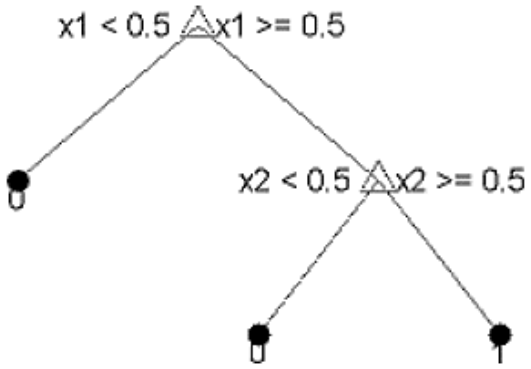

***Figura 18**: Esempio del grafico di un modello di Decsion Tree.*

Prendendo spunto dall'esempio sopra riportato, la scelta della prosecuzione in un cammino rispetto ad un altro dipende dal valore assunto dalle due variabili mostrate, così facendo si arriva alla predizione della variabile di risposta (nodo foglia), la quale è indicata dal punto nero in figura. Nello specifico, il *Coarse Decision Tree* è un tipo di modello decisionale piuttosto semplice e poco flessibile alle modifiche. Consta di pochi nodi foglia che gli permettono solo distinzioni grossolane tra le classi in esame. Ciò è principalmente dovuto al basso numero di *branch point* (ossia i punti di divisione dell'albero) che ne limitano la profondità della struttura e, di pari passo, l'accuratezza ottenuta durante la fase di allenamento del modello. Tuttavia, al contempo, tale modello mostra una più semplice interpretazione dei risultati ed una maggiore robustezza rispetto agli altri modelli.
Nell'elaborazione di questa tesi è stato scelto di affidare l'implementazione di tale modello all'applicazione insita all'ambiente Matlab chiamata "Classification Learner", la quale svolge tutte le procedure di preparazione dei dati e del relativo modello. La minima configurazione manuale necessaria è stata svolta con i seguenti passi:

- Il dataset necessario all'esecuzione del modello è stato definito partendo dalla matrice *Data*, a cui è stata concatenata una colonna rappresentante il vettore delle variabili di risposta;
- Questo vettore è stato costituito avvalendosi del parametro I della scala INFVo fornito per ogni paziente. In particolare, a ciascuno è stato assegnato valore 1 se il rispettivo parametro di intellegibilità risulta essere superiore a 5, ossia la soglia precedentemente scelta per distinguere le due classi di segnali vocali (paragrafo

2.4.3). Viceversa, è stato assegnato valore 0 se risultano avere il parametro I minore della soglia indicata;

- Una volta che tale vettore è stato ottenuto ed inserito nella matrice delle metriche statistiche, è stato possibile iniziare ad operare con il "Classification Learner". All'interno dell'ambiente di tale applicazione è stato necessario caricare il dataset creato ed indicare quale delle sue colonne contenga il vettore delle variabili di risposta. In una prima implementazione, è stato scelto di adoperare tutto il dataset a disposizione sia per il training che per la validazione ed è stato scelto di suddividere l'albero decisionale mediante il criterio dell'indice di diversità di Gini. Esso, come tutti i criteri di suddivisione, si basa sull'errore associato ai nodi e, in particolare, come riportato nella libreria "Help" di Matlab, è calcolato nel seguente modo

$$1 - \sum_{i} p^2(i) \tag{2.32}$$

dove la sommatoria è eseguita al nodo in esame su tutta la classe i-esima, con $p^2(i)$ che rappresenta la frazione osservata della classe i all'interno di tutte le classi considerate. In sostanza, l'indice di Gini ha valore 0 in corrispondenza dei nodi contenenti una sola classe (detti nodi puri), mentre in tutti le altre ha valore positivo restituendo, così, una misura dell'impurità dei nodi. Concluse queste brevi operazioni di configurazione, è stato scelto il tipo di modello di predizione desiderato (*Coarse Decision Tree*) e sono stati immediatamente ricavati i risultati della predizione.

Anche in questo caso, come in quello della regressione logistica, è stata svolta inizialmente una procedura di classificazione con il dataset completo e, successivamente, dai risultati ricavati è stata svolta la fase di validazione del modello. Inoltre, il "Classification Learner" è stato anche in grado di presentare i diversi tipi di grafici, come *confusion matrix*, *scatter plot* o curve ROC, atti alla verifica delle caratteristiche risultanti del modello esaminato. Così com'è stato fatto per gli altri, tutti i risultati ottenuti sono stati riportati e largamente commentati nel capitolo 3, al paragrafo 3.3.

Capitolo 3

Risultati

Nel seguente capitolo sono presentati tutti i risultati ottenuti da questo lavoro di tesi. In particolare, sono stati riportati alcuni dei grafici più esplicativi riscontrati nella procedura dei confronti al paragrafo 2.4.3 e tutti i risultati ottenuti dalle classificazioni descritte al paragrafo 2.4.4. In merito a quest'ultimi, sono state mostrate le classificazioni ottenute tramite il modello di regressione logistica, configurato per la combinazione di due, tre e quattro parametri per volta, ognuno valutato anche nel caso di soglia mobile, ed il modello del Coarse Decision Tree. Entrambi i modelli sono stati valutati sia nel caso della procedura con l'intero dataset che in quella con dataset frazionato, che ne permette la validazione. Infine, per ogni modello e procedura implementata sono stati riportati tutti i grafici che ne descrivono le caratteristiche.

3.1 Riscontro dell'operazione di confronto

Dalla trattazione descritta al paragrafo 2.4.3, alla conclusione di tutte le iterazioni sono stati ottenuti i risultati riguardanti i confronti degli andamenti di analoghi parametri estratti da pazienti dello stesso gruppo (confronti intraclasse), sia per la classe 0 che per la classe 1, ed i confronti di andamenti analoghi tra pazienti delle due classi differenti (confronti interclasse). In sostanza, nel primo caso è stato contato il numero totale di volte in cui l'algoritmo ha evidenziato una corrispondenza tra gli andamenti dello stesso parametro estratto in pazienti con ugual qualità vocale. Il conteggio delle corrispondenze individuate nelle due classi è stato accorpato e mostrato come unico risultato nella tabella conclusiva. Per i confronti interclasse, invece, sono state conteggiate tutte le non corrispondenze accertate tra gli andamenti di medesimi parametri estratti da soggetti con qualità vocali differenti.
Ciò è stato svolto allo scopo di mostrare un'eventuale natura delle features di riuscire a discriminare la diversità delle classi in esame oppure di essere una grandezza di misura caratteristica di una o dell'altra. In tabella 10 è riportato l'esito dei due confronti:

***Tabella 10**: Risultati dei confronti intraclasse ed interclasse.*

Parametri	Confronti Intraclasse	Confronti Interclasse
MFCC 1	7 su 110	116 su 121
MFCC 2	3 su 110	112 su 121
MFCC 3	2 su 110	116 su 121
MFCC 4	2 su 110	114 su 121
MFCC 5	3 su 110	116 su 121
MFCC 6	2 su 110	117 su 121
MFCC 7	3 su 110	118 su 121
MFCC 8	1 su 110	118 su 121
MFCC 9	2 su 110	118 su 121
MFCC 10	2 su 110	117 su 121
MFCC 11	3 su 110	114 su 121
MFCC 12	1 su 110	119 su 121
MFCC 13	2 su 110	116 su 121
Entropia spettr.	2 su 110	117 su 121
Kurtosis spettr.	1 su 110	117 su 121
Log Spectral tilt	3 su 110	115 su 121
SPI	9 su 110	116 su 121
CPPS	15 su 110	100 su 121

Da notare che il numero totale di confronti effettuati nella procedura intraclasse è pari a 110, mentre per quella interclasse è 121. Il primo è stato ricavato dal conteggio di tutte le possibili combinazioni rinvenute confrontando due pazienti per volta senza alcuna ripetizione, il quale rappresenta il calcolo del coefficiente binomiale (formula 2.26). Poiché si hanno undici parametri da combinare due per volta, il risultato del coefficiente binomiale è pari a 55 per ogni classe, quindi, 110 in totale. Il secondo risultato rappresenta anch'esso tutte le possibili combinazioni ottenibili, ma, nel caso dei confronti interclasse, ciascun soggetto di una classe viene comparato con ogni paziente dell'altra, quindi, il numero totale di confronti eseguiti è pari a 11×11=121.

Osservando la colonna relativa ai confronti intraclasse è stato possibile notare un basso numero di test superati, il che si traduce in una scarsa capacità dei parametri di essere

delle grandezze di misura descrittive di una delle due classi. In altre parole, essi non sono stati in grado di appurare, nella maggior parte dei casi, se un determinato andamento sia identificativo di una classe o dell'altra. Una debole contro-tendenza è stata rilevata per il CPPS, il quale evidenzia un grado di correlazione leggermente maggiore rispetto agli altri parametri considerati, ma non sufficiente ad indicare un qualche potere discriminante. In figura 19 sono riportati alcuni di questi andamenti correlati del CPPS, estratti dai confronti all'interno della classe 0, e riportati sottoforma di istogrammi, i quali mettono in correlazione tutti i valori assunti dal parametro cepstrale (posizionati nelle ascisse) con il numero di volte che questo risulta assumerli lungo i frame analizzati (nelle ordinate indicato con N):

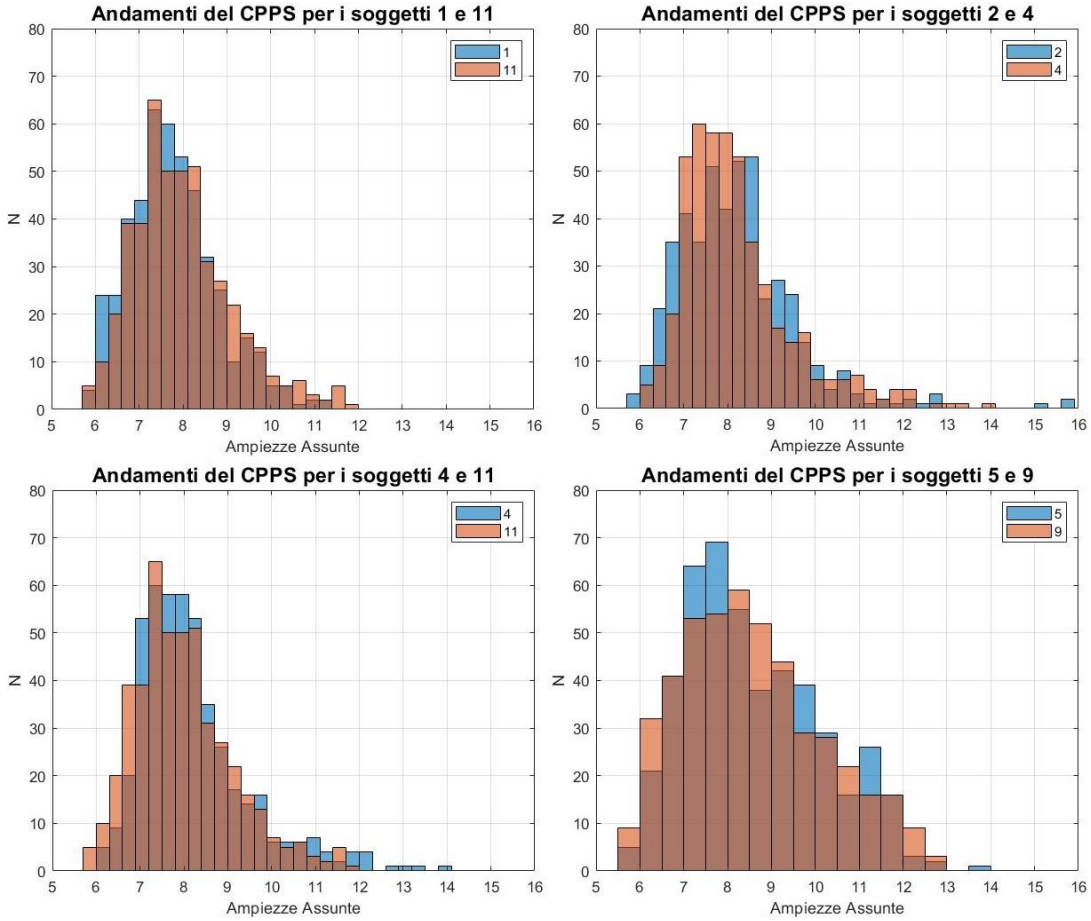

***Figura 19**: Esempio di correlazione degli andamenti del CPPS per la classe di pazienti con qualità vocale buona.*

Il soggetto 1 indica il terzo paziente del gruppo degli OPHL II, il soggetto 2 il quarto paziente degli OPHL II, il soggetto 4 l'ottavo paziente degli OPHL II, il soggetto 5 il decimo paziente degli OPHL II, il soggetto 9 il ventunesimo paziente degli OPHL II ed il soggetto 11 il secondo pazienti degli OPHL III.

In totale contrapposizione è l'andamento dei confronti interclasse, i quali sono stati in grado di produrre risultati molto più promettenti, mostrando nella relativa colonna un elevato numero di test superati. Ciò è stato interpretato come una miglior predisposizione dei parametri a individuare le differenze esistenti negli andamenti appartenenti a pazienti di classe diverse. Nella figura 20 sono riportati a titolo d'esempio alcuni andamenti di MFCC12, in forma di istogrammi:

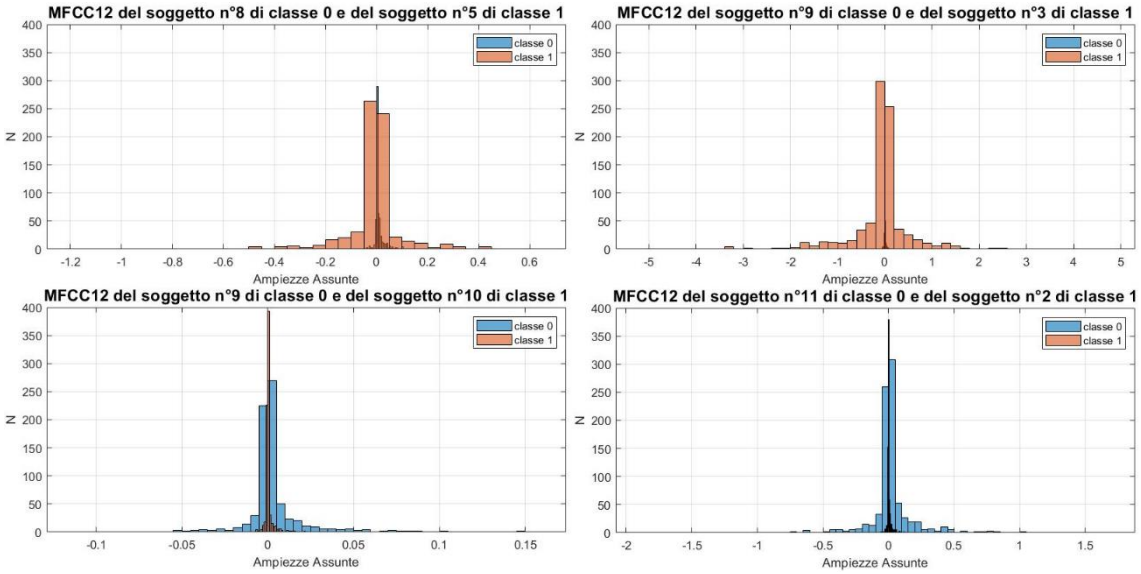

***Figura 20**: Esempio di non correlazione degli andamenti di MFCC12 per la classe di pazienti con bassa qualità vocale.*

3.2 Classificazione mediante il metodo della regressione logistica

A seguito dei passaggi descritti nel paragrafo 2.4.4 sono stati ottenuti, per ciascuna configurazione di combinazione dei parametri e con il dataset completo, i relativi set di parametri che hanno restituito una buona accuratezza di classificazione, i valori di precisione, sensitività e specificità. Inoltre, per ciascuna configurazione di combinazioni è stato preso il set con le migliori caratteristiche di classificazione ed è stato valutato con soglie diverse, vicine a quella originariamente impostata. Questo è stato fatto allo scopo di mostrare le diverse caratteristiche di predizione acquisite dal modello al variare della soglia. Infine, prendendo la combinazione che ha riportato il miglior risultato di accuratezza tra tutte quelle vagliate, è stata suddiviso il dataset ed è stata condotta l'effettiva validazione del modello generale.

3.2.1 Combinazioni di due parametri per volta

Dall'equazione 2.26 è stato calcolato il numero totale di combinazioni ottenute dalle 162 features associate due per volta, il quale risulta pari a 13041 e che rappresenta anche il numero delle iterazioni da compiere. Alla fine di tutte queste operazioni sono stati riscontrati i risultati riassunti in tabella 11:

***Tabella 11**: Risultato della classificazione con combinazione di due parametri a soglia fissa.*

Parametri		Precisione	Sensitività	Specificità	Accuratezza
10	*25*	87.5%	63.64%	95.24%	84.38%
9	*122*	85.71%	54.55%	95.24%	81.25%
10	*122*	85.71%	54.55%	95.24%	81.25%
10	*149*	77.78%	63.64%	90.48%	81.25%
18	*152*	77.78%	63.64%	90.48%	81.25%
19	*152*	77.78%	63.64%	90.48%	81.25%

dove il parametro 9 indica la skewness calcolata sull'andamento di MFCC1, il parametro 10 la kurtosis di MFCC1, il parametro 18 la skewness di MFCC2, il parametro 19 la kurtosis di MFCC2, il parametro 25 il 5° percentile di MFCC3, il parametro 122 la deviazione tipo dell'entropia spettrale, il parametro 149 la deviazione tipo di SPI ed il parametro 152 il 95° percentile di SPI. È stato possibile notare che la coppia di parametri 10 e 25 mostra, oltre massima all'accuratezza, anche il miglior compromesso tra i valori di sensitività e specificità, in ragion di ciò tale set è stato scelto come candidato per la valutazione della soglia mobile. Anche la coppia 10-149, 18-152 e 19-152 mostrano un buon equilibrio tra i due valori, ma alla luce della loro più bassa accuratezza è stato deciso di non tenerli in considerazione. La procedura con soglia mobile è stata basata sulla ripetizione della classificazione per i soli due parametri selezionati, con un ciclo che ad ogni iterazione faccia variare la soglia utilizzata per catalogare i valori della predizione del modello, originariamente posta a 0.5, in un range da 0.4 a 0.6 a passi di 0.05. In questo modo sono stati ottenuti i risultati presentati in tabella 12:

***Tabella 12**: Risultato della classificazione con i parametri 10 e 25 con soglia mobile.*

Soglia	Precisione	Sensitività	Specificità	Accuratezza
0.4	72.73%	72.73%	85.71%	81.25%
0.45	88.89%	72.73%	95.24%	87.5%
0.5	87.5%	63.64%	95.24%	84.38%
0.55	80%	36.36%	95.24%	75%
0.6	75%	27.27%	95.24%	71.88%

da cui è stato possibile evidenziare un notevole calo dell'accuratezza e delle altre caratteristiche per le soglie più alte, un miglioramento nell'equilibrio dei valori di sensitività e specificità per la soglia più bassa ed un aumento dell'accuratezza fino all'87.5%, per il valore di sbarramento di 0.45.

Infine, in tabella 13 è mostrata la *confusion matrix* del modello della combinazione 10-25 a soglia fissa per presentare con maggior chiarezza la bontà della classificazione ottenuta:

***Tabella 13**: Confusion matrix per la combinazione dei parametri 10 e 25.*

		Classe Reale	
		7	1
Classe predetta		4	20

3.2.2 Combinazioni di tre parametri per volta

Dalla formula 2.26 si evince che è stato svolto un numero di iterazioni, corrispondente anche al numero di combinazioni ottenute per questa configurazione, pari a 695520. In seguito all'elevato numero di cicli compiuti dall'algoritmo, sono stati ottenuti molti modelli adeguati, ma in tabella 14 sono riassunti solo quelli più significativi:

***Tabella 14**: Risultato della classificazione con combinazione di tre parametri a soglia fissa.*

Parametri			Precisione	Sensitività	Specificità	Accuratezza
9	*154*	*156*	80%	72.73%	90.48%	84.38%
18	*149*	*156*	87.5%	63.64%	95.24%	84.38%
25	*37*	*156*	87.5%	63.64%	95.24%	84.38%
109	*145*	*156*	87.5%	63.64%	95.24%	84.38%
3	*10*	*30*	72.73%	72.73%	85.71%	81.25%
8	*15*	*106*	85.71%	54.55%	95.24%	81.25%

9	*127*	*149*	85.71%	54.55%	95.24%	81.25%
9	*146*	*152*	77.78%	63.64%	90.48%	81.25%
10	*154*	*156*	77.78%	63.64%	90.48%	81.25%
18	*122*	*156*	77.78%	63.64%	90.48%	81.25%
19	*100*	*109*	72.73%	72.73%	85.71%	81.25%
27	*37*	*159*	77.78%	63.64%	90.48%	81.25%

in cui i set di parametri di maggior interesse risultano essere il 9-154-156, il 3-10-30 ed il 19-100-109, i quali valori rappresentano rispettivamente: il 3 la mediana calcolata sull'andamento di MFCC1, il 9 è la skewness di MFCC1, il 10 è la kurtosis di MFCC1, il 19 è la kurtosis di MFCC2, il 30 è la mediana di MFCC4, il 100 è la kurtosis di MFCC11, il 109 è la kurtosis di MFCC12, il 154 è la kurtosis del SPI ed il 156 è la mediana del CPPS. La prima configurazione presentata è stata quella che ha ottenuto il grado maggiore di accuratezza, ma non un perfetto bilanciamento tra i valori di sensitività e specificità. Al contrario, le altre due configurazioni sono state in grado di evidenziare un netto miglioramento nell'equilibrio delle due caratteristiche, a fronte, però, di un leggero calo di accuratezza. Tutte e tre sono state oggetto della procedura di classificazione con soglia mobile, ma la prima e la seconda non hanno portato significativi miglioramenti dei risultati, bensì a dei peggioramenti più o meno sostanziali; perciò, in tabella 15 sono presentati solo i risultati derivati dalla terza configurazione:

***Tabella 15**: Risultati della classificazione con i parametri 19, 100 e 109 con soglia mobile.*

Soglia	Precisione	Sensitività	Specificità	Accuratezza
0.4	69.23%	81.82%	80.95%	81.25%
0.45	69.23%	81.82%	80.95%	81.25%
0.5	72.73%	72.73%	85.71%	81.25%
0.55	88.89%	72.73%	95.24%	87.5%
0.6	88.89%	72.73%	95.24%	87.5%

È stato riscontrato un ottimo bilanciamento delle diverse caratteristiche nei valori soglia più bassi, mentre in quelli più alti sono state ottenute delle caratteristiche più disomogenee, a fronte di un netto aumento dell'accuratezza fino a percentuali dell'87.5%.
In conclusione, la tabella 16 mostra la *confusion matrix* relativa al set di parametri che ha mostrato l'accuratezza maggiore in assoluto nella procedura con soglia fissa:

***Tabella 16**: Confusion matrix per la combinazione dei parametri 9, 154 e 156.*

		Classe Reale	
		Classe reale	
Classe predetta		8	2
		3	19

3.2.3 Combinazioni di quattro parametri per volta

Per tale configurazione è stato ottenuto un numero di combinazioni pari a 27646920 e, anche qui, a seguito di questo elevato numero di iterazioni, sono stati riscontrati molti modelli in grado di presentare un buon livello di classificazione, ma in tabella 17 sono riportati solo quelli con le caratteristiche più efficienti:

***Tabella 17**: Risultato della classificazione con combinazione di quattro parametri a soglia fissa.*

Parametri				Precisione	Sensitività	Specificità	Accuratezza
21	*33*	*53*	*80*	100%	81.82%	100%	93.75%

3	*21*	*47*	*103*	90%	81.82%	95.24%	90.63%
26	*52*	*67*	*103*	90%	81.82%	95,24%	90.63%
8	*48*	*71*	*103*	88.89%	72.73%	95.24%	87.5%
15	*21*	*78*	*106*	100%	63.64%	100%	87.5%
18	*124*	*137*	*156*	88.89%	72.73%	95.24%	87.5%
21	*49*	*105*	*107*	88.89%	72.73%	95.24%	87.5%
21	*52*	*74*	*95*	88.89%	72.73%	95.24%	87.5%
21	*52*	*103*	*115*	81.82%	81.82%	90.48%	87.5%

Da quanto riportato in tabella 17, sono stati individuati due modelli aventi le migliori caratteristiche di classificazione: il primo è quello relativo al set di variabili 21, 33, 53, 80, le quali rappresentano rispettivamente la mediana di MFCC3, il range di MFCC4, il 95° percentile di MFCC6 ed il 95° percentile di MFCC9. In esso è stata ottenuta l'accuratezza massima, sia all'interno della configurazione stessa che per tutte le configurazioni di combinazioni di parametri implementate. Inoltre, è stato in grado di mostrare un bilanciamento quasi ottimale tra sensitività e specificità; il secondo è quello con il set 3, 21, 47, 103, le quali rappresentano la mediana di MFCC1, la mediana di MFCC3, la media di MFCC6, la moda di MFCC12, ed è stata ottenuta un'accuratezza quasi massima, ma un miglior equilibrio tra sensitività e specificità. Entrambi i modelli sono stati scelti per l'analisi con soglia mobile, ma il secondo non ha mostrato alcuna variazione lungo i diversi sbarramenti imposti, perciò, in tabella 18 si riporta unicamente l'esito del primo modello:

***Tabella 18**: Risultato della classificazione con i parametri 21, 33, 53 e 80 con soglia mobile.*

Soglia	Precisione	Sensitività	Specificità	Accuratezza
0.4	81.82%	81.82%	90.48%	87.5%
0.45	81.82%	81.82%	90.48%	87.5%
0.5	100%	81.82%	100%	93.75%
0.55	100%	81.82%	100%	93.75%
0.6	100%	81.82%	100%	93.75%

Come evidenziato dalla tabella sopra riportata, a valori di soglia più alti sono state ricavate accuratezze e specificità più elevate, mentre per valori più bassi sono state ottenute coppie di sensitività-specificità maggiormente equilibrate, a discapito di un sensibile calo dell'accuratezza. Tutto ciò è stata un'ulteriormente prova della grande efficienza di tale configurazione e dei parametri su cui si basa.
Infine, è stato deciso di riportare in tabella 19 la *confusion matrix* del modello a soglia fissa caratterizzato dall'accuratezza più alta tra le varie ottenute:

***Tabella 19**: Confusion matrix per la combinazione dei parametri 21, 33, 53 e 80.*

		Classe Reale	
Classe predetta		9	0
		2	21

3.2.4 Validazione generale del modello

Tale fase, per praticità, è stata totalmente svolta all'interno dell'applicazione "Classification Learner". Al momento del caricamento della matrice *Data* contenente tutti i dati a disposizione, è stata scelta l'impostazione di eseguire una fase di validazione mediante la cross-validation. Essa ha permesso di suddividere il dataset in cinque porzioni equivalenti in termini di capienza ed in ciascuna è stato eseguito il training su alcune osservazioni e la validazione su delle altre totalmente diverse rispetto alle prime. Infine, è stato selezionato nel relativo menù il modello della regressione logistica e sono state selezionate le features 21 (mediana di MFCC3), 33 (range di MFCC4), 53 (95° percentile di MFCC6) e 80 (95° percentile di MFCC9), le quali hanno riportato i migliori risultati di classificazione, come mostrato in tabella 17. Alla fine di questa veloce

preparazione, è stato dato inizio alle iterazioni ed è stata restituita un'accuratezza di validazione pari all'81.2%. Ciò rileva una leggera flessione rispetto a quanto ottenuto con le classificazioni mediante l'intero dataset, ma il risultato è ancora statisticamente accettabile e nettamente più realistico in confronto al precedente. Infine, sono state riportate le caratteristiche del modello mediante la rappresentazione della *confusion matrix*, in tabella 20, e della curva ROC (*Receiver Operating Characteristic curve*), in figura 21. Quest'ultima traccia il valore di sensitività del classificatore in funzione della sua specificità e ne restituisce una curva, la quale mostra le prestazioni del modello in esame per qualsiasi soglia di classificazione si scelga di impostare. In questo modo, è possibile individuare visivamente le caratteristiche presentate dal modello corrente, in termini di elementi classificati come corretti positivi e corretti negativi, e le diverse caratteristiche ottenibili con soglie di classificazioni differenti. Inoltre, in tale curva ROC è possibile evidenziare il valore di AUC (*Area Under Curve*), il quale misura l'area sottesa dalla curva ROC e valuta il potere di differenziazione del modello tra le due classi esaminate.

***Tabella 20**: Confusion matrix del modello di regressione logistica validato.*

		Classe Reale
Classe predetta	18	3
	3	8

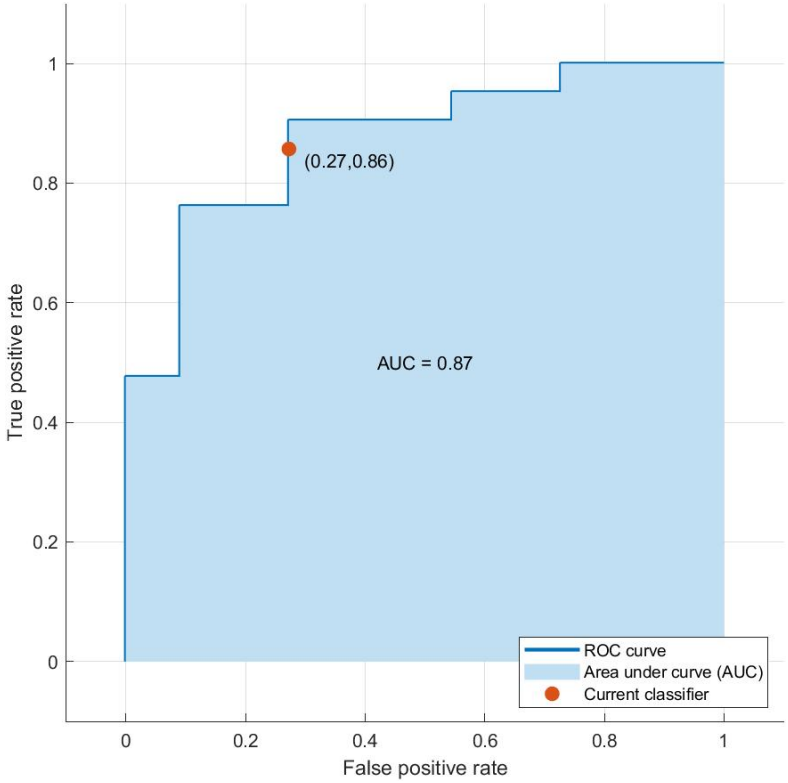

***Figura 21**: Curva ROC del modello di regressione logistica validato.*

È stato possibile osservare un valore di true positive rate (ossia di sensitività) pari all'86%, il quale si traduce in una corretta classificazione per buona parte dei soggetti della classe 0, ed un valore di AUC pari all'87% che esprime un'elevata qualità di classificazione delle previsioni ottenute.

3.3 Classificazione mediante il modello del Coarse Decision Tree

In seguito alla procedura ampiamente descritta nel paragrafo 2.4.4, dal modello decisionale implementato attraverso l'applicazione Matlab è stata ottenuta

un'accuratezza di classificazione pari al 93.8%, un valore che sembra confermare quanto ottenuto per la regressione logistica.

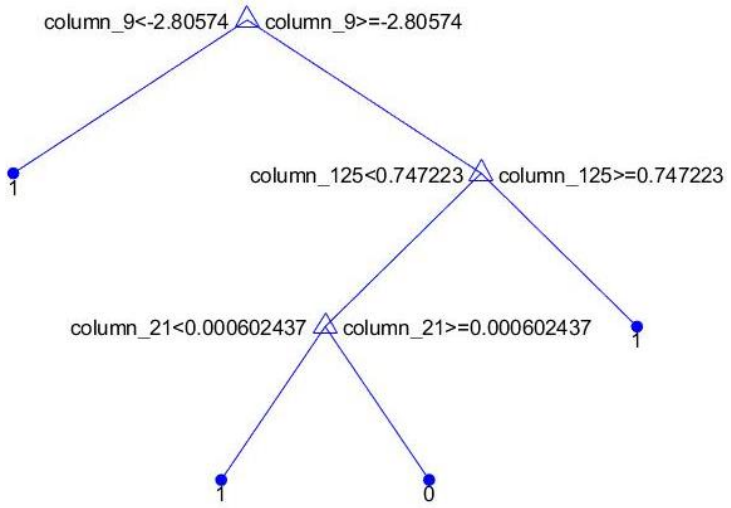

***Figura 22**: Grafico del Coarse Decision Tree implementato.*

Com'è possibile notare dalla figura 22 rappresentate il grafico decisionale del modello implementato, sono state eseguite tre suddivisioni, nella quali sono stati considerati tre diversi parametri, ossia: la skewness di MFCC1, indicata dalla colonna 9, il 95° percentile dell'entropia spettrale, riferita alla colonna 125 e la mediana di MFCC3, indicata dalla colonna 21. In base ai valori assunti da questi parametri per ciascun paziente analizzato, è stata ricavata la predizione dell'appartenenza di ognuno alla classe 0 o alla classe 1. Di seguito è stata presentata anche la curva ROC:

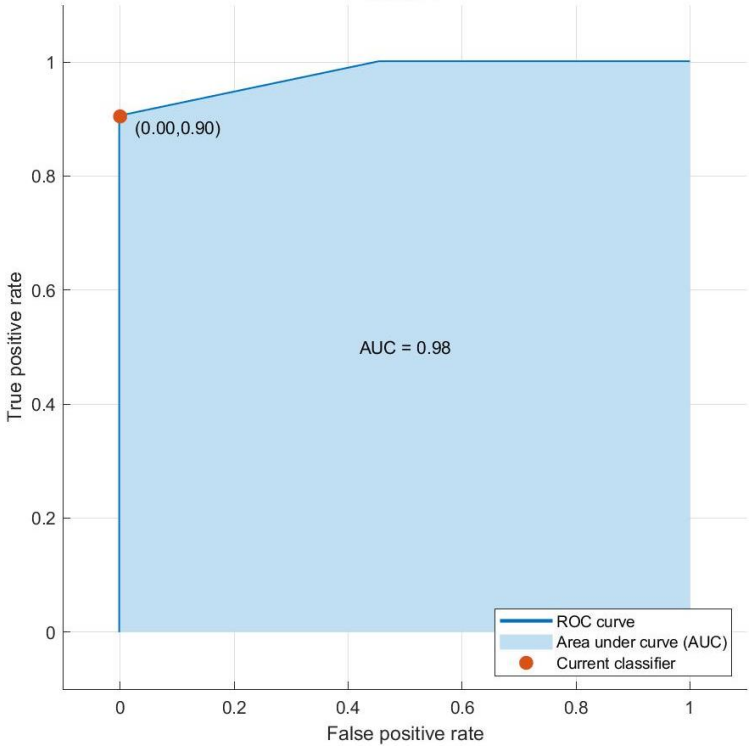

***Figura 23**: Curva ROC del modello decisionale ottenuto.*

in cui è stato possibile osservare un valore di sensitività del 90% ed un valore di AUC pari al 98%, dimostrando così le ottime prestazioni possedute dal classificatore in esame. Infine, in tabella 21 è riportata la relativa *confusion matrix*:

***Tabella 21**: Confusion matrix del Coarse Decision Tree ottenuto.*

	Classe Reale	
Classe predetta	19	2
	0	11

3.3.1 Processo di validazione

Come per il modello della regressione logistica, il processo di validazione per il modello del *Coarse Decision Tree* è stato effettuato all'interno del "Classification Learner". In particolare, sono state adottate le medesime scelte iniziali del primo modello, ossia la selezione della *cross-validation* con suddivisione del dataset in cinque parti, ma è stato selezionato il modulo relativo al Decision Tree e sono state reclutate le stesse features presentate in figura 23, ovvero: la skewness di MFCC1 (parametro 9), il 95° percentile dell'entropia spettrale (parametro 125) e la mediana di MFCC3 (parametro 21). Alla conclusione delle iterazioni è stato rinvenuto un valore di accuratezza di tale modello validato pari all'81.2%. Anche in questo caso è stato constato un calo delle prestazioni, ma tale risultato è stato considerato ancora statisticamente accettabile e, soprattutto, molto più veritiero e realistico di quanto ottenuto con la procedura che coinvolge l'intero dataset.
Nella tabella 22 è riportata la corrispondente *confusion matrix*, mentre la figura 24 mostra la relativa curva ROC.

***Tabella 22**: Confusion matrix del Coarse Decision Tree validato.*

		Classe Reale
Classe predetta	17	4
	2	9

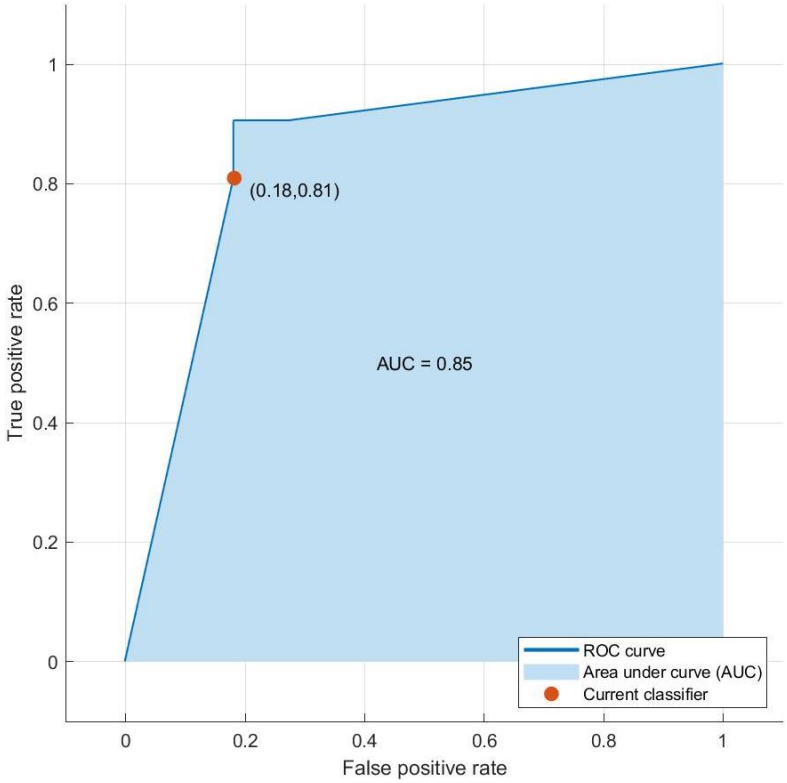

***Figura 24**: Curva ROC del Coarse Decision Tree validato.*

Capitolo 4

Conclusioni e Sviluppi futuri

Il seguente lavoro, con i relativi risultati ottenuti, è stato in grado di portare in evidenza alcuni aspetti importanti riguardanti la trattazione dei segnali vocali di pazienti laringectomizzati e, in particolare modo, di registrazioni relative all'eloquio libero di tali soggetti. È stato appurato che la frammentazione del segnale, volta alla ricerca delle sole finestre armoniche, si è rilevata fondamentale per ottenere delle informazioni proficue dalle successive analisi. L'intuizione di tale studio, in merito a questo passo, è stata quella di valutare il parametro della kurtosis spettrale ad ogni frame di segnale e su esso basare il criterio di scelta. Questo procedimento si è rivelato molto promettente, difatti, in seguito ad un attento ascolto dei brani pre e post-elaborazione, è risultata evidente la precisione della suddivisione dei frame armonici da quelli inarmonici, sia quando adoperato per soggetti sani che per laringectomizzati. Una pecca, tuttavia, è che tale metodologia risulta statisticamente poco salda a causa delle variabili modalità di acquisizione delle registrazioni dei soggetti sani, dai quali è stata validata la soglia definitiva, e a causa dell'esigua quantità di dati di pazienti disfonici su cui è stata applicata. Sarebbe consigliabile comprovarne la fedeltà su un numero di dati più consistente provenienti sia da soggetti sani, acquisiti con metodi più accurati in modo da ricavare una soglia globale analiticamente e statisticamente significativa, che da pazienti con diverse tipologie di disfonia, così da validarne l'efficacia in ogni situazione d'interesse.

I parametri selezionati per tale lavoro hanno mostrato un evidente adeguatezza per l'analisi degli eloqui. Da ciò, quindi, è stato possibile dimostrare che il dominio spettrale e cepstrale sono in grado di fornire informazioni più fruttuose per i pazienti analizzati rispetto ai consueti parametri di stabilità del periodo o dell'ampiezza, nonostante gli esigui intervalli armonici riscontrati all'interno delle registrazioni, e di ottenere dati non eccessivamente compromessi, malgrado l'elevata presenza di disturbi nelle voci dei soggetti studiati. In evidenza di ciò, sono state rilevate delle accuratezze di classificazione relativamente elevate, superando tutti i risultati ottenuti con la prima versione dell'elaborato. Tuttavia, sempre a causa del ristretto dataset adoperato e al

ridotto numero di tipologie di laringectomizzati trattati, i risultati ottenuti non sono da considerarsi universalmente validi, quindi, potrebbe essere interessante saggiare la loro validità su dataset più ampi ed estendere l'analisi coinvolgendo anche altri tipi di disfonie. Oltre ai possibili sviluppi fin ora citati, potrebbe risultare utile un riarrangiamento degli algoritmi di classificazione, in modo da far fronte ad una più ampia mole di dati da trattare. Potrebbe essere vantaggioso ricercare ulteriori parametri nel dominio spettrale e cepstrale, così da ovviare eventuali problematiche causate dall'eterogeneità dei dati dalle diverse tipologie di disfonia. Infine, da un punto di vista clinico, potrebbe essere consigliabile affiancare informazioni addizionali da più scale qualitative, in modo da allenare con maggior precisione gli algoritmi di classificazione e far fronte all'eterogeneità dei dati prima menzionata.

Tale studio, nella prospettiva di un progetto più ampio riguardo la costruzione di un dispositivo riabilitativo, non mira a sostituirsi ai metodi implementati in passato, bensì tenta di integrarsi a loro. Fatta presente tale precisazione, il lavoro svolto si è, quindi, sforzato di perfezionare il corredo di procedure disponibili per la trattazione delle voci disfoniche, offrendo un metodo di valutazione per le registrazioni di eloqui e, di conseguenza, per cercare di effettuare un altro passo nella ricerca al raggiungimento di analisi quantitative delle attività logopediche. Con ciò, e con i futuri studi che seguiranno questo lavoro, si auspica la costruzione di tale dispositivo riabilitativo in tempi brevi, che garantisca lo svolgimento di attività rieducative efficaci in qualsiasi situazione, il massimo recupero di voce consentito ad ogni paziente ed il ritorno, per ciascuno di essi, ad una più regolare conduzione delle interazioni sociali.

Bibliografia

[1] "*Come si produce la voce.*" (2011), URL: http://fisicaondemusica.unimore.it/Voce_umana.html

[2] *"Tumore alla laringe"*, URL: https://www.ultraspecialisti.com/aree-sanitarie/oncologia/tumore-alla-laringe/

[3] Li Stan Z, Jain Anil K, "*Encyclopedia of Biometrics*" (2009), Springer, pp. 1323.

[4] Castellana A, Carullo A, "*Discriminating Pathological Voice from Healty Voice using Cepstral Peak Prominence Smoothed Distribution in Sustained Vowel*" (2018), IEEE Transactions on Instrumentation and Measurement, vol. 67 (3), 646-654.

[5] Fondazione AIRC, *"Tumori di laringe e faringe"* (2018), URL: https://www.airc,it/cancro/informazioni-tumori/guida-ai-tumori/tumori-faringe-laringe

[6] Istituto Clinico Humanitas *"Laringectomia"*, URL: https://www.humanitas.it/cure/laringectomia/

[7] Società Italiana di Otorinolaringologia e Chirurgia Cervico-Facciale *"Le laringectomie parziali orizzontali"*, URL: https://www.sioechcf.it/wp-content/uploads/2017/06/Laringectomia_Parziali-Orizzontali.pdf

[8] Fondazione IRCCS Istituto Nazionale dei Tumore *"Percorso riabilitativo del paziente laringectomizzato"*, URL: https://www.istitutotumori.mi.it/percorso-riabilitativo-del-paziente-laringectomizzato-s.c.-otorinolaringoiatria

[9] Di Nicola V, Fiorella M.L, *"acoustic analysis of voice in patients traeted by reconstructive subtotal laryngectomy. Evaluation and critical review"* (2006),

Acta Otorhinolaryngol Ital 26, 59-68.

[10] Makeieff M, Barbotte E, *"Acoustic and Aerodynamic Measurement of Speech Production after Supracricoid Partial Laryngectomy"* (2005), The Laryngoscope, 115:546-551, DOI 10.1097/01.mlg.0000157848.78530.ee.

[11] Kosztyła-Hojna B, Łuczaj J, *"Perceptual and acoustic voice analysis in patient with glottis cancer after endoscopic laser cordectomy"* (2020), Otolaryngol Pol, 74 (3), 23-28, DOI 10.5604/01.3001.0013.7850.

[12] Krengli M, Policarpo M, *"Voice quality after treatment for T1a glottic carcinoma Radiotherapy Versus Laser Cordectomy"* (2004), Acta Oncologica, 43 (3), 284-289, DOI: 10.1080/02841860410026233.

[13] E. van Sluis K, van dei Molen L, *"Objective and Subjecctive Voice Outcomes after Total Laryngectomy: a systematic review"* (2018), Eur Arch Otorhinolaryngol, 275, 11-26, DOI 10.1007/s00405-017-4790-6.

[14] Chhetri S. S, Gautam R, *"Acoustic analysis before and after voice therapy for laryngeal pathology"* (2015), Kathmandu Univ Med J, 52(4), 323-327.

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

[16] Software Instruction Manual *"Multi-Dimensional Voice Program (MDVP)"*, Model 5105, Appendix C, 135-189.

[17] Atzori A, Carullo A, *"Parkinson disease voice features for rehabilitation therapy and screening purposes"* (2019), IEEE International Symposium on Medical Measurements and Applications (MeMeA), Istanbul, Turkey, DOI: 10.1109/MeMeA.2019.8802223.

[18] Soumaya Z, Taoufiq B. D, *"The detection of Parkinson using the genetic algorithm and SVM classifier"* (2021), Applied Acoustic 171, https://doi.org/10.1016/j.apacoust.2020.107528

[19] Karan B, Sahu S. S, *"Parkinson disease prediction using intrinsic mode function based features from speech signal"* (2019), Biocybernectics and Biomedical Engineerring, 40, 249-264.

[20] César M, Rufiner H. L, *"Acoustic analysis of speech for detection of laryngeal pathologies"* (2000), IEEE Xplore, 2369-2372, DOI: 10.1109/IEMBS.2000.900621.

[21] Bogert B. P, Healy M. J. R, *"The quefrency alanysis of time series for echos: cepstrum, pseudo autocovariance, cross cepstrum and saphe"* (1963), Proceedings of Symposium on Time Series Analysis, 209-243.

[22] Childers D. G, Skinner D. P, *"The cepstrum: A guide to processing"* (1977), Proceedings of the IEEE, vol. 65 (10), 1428-1443.

[23] Randall R. B, *"A history of cepstrum analysis and its application to mechanical problems"* (2017), Mechanical System and Signal Processing, vol. 97, 3-19.

[24] Hillenbrand J, Houde R. A, *"Acoustic Correlates of Breathy Vocal Quality: Dysphonic Voices and Continuous Speech"* (1996), Journal of Speech and Hearing Research, vol. 39, 311-321.

[25] Pietruch R. W, Grzanka A. D, *"Vowel recognition of patient after total laryngectomy using Mel Frequency Cepstral Coefficients and mouth contour"* (2010), Journal of Automatic Control, University of Belgrade, vol. 20 (1), 33-38, DOI: 10.2298/JAC1001033P.

[26] Peeters G, *"A Large Set of Audio Features for Sound Description (Similarity and Classification) in the CUIDADO Project."* (2004), Technical Report, IRCAM.

[27] Muda L, Begam M, *"Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW)"* (2010), Journal of Computing, vol. 2 (3), 138-143, ISSN 2151-9617.

[28] Chakroborty S, Roy A, *"Fusion of a Complementary Feature Set with MFCC for Improved Closed Set Text-Independent Speaker Identification"* (2006), IEEE International Conference on Industrial Technology, 387-390, DOI: 10.1109/ICIT.2006.372388.

[29] Manwa L. Ng, Liu H, Zhao Q, *"Long-term average spectral characteristics of Cantonese alaryngeal speech"* (2009), Auris, nasus, larynx, vol. 36 (5), 571–577, DOI:10.1016/j.anl.2008.12.005.

[30] Mathew M, Bhat J, *"Soft phonation index - A sensitive parameter?"* (2009), Indian Journal of Otolaryngology and Head and Neck Surgery: official publication of the Association of Otolaryngologists of India, vol. 61, 127-130, DOI:10.1007/s12070-009-0050-4.

[31] Watada J, *"Kolmogorov-Smirnov Two Sample Test with Continuous Fuzzy Data"* (2010), 175-186.

[32] Leon A. C, *"Comprehensive Clinical Psychology"* (1998), Elsevier Science Ltd, 278-279.

[33] Shao Y. E, Hou C. D, *"Hybrid intelligent modeling schemes for heart disease classification"* (2014), Applied Soft Computing, vol.14 (A), 47-52, ISSN 1568-4946, doi.org/10.1016/j.asoc.2013.09.020.