

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE
Corso di Laurea Magistrale in Matematica

Analisi prosodica della frase
mediante omologia persistente

Tesi di Laurea in Topologia Algebrica

Relatore:
Chiar.mo Prof.
Massimo Ferri

Correlatore:
Dott.
Marco Barone

Presentata da:
Elena Gironi

Correlatore:
Dott.
Ivan Tomba

III Sessione
Anno Accademico 2016/2017

*Alle voci
che vibrano
e che si affermano*

Indice

Introduzione	i
1 Omologia Persistente	1
2 Prosodia	11
2.1 Elementi di acustica del parlato	11
2.2 Elementi di linguistica	13
2.3 Analisi del contorno intonativo	15
2.4 ToBI: sistema di annotazione	18
3 Descrizione della ricerca	21
3.1 Studi precedenti	21
3.2 Obiettivi della ricerca	23
3.3 Strumenti della ricerca	24
3.4 Analisi effettuate	29
4 Risultati	33
4.1 Risultati ottenuti	33
4.2 Conclusioni e obiettivi futuri	43
A Codici Dendrogrammi	45
B Dendrogrammi	51
C Confronto diagrammi di persistenza	65
Bibliografia	73
Ringraziamenti	75

Introduzione

La presente tesi nasce dall'idea di applicare le tecniche di omologia persistente allo studio della prosodia e si pone come obiettivo la ricerca di funzioni filtranti in grado di catturare gli aspetti intonativi chiave che consentono la classificazione di profili intonativi e la conseguente distinzione fra le diverse modalità frasali. L'analisi dell'andamento intonativo, nel corso dell'emissione di un contenuto, è da tempo oggetto di studio da parte dei linguisti con lo scopo di identificare le caratteristiche comuni a una certa modalità (affermazione, domanda, ecc) e un certo intento pragmatico. Gli studi condotti hanno portato ad un'effettiva classifica dei profili intonativi ricorrenti per una certa modalità a seconda della varietà linguistica esaminata.

Il fulcro del progetto è il test, e la conseguente analisi, di alcune funzioni filtranti su un set di audio campione tratti dalla ricerca del Dott. Marco Barone. L'insieme dei 56 dati audio si suddivide in tre categorie equiparate ciascuna relativa ad una tipologia di frase: domande polari, domande parziali e liste. L'analisi svolta mira alla creazione di dendrogrammi attraverso algoritmi di clustering gerarchico che rendano visibile la classificazione dei dati ottenuta per mezzo delle funzioni filtranti considerate.

Nello specifico, la tesi è suddivisa in quattro capitoli: i primi due forniscono nozioni introduttive alle materie coinvolte mentre i restanti trattano la sperimentazione effettuata. Il primo capitolo dell'elaborato riprende i concetti propri dell'omologia persistente con attenzione alla descrizione degli argomenti utilizzati nel seguito. Tale teoria sta infatti alla base delle tecniche usate per raggiungere gli scopi descritti.

Nel secondo capitolo si introducono le definizioni basilari relative all'ambito linguistico di interesse: vengono infatti forniti elementi di acustica del parlato e di linguistica. Viene inoltre esposta l'analisi dei contorni intonativi svolta sin ora introducendo anche il sistema di annotazione utilizzato in tale ambito.

Il terzo capitolo introduce la ricerca effettuata nel progetto di tesi descrivendone gli strumenti usati e le analisi svolte. Ci si sofferma sulla descrizione del set di dati e sulle sue caratteristiche; attraverso tale osservazione si sono decise le funzioni filtranti di cui servirsi per l'analisi topologica dei dati. Vengono poi descritte le funzioni scelte che sono relative ai parametri di frequenza fondamentale e di volume di un enunciato. In partico-

lare vengono usate la funzione di altezza della frequenza fondamentale, la sua derivata e la funzione di altezza del volume.

Il quarto capitolo racchiude i risultati ottenuti dall'analisi dei dendrogrammi computati con diversi metodi di linkage e relativi alle funzioni filtranti definite ed a funzioni somma implementate per riassumere le informazioni apportate dalle tre funzioni filtranti. I commenti relativi si soffermano ad analizzare i punti di forza della ricerca avvenuta così come i punti problematici. Si forniscono infine indicazioni per eventuali future ricerche motivandole sulla base delle analisi raccolte.

Capitolo 1

Omologia Persistente

Nel presente capitolo introduciamo i concetti matematici relativi all'ambito della topologia persistente, sui quali si fonda la nostra ricerca. Il contenuto di questo capitolo è tratto da alcuni articoli [1], [2] e dal testo [4].

La topologia persistente viene usata per riconoscere e classificare forme diverse tra loro che hanno caratteristiche comuni.

Questo problema viene affrontato dalla geometria che riesce ad associare ed analizzare oggetti rigidi attraverso trasformazioni che usano matrici. Ad esempio esistono trasformazioni che consentono di sovrapporre immagini di oggetti aventi la stessa struttura usando come descrittori di forma delle misure geometriche.

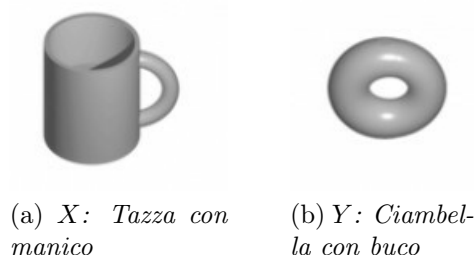
La geometria di questo tipo però è troppo rigida per riuscire a riconoscere le somiglianze tra oggetti, ad esempio, di origine naturale. Se si pensa alle immagini di un uomo in piedi e di un uomo seduto, la geometria delle trasformazioni matriciali non è in grado di riconoscere delle somiglianze mentre attraverso la topologia è possibile definire un omeomorfismo che associ le due figure. Infatti in ambito topologico si considera l'equivalenza tra spazi topologici piuttosto che la sovrapposibilità mediante trasformazioni.

In ogni caso la topologia consente di fare associazioni e collegamenti troppo ampi e generici, ecco allora che si introduce la topologia persistente che va a studiare non solo gli spazi topologici ma le coppie (X, f) in cui X è lo spazio topologico mentre f è una funzione continua, detta funzione filtrante, che consente di focalizzare alcuni aspetti, relativi alla forma, di cui si vuole tener conto. La topologia persistente dunque ci consente di usare la topologia e in più offre dei descrittori topologici che ci permettono di preservare certe caratteristiche geometriche richieste.

Questo strumento matematico è già largamente usato nelle applicazioni in campo biomedico, geologico e musicale per classificare e analizzare diverse tipologie di forme provenienti da immagini o altre fonti. Nel presente elaborato abbiamo usato questo strumento in ambito prosodico.

Procediamo ora a definire matematicamente i principali concetti propri della topolo-

gia persistente. Accompagniamo la trattazione con un esempio che ci aiuti a visualizzare quanto andremo a definire. Consideriamo a tal proposito due oggetti: una tazza con manico e una ciambella con un buco centrale.



Definizione 1.0.1 (Omeomorfismo).

Siano X e Y due spazi topologici e sia $\phi : X \rightarrow Y$ funzione continua con inversa continua. Allora chiamiamo ϕ *omeomorfismo* da X in Y e diciamo che X e Y sono *spazi omeomorfi*.

Relativamente al nostro esempio possiamo affermare che la tazza è omeomorfa alla ciambella infatti esistono deformazioni continue che permettono di trasformare l'oggetto X nell'oggetto Y . Esiste dunque un omeomorfismo da X a Y . Visivamente possiamo aiutarci con la seguente immagine 1.1.

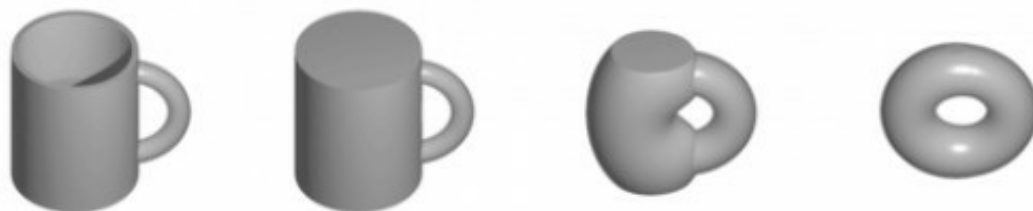


Figura 1.1: *Deformazioni continue che ci consentono di visualizzare l'omeomorfismo*

Quando parliamo di *omologia* intendiamo un formalismo matematico che ci consente di parlare in modo quantitativo di come uno spazio sia connesso. L'omologia in confronto ad altri formalismi, possiede algoritmi più veloci che però catturano meno informazioni topologiche. Nonostante ciò rimane un valido strumento per il confronto di spazi in quanto è in grado di ottenere informazioni relative ad una certa caratteristica.

Per definire correttamente i gruppi di omologia servono diverse nozioni preliminari per le quali rimandiamo al testo [4]. Di seguito ci limiteremo a definire intuitivamente i principali elementi per riuscire a definire poi cosa sia un gruppo di omologia nel contesto dell'omologia simpliciale a coefficienti in \mathbb{Z}_2 .

Definizione 1.0.2 (Simplesso).

Chiamiamo p -simplesso σ l'involuppo convesso, in uno spazio Euclideo, di un insieme di $p+1$ punti, detti *vertici* del simplesso, che non sono contenuti in un sottospazio Euclideo $(p-1)$ -dimensionale. Diciamo che il simplesso è generato dai suoi vertici e chiameremo *faccia* di un simplesso σ il simplesso τ generato da un sottoinsieme non vuoto di vertici di σ . In tal caso scriviamo $\tau \leq \sigma$.

Definizione 1.0.3 (Complesso simpliciale).

Un *complesso simpliciale* (geometrico) è una collezione finita K di semplici di un dato spazio Euclideo tali che:

1. Se $\sigma \in K$ e $\tau \leq \sigma$, $\Rightarrow \tau \in K$.
2. Se $\sigma_1, \sigma_2 \in K$ e $\sigma_1 \cap \sigma_2 \neq \emptyset$, $\Rightarrow \sigma_1 \cap \sigma_2$ è faccia di σ_1 e σ_2 .

Lo spazio di un complesso simpliciale K è il sottospazio topologico dello spazio Euclideo $|K|$ costituito dall'unione dei tutti i semplici di K .

Definizione 1.0.4 (p -catena).

Dato un complesso simpliciale finito K , chiamiamo p -catena ogni combinazione lineare formale di p -simplessi con coefficienti in \mathbb{Z}_2 . Dunque possiamo scrivere

$$\sum_i a_i \sigma_i$$

con $a_i \in \mathbb{Z}_2$ e σ_i p -simplesso.

Le p -catene formano uno spazio vettoriale a coefficienti in \mathbb{Z}_2 indicato con C_p .

Si osservi che ogni p -catena identifica un insieme di p -simplessi di K e che la somma di due p -catene corrisponde alla differenza simmetrica dei rispettivi insiemi.

Definizione 1.0.5 (Operatore di bordo).

Definiamo *operatore di bordo* la trasformazione lineare $\partial_p : C_p \rightarrow C_{p-1}$, $\forall p \in \mathbb{Z}$. È sufficiente definire tale operatore sui p -simplessi e poi estenderlo per linearità alle p -catene. Scriviamo allora il p -simplesso σ come $\sigma = [u_0, u_1, \dots, u_p]$ e indichiamo con $[u_0, \dots, \hat{u}_j, \dots, u_p]$ una particolare faccia di σ generata da tutti i suoi vertici, meno u_j , $j = (0, \dots, p)$. Dunque definiamo

$$\partial_p(\sigma) = \sum_{j=0}^n [u_0, \dots, \hat{u}_j, \dots, u_p].$$

Si dimostra che $\partial_p \partial_{p+1} = 0$, dunque si ha che l'immagine di ∂_{p+1} è contenuta nel nucleo di ∂_p , cioè $\text{Im} \partial_{p+1} \subset \text{Ker} \partial_p$.

Definizione 1.0.6 (*p-bordi e p-cicli*).

Poniamo $\text{Im}\partial_{p+1} = B_p$ e chiamiamo gli elementi di tale insieme *p-bordi*. Pertanto un *p-bordo* b è una p -catena per la quale esiste una $(p+1)$ -catena c tale che $b = \partial_{p+1}c$.

Poniamo inoltre $\text{Ker}\partial_p = Z_p$ e chiamiamo gli elementi di Z_p *p-cicli*. Avremo che un *p-ciclo* c è una p -catena che ha bordo nullo, cioè tale che $\partial_p c = 0$.

Per una migliore comprensione osserviamo la figura 1.2 che rappresenta un complesso simpliciale K composto dai triangoli in grigio e dalle loro facce. Possiamo osservare che in blu è segnata la catena b che è un 1-ciclo che è anche un bordo; mentre le catene in rosso c e in verde c' sono 1-cicli ma non sono bordi.

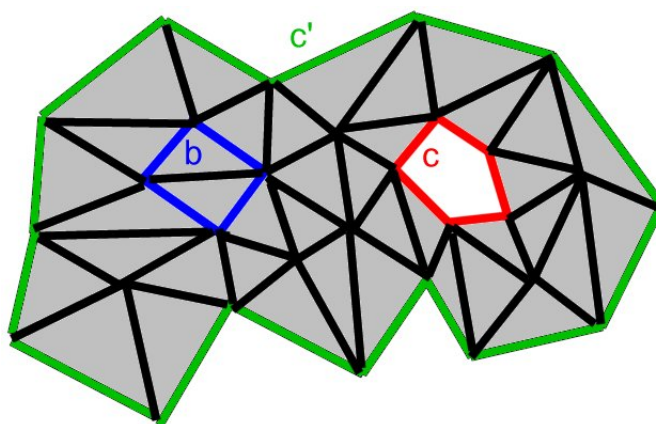


Figura 1.2: Cicli e bordi.

Definizione 1.0.7 (*Gruppo di omologia*).

Secondo quanto definito diciamo dunque che il *k-esimo gruppo di omologia* è il gruppo quoziente ottenuto quozientando il *k-esimo gruppo dei cicli* con il *k-esimo gruppo dei bordi*, con $k \in \mathbb{Z}$, cioè

$$H_k(K) = \frac{Z_k(K)}{B_k(K)}.$$

In altre parole possiamo suddividere ogni gruppo di cicli in classi di cicli che differiscono tra loro dal bordo.

I gruppi di omologia forniscono un linguaggio matematico per descrivere i ‘buchi’ di uno spazio topologico. Tale descrizione avviene non osservando direttamente il buco ma concentrandosi su cosa li circonda.

Definizione 1.0.8 (*Numeri di Betti*).

Sia $k \in \mathbb{Z}$, definiamo *k-esimo numero di Betti* di uno spazio topologico X , indicato con $\beta_k(X)$, la dimensione del *k-esimo spazio vettoriale di omologia* $H_k(X)$, ossia il numero di generatori indipendenti (le classi di omologia dei *k-cicli*) dello spazio. Intuitivamente

$\beta_0(X)$ conta il numero di componenti connesse di cui è composto lo spazio, $\beta_1(X)$ conta il numero di 1-cicli indipendenti (i buchi dello spazio in esame), $\beta_2(X)$ conta i 2-cicli indipendenti di X ossia gli spazi vuoti, ecc...

Per capire meglio riprendiamo l'esempio e calcoliamo i numeri di Betti degli oggetti considerati. Abbiamo già provato che i due oggetti sono omeomorfi, pertanto avranno gli stessi numeri di Betti β_k con $k = 0, 1, 2$, essendo X e Y oggetti in \mathbb{R}^3 .

$$\begin{aligned} \beta_0(X) = \beta_0(Y) = 1 & \quad \text{Entrambi gli oggetti sono infatti composti} \\ & \quad \text{da una sola componente connessa.} \\ \beta_1(X) = \beta_1(Y) = 1 & \quad \text{Entrambi gli oggetti presentano un buco,} \\ & \quad \text{ossia il manico in } X \text{ e il centro della ciambella in } Y. \\ \beta_2(X) = \beta_2(Y) = 0 & \quad \text{Nessuno degli oggetti presenta cavità all'interno.} \end{aligned}$$

Definizione 1.0.9 (Funzione filtrante).

Sia X spazio topologico e sia $f : X \rightarrow \mathbb{R}$ una funzione continua. Chiamiamo la coppia (X, f) *coppia di taglia* e la funzione f *funzione filtrante*.

Definizione 1.0.10 (Insieme di sottolivello).

Data una coppia di taglia (X, f) , e dato $u \in \mathbb{R}$, definiamo *insieme di sottolivello* sotto u l'insieme

$$X_u = \{x \in X \mid f(x) \leq u\}.$$

La funzione filtrante che potremmo usare per andare ad indagare sui nostri due oggetti X e Y , rispettivamente la tazza e la ciambella, potrebbe essere l'altezza. Possiamo considerare cioè una funzione f da X e Y a valori in \mathbb{R} che associ ad ogni punto dei due oggetti la corrispondente ordinata. Supponiamo allora di allineare X e Y come in figura 1.3 in modo che abbiano entrambi inizio a un livello a e fine a un livello c . Supponiamo inoltre che il buco del manico della tazza e il buco della ciambella si aprano e si chiudano allo stesso livello.

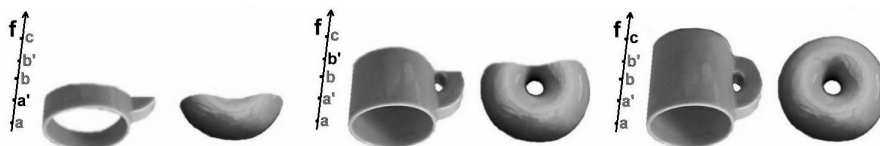


Figura 1.3: *Funzione filtrante e sottolivelli relativi di X e Y .*

Avremo dunque che le due figure, al livello $a - \epsilon$ hanno insieme di livello vuoto, mentre al livello c entrambe le figure sono complete:

$$X_{a-\epsilon} = Y_{a-\epsilon} = \emptyset, \quad X_c = X, \quad Y_c = Y.$$

Definizione 1.0.11 (*Omologia persistente*).

L'aggettivo persistente indica che con questo strumento si potranno osservare le caratteristiche di uno spazio che persistono anche quando si cambia il livello di dettaglio. Data la coppia di taglia (X, f) ed indicato con X_u il sottolivello di X rispetto a f , otteniamo una famiglia $\{X_n\}_{n \in \mathbb{R}}$ di sottospazi di X tali che $\forall u, v \in \mathbb{R}, u < v$ esiste una mappa iniettiva

$$i_{uv} : X_u \rightarrow X_v.$$

La famiglia $\{X_n\}$ dei sottospazi assieme alle mappe di inclusione i_{uv} formano una filtrazione di X indotta da f . Abbiamo inoltre che ogni inclusione i_{uv} induce un omomorfismo tra i relativi gruppi di omologia dei sottolivelli, cioè

$$i_{k(uv)}^* : H_k(X_u) \rightarrow H_k(X_v).$$

Definiamo dunque il k -esimo gruppo di omologia persistente calcolato in un punto (u, v) il gruppo

$$PH_k(u, v) = i_{k(uv)}^*(H_k(X_u)) \subseteq H_k(X_v).$$

Definizione 1.0.12 (*Numeri di Betti persistenti*).

Mantenendo gli elementi come definiti sopra, il numero naturale

$$\beta_k(u, v) = \text{Dim}(PH_k(u, v))$$

viene detto k -esimo numero di Betti persistente (k -PBN) calcolato in (u, v) . Il k -esimo numero di Betti conta dunque il numero di classi di k -cicli di $H_k(X_u)$ che sopravvivono in $H_k(X_v)$. Chiamiamo k -Persistent Betti Number functions (k -PBN functions) le funzioni che associano ad ogni coppia (u, v) il relativo numero di Betti persistente calcolato in quel punto.

Definizione 1.0.13 (*Diagrammi di persistenza*).

Le k -PBN functions sono completamente determinate da alcune linee e punti di discontinuità chiamate *cornerlines* e *cornerpoints*. Le coordinate (u, v) di un cornerpoint rappresentano rispettivamente i livelli di 'nascita' e 'morte' di un generatore e l'ascissa di una cornerline indica il livello di nascita di un generatore che 'non muore mai'. Definiamo inoltre *persistenza* di un cornerpoint la differenza $v - u$ delle coordinate. Il k -diagramma di persistenza $D_k(X, f)$ è formato dalle cornerlines e dai cornerpoints relativi alla corrispondente k -PBN function.

Procediamo ora all'osservazione dei diagrammi di persistenza in grado 1 relativi alle figure X e Y prese in esempio. Consideriamo inizialmente i grafici delle 1-Persistent Betti Number function per i due oggetti. Tali grafici sono rappresentati in 1.4.

Dai grafici, ponendo $c' = c - (\text{spessore del fondo della tazza})$, otteniamo diverse informazioni. Ad esempio relativamente alla 1-PBN function della tazza, il trapezio di vertici

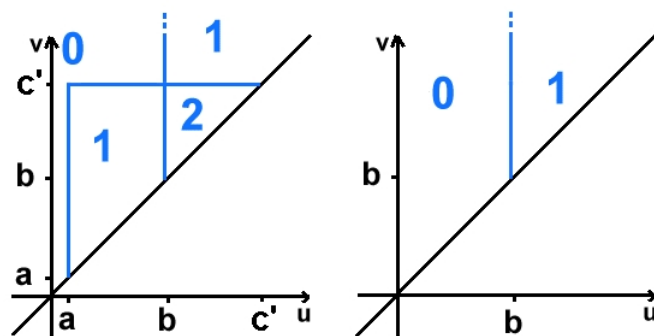


Figura 1.4: 1-PBN function per X e Y .

$(a, a), (b, b), (a, c'), (b, c')$ ci dice che l'1-ciclo presente nei sottolivelli $X_{a'}$, con $a \leq a' < b$, sopravvive nei sottolivelli $X_{b'}$ con $a' < b' < c$ fino a scomparire in $X_{c'}$ e oltre, a causa del fondo della tazza. Il triangolino adiacente ci dice inoltre che gli 1-cicli di un $X_{b'}$ con $b \leq b' < c'$ permangono in tutti i sottolivelli $X_{b''}$ per $b' < b'' < c'$, ma non per $b'' > c'$ in cui rimane solo l'1-ciclo del manico.

Il grafico della ciambella, più semplicemente, ci dice che al livello b nasce un 1-ciclo (il buco della ciambella) che non muore mai.

Osserviamo ora gli 1-diagrammi di persistenza dei due oggetti nella figura 1.5.

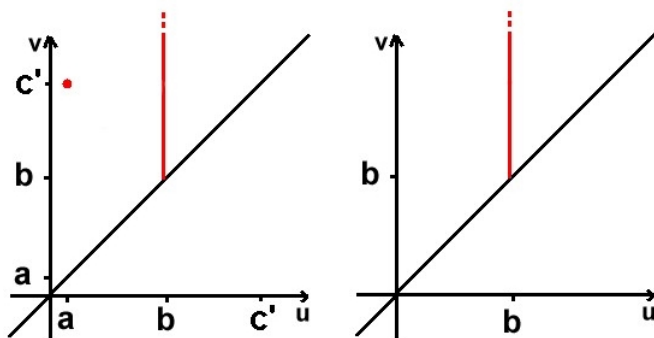


Figura 1.5: Diagrammi di persistenza per X e Y .

I diagrammi ci danno esattamente la stessa informazione già commentata nel caso delle 1-PBN function. Infatti nel caso della tazza osserviamo la presenza di un corner-point in (a, c') indica che un 1-ciclo è nato a livello a e scomparso a livello c' mentre la presenza di una cornerline che comincia dal punto (b, b) dice che esiste un 1-ciclo che nasce al livello b e non muore mai. Analogamente per il diagramma di persistenza della

ciambella osserviamo una cornerline che ha origine in (b, b) che è l'ordinata alla quale corrisponde l'inizio del buco della ciambella.

Una volta computati i diagrammi di persistenza di due oggetti relativi a certe funzioni filtranti possiamo procedere ad un confronto tra i diagrammi per mezzo della *distanza di matching* (o *bottleneck distance*) che mette in relazione i cornerpoints e le cornerlines dei due grafici. Possiamo dire che la distanza di matching è considerata una metrica per i diagrammi di persistenza.

Definizione 1.0.14 (*Distanza di matching*).

Dati due diagrammi di persistenza $D_k(X, f)$ e $D_k(Y, g)$ relativi alle due coppie di taglia (X, f) e (Y, g) , allora $d_{match}(D_k(X, f), D_k(Y, g))$ è il minimo costo di una corrispondenza biunivoca tra in due diagrammi, dove il costo di una corrispondenza biunivoca $\phi : D_k(X, f) \rightarrow D_k(Y, g)$ è dato da

$$\sup_{p \in D_k(X, f)} \|p - \phi(p)\|_\infty.$$

Operativamente possiamo pensare di collegare i cornerpoints di $D_k(X, f)$ con i cornerpoints di $D_k(Y, g)$ o con la loro proiezione sulla diagonale $u = v$, il peso di questo collegamento, come detto, è il sup della distanza L_∞ dei punti collegati. La distanza di matching è il minore di questi costi tra tutti i possibili collegamenti.

Osserviamo per una migliore comprensione l'immagine 1.6 qua di seguito, in cui è ben evidente come viene calcolata la distanza di matching di due diagrammi di persistenza.

Definiamo ora una metrica utile al confronto di due spazi omeomorfi, la *pseudodistanza naturale*. Relativamente a tale argomento, si rimanda a [3].

Definizione 1.0.15 (*Pseudodistanza naturale*).

Si considerino due coppie di taglia (X, f) e (Y, g) con X e Y omeomorfi. Sia inoltre $\varphi : X \rightarrow Y$ un omeomorfismo di cui calcoliamo il peso come $\sup_{x \in X} |g(\varphi(x)) - f(x)|$. Definiamo pseudodistanza naturale di (X, f) e (Y, g) come il minore di questi pesi tra tutti i possibili omeomorfismi, pertanto

$$d((X, f), (Y, g)) = \inf_{\varphi \in \text{Homeo}(X, Y)} \sup_{x \in X} |g(\varphi(x)) - f(x)|.$$

Nonostante questo tipo di distanza sia uno strumento molto potente per osservare le differenze di due spazi omeomorfi, è poco pratica da calcolare in quanto richiede la conoscenza di tutti i possibili omeomorfismi tra i due spazi.

Dati i diagrammi di persistenza relativi a due coppie di taglia (X, f) e (Y, g) , la loro distanza di matching sarà sempre minore rispetto alla pseudodistanza naturale, vale dunque

$$d_{match}(D_k(X, f), D_k(Y, g)) \leq d((X, f), (Y, g)).$$

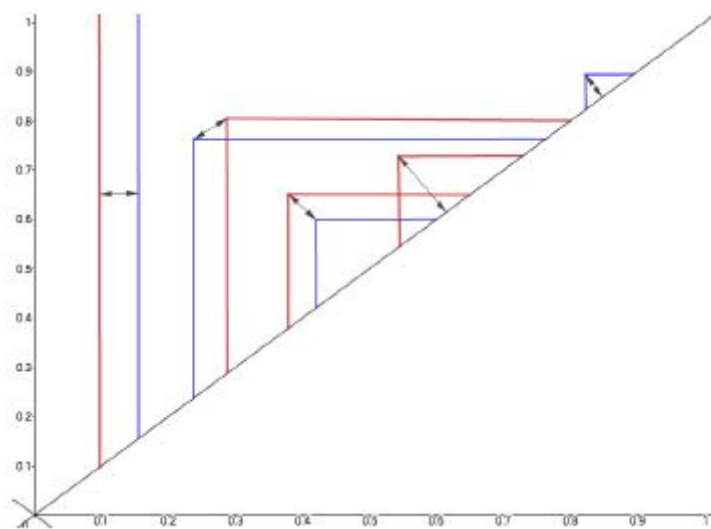


Figura 1.6: *Calcolo distanza di matching*

Definiamo ora il concetto di *grafo di taglia* perché ci servirà in seguito per costruire il diagramma di persistenza relativo a un certo oggetto secondo una funzione filtrante definita.

Definizione 1.0.16 (Grafo di taglia).

Sia (G, ϕ) una coppia in cui G è un grafo mentre ϕ è una funzione definita sui vertici del grafo che assume valori reali. Tale coppia (G, ϕ) viene chiamata grafo di taglia.

Per ulteriori approfondimenti in materia si rimanda al testo [4].

Capitolo 2

Prosodia

Nel presente capitolo introdurremo i concetti relativi alla prosodia e alla fonazione che permettono di comprendere gli aspetti del parlato che saranno materia di indagine. Innanzitutto verrà infatti spiegato come avviene l'atto di emissione vocale e quali sono le caratteristiche del suono prodotto che presenteranno un interesse nel seguito. Verranno poi esposte nozioni basilari di linguistica attraverso le quali vengono spiegati importanti concetti legati alla comprensione e analisi del linguaggio parlato. Successivamente verranno presentate le caratteristiche dello studio dei contorni intonativi: tale studio sta alla base della nostra ricerca. Infine introduciamo un sistema per la trascrizione degli elementi prosodici di interesse. La trattazione fa riferimento, in tutti gli aspetti analizzati, alla lingua italiana.

2.1 Elementi di acustica del parlato

Il nostro corpo, quando decidiamo di emettere un suono, attua un processo detto *fonazione* nel quale l'aria spinta dai polmoni e dal diaframma esce attraverso la trachea verso la laringe dove passa attraverso le corde vocali mettendole in vibrazione. Il suono che la laringe produce è una serie armonica naturale e la variazione della frequenza di vibrazione delle corde vocali, congiuntamente al lavoro della laringe e della bocca, porta all'articolazione della parola e all'emissione coordinata di suoni [7]. In particolare le vocali sono prodotte dalla vibrazione delle corde vocali mentre le consonanti sono prodotte attraverso movimenti delle componenti della cavità orale e vengono percepite come una discontinuità nel corso dell'emissione delle vocali.

Ricordiamo che un suono è descritto da un'onda longitudinale detta *onda acustica* (o onda sonora) che è caratterizzata dai parametri interconnessi di lunghezza d'onda, periodo e frequenza. La *frequenza* è una grandezza fisica, misurata in Hertz (Hz) che esprime il numero delle variazioni di pressione al secondo ed è il parametro che associamo alla

sensazione di altezza di un tono, cioè il parametro che ci permette di distinguere un suono acuto da uno grave. Avremo che maggiore è la frequenza di un suono, maggiore risulta la sensazione di acutezza. Il parlato ha un range di frequenza da 80 a 600 Hz, anche se l'orecchio umano è in grado di udire un più ampio range di frequenze.

Un altro parametro acustico interessante è l'ampiezza dell'onda acustica. Questo parametro ci porta a poter parlare di *volume* di un certo suono. Il volume o intensità è la qualità acustica e psicoacustica associata alla forza di un suono, determinata dalla pressione che l'onda sonora esercita sul timpano, la quale è a sua volta determinata dall'ampiezza della vibrazione e dalla distanza del punto di percezione da quello di emissione del suono. Il volume di un suono è la categoria a cui appartengono tutti gli attributi secondo cui lo ordiniamo da debole a forte. Esso non è propriamente una grandezza fisica tipica del suono, ma piuttosto un'entità soggettiva, pertanto per misurare il volume percepito di un suono si fa spesso riferimento al *livello sonoro*, che viene calcolato in Decibel (Db) secondo una scala logaritmica [8].

I suoni prodotti dalla fonazione del parlato non saranno semplici onde acustiche periodiche sinusoidali ma piuttosto onde complesse che possono essere analizzate solamente a seguito della loro decomposizione in serie di onde sinusoidali semplici che con le loro frequenze ed ampiezze globalmente formano lo spettro acustico dei foni [6]. La decomposizione del segnale complesso nelle sue componenti semplici (detta *analisi in frequenza*) avviene mediante processi che coinvolgono l'utilizzo della trasformata di Fourier e viene visualizzata in uno spettrogramma. Un segnale complesso come quello prodotto dalla fonazione è dunque composto dalla presenza di diverse onde acustiche semplici ciascuna delle quali porta un contributo in termini di frequenza. La frequenza principale del suono viene definita *frequenza fondamentale* (f_0) e tutte le altre frequenze multiple di essa sono dette *armoniche*. La frequenza fondamentale equivale alla frequenza della sua componente più bassa e corrisponde, sul piano percettivo, all'altezza del suono prodotto. Dunque la f_0 è la frequenza fondamentale di vibrazione delle corde vocali.

Nel parlato la frequenza fondamentale della voce varia tra 75 Hz e 500 Hz ed in particolare varia a seconda delle caratteristiche fisiche delle corde vocali di ciascuno, infatti per un uomo sarà compresa tra 70 Hz e 10 Hz, per una donna tra 150 Hz e 250 Hz mentre una voce bianca ha frequenza fondamentale compresa tra i 250 Hz e i 350 Hz [9].

Nella cavità orale, che funziona da cassa di risonanza, alcune frequenze delle armoniche del suono in produzione vengono ridotte, altre eliminate, altre rinforzate (frequenze di risonanza) in relazione alla lunghezza ed alla forma assunta dalla bocca. Le armoniche amplificate danno luogo ad una serie di picchi di frequenza dette *formanti* [6]. Le formanti sono le frequenze di risonanza che consentono il riconoscimento delle vocali. Numerandole in ordine crescente di frequenza, si ha che una specifica vocale è determinata dal rapporto F_1/F_2 tra le prime due formanti del fono considerato.

2.2 Elementi di linguistica

Il termine prosodia (dal greco *prosoidìa*, composto di *pròs* ‘verso’ e *oidè* ‘canto’) indica una branca della linguistica che si occupa dello studio dei fenomeni fonetici come accento, intensità, intonazione, durata del linguaggio parlato. La prosodia si occupa del livello *soprasegmentale* del parlato ossia riguarda ciò che viene detto nel modo in cui viene articolato e si distingue dal livello *segmentale* del parlato che riguarda invece i singoli suoni. Infatti, citando l’enciclopedia Treccani [5], una parola come *casa* può essere resa in forma scritta come una sequenza di quattro caratteri alfabetici distinti e disposti secondo un ordine, i suoni così rappresentati sono considerati segmenti, cioè porzioni di una parola. La loro trascrizione è perciò detta segmentale. Ma esistono alcune caratteristiche del modo in cui questa parola può essere pronunciata che non sono rappresentabili con la trascrizione segmentale: si tratta, ad esempio, della forza, del tono, dell’accento, della durata, del ritmo. Tali caratteristiche, di solito, si estendono su un dominio più ampio dei confini di un singolo segmento sonoro e sono perciò denominate soprasegmentali.

Tra gli aspetti prosodici di un certo suono, vi è il tono. Il tono indica l’altezza, in termini di frequenza, del suono di una certa unità fonetica, in generale ci si riferisce al tono di una certa sillaba mentre si parla di *intonazione* per esprimere la variazione dei toni di un intero enunciato. Dunque parliamo di intonazione per descrivere il movimento melodico di una frase. Possiamo infatti dire che l’intonazione si occupa dell’effetto percettivo di tipo melodico prodotto dalla variazione delle altezze dei suoni durante la pronuncia delle parole.

A seconda del significato che assume l’intonazione nel parlato, distinguiamo tra lingue *tonali* ed *intonative*. In una lingua tonale (come ad esempio il cinese mandarino) la diversa intonazione di una sillaba o di una parola ne cambia l’intero significato. Una lingua intonativa è una lingua in cui l’intonazione permette di comprendere la modalità di una frase (interrogativa, affermativa, ...) e le sue sfumature pragmatiche (sorpresa, dubbio, ...). L’italiano è una lingua intonativa ed è infatti esperienza comune il riconoscimento delle caratteristiche descritte attraverso la corretta analisi dell’intonazione usata.

L’andamento melodico di una frase o di un testo è sintomatico della presenza di accenti che hanno la funzione di mettere in risalto sillabe o parole. L’*accento* infatti è un tratto prosodico che indica la prominenza (o enfasi) intonativa all’interno di una certa emissione fonologica. Tale focus può essere attuato con l’aumento dell’altezza della voce, con l’aumento della durata e con un rafforzamento dell’intensità. Distingueremo tra due tipi di accento, l’accento di parola e l’accento di frase.

L’*accento di parola* è l’accento che causa un rinforzo di una particolare sillaba (più precisamente di una vocale) all’interno di una parola, tale sillaba viene detta sillaba tonica della parola anche se è più corretto parlare di posizione tonica (stressed position) all’interno della parola. Non tutte le parole scritte acquisiscono un accento di parola, solo le parole fonologiche che sono le unità (meglio definite come gruppi clitici [11])

che possono essere lette come una stessa parola scritta. Ad esempio non sono parole fonologiche (e pertanto non prendono accento di parola) gli articoli o le preposizioni semplici. Infatti nella pronuncia di una frase contenente un articolo e un nome, l'insieme di *articolo+nome* viene pronunciato come fosse un'unica parola, ad esempio ne *'Il cane è passato sotto il tavolo'* avremo che *'il cane'* e *'il tavolo'* sono pronunciati come fossero un'unica unità e costituiscono parole fonologiche. Questo tipo di accento porta alla realizzazione di una catalogazione in base alla posizione tonica di una parola, si hanno così parole tronche, piane, sdrucchiole, bisdrucchiole e, raramente, trisdrucchiole. Ad esempio *'Piangere'* ha sillaba tonica *'Pian'* e vocale tonica *'a'* ed è una parola sdrucchiola.

L'*accento di frase* invece è l'accento che richiama l'attenzione su particolari parole all'interno di una frase. Infatti l'accento di frase consiste nella messa in relazione dei singoli accenti di parola così da determinare una struttura prosodica, un vero e proprio schema intonativo caratteristico per ogni atto comunicativo. È possibile riconoscere attraverso esso la modalità (affermativa, interrogativa,..) di una frase [6]. Come detto anche questo accento si realizza attraverso l'aumento di intensità, di durata e di frequenza della sezione accentata. Si parla di *pitch accent* per indicare l'*accento di frequenza* che è la caratteristica in termini, appunto, di frequenza dell'accento di frase. Il *pitch accent* descrive dunque il movimento intonativo (in relazione alla sensazione di altezza) nelle adiacenze di una sillaba che riceve un accento di frase. Parliamo di *nucleo* di una frase per intendere il sintagma posto in genere alla fine della frase intonativa. La posizione nucleare è la più forte all'interno della frase e la sua intonazione, assieme a quella del tono di chiusura (*boundary tone*), contiene la maggior parte dell'informazione sul valore pragmatico della frase. Avendo definito ciò, possiamo ora specificare che esistono accenti di frase di tipo nucleare, prenucleare e postnucleare. Per chiarire questo concetto consideriamo la frase *'Maria mangia la mela'*. Il nucleo della frase è *'la mela'* ma accenti di frase possono cadere anche su *'Maria'* e *'mangia'* che occupano posizioni prenucleari. Infine sicuramente cadrà un accento di frase su *'la mela'* e sarà un accento di frase cosiddetto nucleare. La stessa suddivisione la si ritrova anche in relazione al *pitch accent*, dunque avremo un *nuclear pitch accent* ed eventualmente *pre-nuclear* e *post-nuclear pitch accent*. Questi termini si riferiscono all'accento di frequenza che cade su quelle posizioni. Infine è interessante notare che si definisce anche un *boundary tone*, un tono di frontiera, che descrive il movimento tonale che cade alla fine della frase intonativa.

La realizzazione degli accenti di frase dipende strettamente dalla pronuncia di un particolare oratore. Infatti l'accentare o meno una certa parola di una frase può essere determinato dalla provenienza geografica del soggetto o anche dal semplice contesto in cui la pronuncia. Un altro elemento che causa un cambiamento nel pattern degli accenti di frase è la lunghezza della frase e la velocità con la quale la si pronuncia. Una frase lunga, o comunque pronunciata velocemente, avrà in genere meno parole con accenti di frase. Quindi non è possibile presentare esempi generici che vengano interpretati ugualmente da ogni lettore. Possiamo però dire che in qualsiasi frase sicuramente non

presentano accento di frase gli elementi che si è detto non essere parole fonologiche.

Quindi in conclusione l'accento di parola indica una posizione tonica che può ricevere accenti di frase e precisamente è una prominente potenziale di una sillaba (chiamata tonica) rispetto alle altre sillabe della parola ed è costituita da 3 elementi che diventano osservabili quando la prominente è realizzata: durata maggiore, intensità (volume) maggiore e movimento di frequenza diverso in confronto alle altre sillabe. Si osservi che 'prominente potenziale rispetto alle altre' significa che quando la prominente (il focus) si realizza può realizzarsi solo in una determinata sillaba. Inoltre il movimento di frequenza non è univocamente determinato dalla parola, ma per capire quale sia il movimento di frequenza diverso (di una sillaba rispetto alle altre) da applicare a tale parola, bisogna collocarla in contesto. L'accento di frase invece è la prominente effettivamente realizzata di certe sillabe di una frase (che devono essere almeno toniche) rispetto alle altre. Al massimo una sillaba per parola può ricevere accento di frase e nel caso ciò accada questa deve essere la sua sillaba tonica. Non è però vero che ogni sillaba tonica abbia necessariamente accento di frase. Tale accento si compone di intensità maggiore, durata maggiore e movimenti di frequenza specifici del modo intonativo.

2.3 Analisi del contorno intonativo

Il disegno degli accenti proprio dell'intonazione di una frase si è detto contenere le informazioni relative alla modalità e alla funzione pragmatica della frase stessa. Dunque si è giunti allo studio e alla classificazione dei *contorni intonativi*, dove per contorno intonativo intendiamo l'andamento intonazionale della voce durante l'emissione di un contenuto. Per indagare su questi elementi si è fatto uso delle caratteristiche acustiche proprie del suono descritte precedentemente. In particolare l'andamento dell'intonazione, in quanto, come visto, causa un certo andamento in termini di frequenza, può essere rappresentato dalla variazione della frequenza fondamentale f_0 nel tempo.

Estrarre l'evoluzione della f_0 è possibile utilizzando software che siano in grado di analizzare il suono della voce. Nel nostro percorso di studio e ricerca si è deciso di usare il software Praat creato da Paul Boersma e David Weenink dell'Università di Amsterdam, che consente l'acquisizione, memorizzazione, analisi, e riproduzione del segnale verbale e fornisce inoltre dati numerici con possibilità di valutazioni oggettive delle caratteristiche vocali [10]. Di seguito 2.1 mostriamo la curva disegnata dalla frequenza fondamentale f_0 estrapolata dal software e relativa alla registrazione della frase test 'Maria mangia la mela' prodotta da me e dunque avente le caratteristiche di un oratore proveniente dall'area di Bologna.

La classificazione dei profili intonativi avviene analizzando le caratteristiche comuni a una certa modalità (affermazione, domanda, ...) e un certo intento pragmatico. Si può

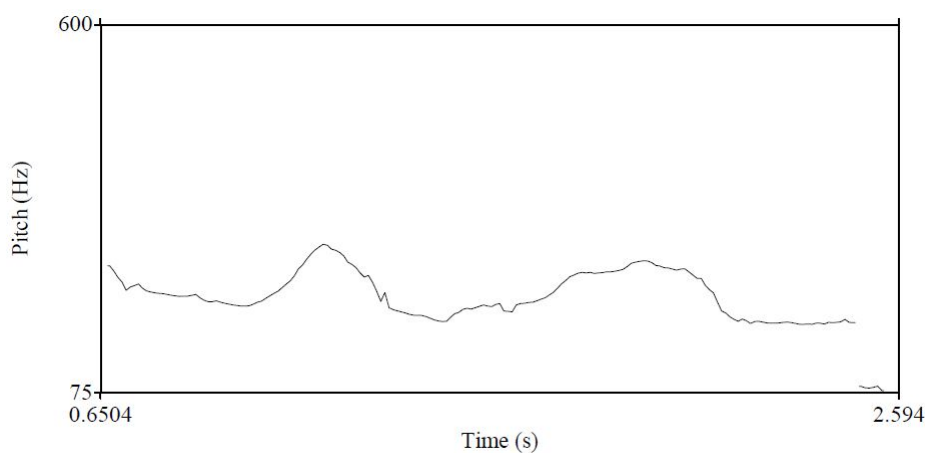


Figura 2.1: *Esemplio curva f_0*

giungere a parlare di *lessico intonazionale* che contiene appunto i diversi contorni, ognuno dei quali ha un particolare significato [11]. Ovviamente data l'innumerabile varietà di possibili profili intonativi ci sono caratteristiche invarianti che non sono centrali nella catalogazione dei contorni. Ad esempio le diverse caratteristiche della voce del oratore (voce maschile, femminile, rauca, ecc) non vanno ad incidere in tal senso. Un'altra caratteristica invariante è il contenuto linguistico associato a un dato profilo e la sua lunghezza; entro certi limiti questi aspetti non vanno ad influire sul tipo di contorno intonativo. Queste invarianze rendono lo studio di questa branca della fonologia ancora più interessante perché si ha a che fare con un tratto fonetico altamente informativo. Inoltre questo studio è valido e possibile per la lingua italiana poiché, come detto, essa è una lingua intonativa e pertanto le singole parole non sono specificate per i toni e dunque non influiscono a livello soprassegmentale.

In questo campo è utile osservare come si legano tra loro le macrounità intonative (sintagmi intonativi) che compongono una frase. La fine di un sintagma intonativo è caratterizzata da vari fenomeni e da pause. Infatti la sillaba finale di un sintagma intonativo è spesso allungata indipendentemente dal tipo di accento e dal contorno ritmico in cui è collocata. L'inizio di un nuovo sintagma intonativo al contrario, ha la tendenza ad essere pronunciato più velocemente. Oltre alle pause o ai vari fenomeni, i limiti dei sintagmi intonativi sono identificabili per il fatto che in loro corrispondenza viene concluso un determinato contorno intonativo. I limiti dei contorni sono infatti spesso caratterizzati da un cambio di altezza tonale simile a quello che caratterizza le sillabe accentate, ma indipendente dalla presenza di accento. All'interno di un contorno intonativo vi è inoltre il fenomeno di declinazione, per cui le sillabe basse non accentate sono più acute all'inizio che alla fine del costituente; la loro frequenza va infatti gradualmente abbassandosi, per rialzarsi poi nel sintagma intonativo successivo [11].

Si osserva in 2.2 il contorno di f_0 in cui si distinguono bene i vari sintagmi intonativi. La registrazione da cui si è estratta tale f_0 si presta bene a questa osservazione poiché si tratta di una lista di nomi: ‘Fragoline, angurie, mandarini, banane, mandorle, amarene, limoni’. Si osserva anche che le parole ‘Banane, mandorle’ e ‘Amarene, limoni’ vengono pronunciate come fossero legate e pertanto generano una sezione di contorno continua.

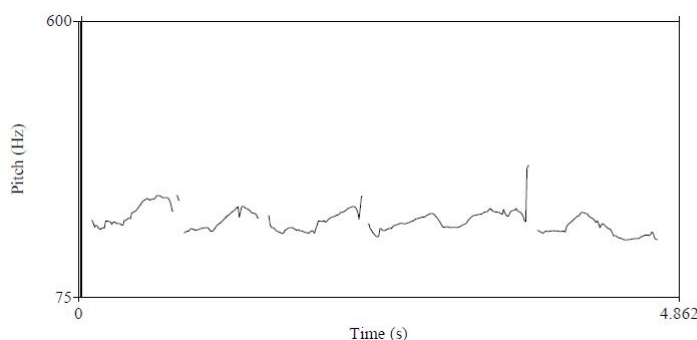


Figura 2.2: *Esempio contorni intonativi.*

Un altro aspetto a cui è necessario prestare attenzione nell’analisi del contorno di f_0 è il modo in cui le consonanti interrompono la regolarità dell’andamento di f_0 . Le consonanti sorde, ad esempio, creano delle discontinuità, dei ‘buchi’, nel profilo di f_0 poiché non corrispondono a un’emissione sonora. Nello studio dei contorni intonativi sarà dunque bene usare frasi test che contengano il minor numero di consonanti ‘problematiche’ al loro interno. Osserviamo il grafico di seguito 2.3 in cui è presente in blu la frequenza fondamentale estratta attraverso Praat della frase ‘Che cosa le regalereste voi?’. La zona evidenziata corrisponde alla discontinuità del profilo intonativo causato dalle consonanti ‘st’ presenti all’interno di ‘regalereste’.

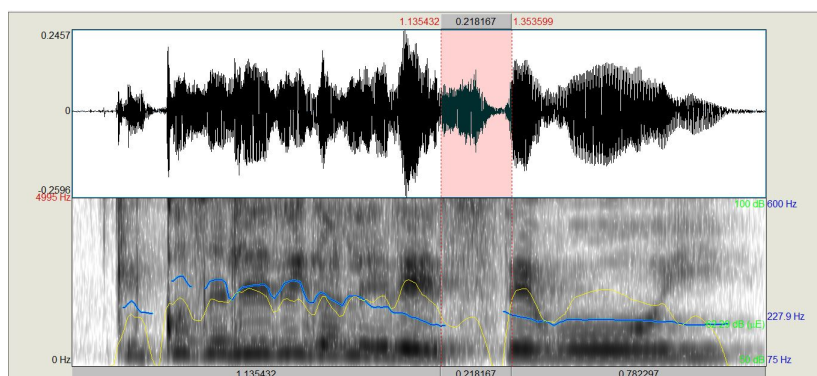


Figura 2.3: *Esempio di discontinuità di f_0 causata da consonanti.*

La materia descritta è oggetto di studi da diversi anni. Questo tipo di analisi, relativamente alla lingua italiana, è stata condotta inizialmente su quello si considerava essere un italiano standard, ma ben presto si è estesa ai vari dialetti presenti. Infatti sebbene esista un'unica lingua italiana scritta è altresì nota la presenza in suolo italiano di numerosi dialetti e cadenze su cui sono stati fatti molti studi linguistici. Relativamente allo studio dei contorni intonativi è importante circoscrivere il dialetto e l'area di provenienza delle frasi in analisi perché l'intonazione ne risente fortemente. Una frase avente un data modalità può essere pronunciata con un certo andamento nel Nord Italia e con andamento completamente diverso nel Sud Italia; non solo, ma studi hanno evidenziato l'esistenza di differenze anche tra pronunce relative a una stessa area geografica. Si è ad esempio osservato che all'interno di una stessa varietà di italiano possono essere presenti due o più tipi di contorni intonativi relativi allo stesso tipo pragmatico di frase. A tal riguardo e per ulteriori approfondimenti si rimanda a [12] e [13].

2.4 ToBI: sistema di annotazione

Ai fini dello studio dei contorni intonativi di cui si è parlato, è necessario introdurre un sistema di scrittura che consenta di descrivere universalmente i diversi profili con una terminologia e una notazione regolamentata. Presentiamo così ToBI, acronimo per Tones and Break Indices, un insieme di convenzioni definite con l'intento di creare un sistema standard di trascrizione dell'intonazione originariamente pensato (1992) per la lingua inglese. Con il tempo sono nate poi delle estensioni che hanno consentito di usare lo stesso sistema di convenzioni anche per altre lingue. Nel 1995 Avesani studia un sistema di annotazione basato su ToBI per la lingua italiana studiando l'italiano fiorentino; più recentemente questo studio è stato ampliato a 13 varietà di italiano e sono state stabilite un certo set di unità fonologiche relative a queste varietà [13]. Diamo dunque una panoramica su come funziona questo sistema di annotazione. Quanto segue è tratto dal sito che descrive le linee guida di tale sistema [14] a cui si rimanda per ulteriori approfondimenti.

ToBI descrive l'andamento del contorno intonativo per mezzo dell'analisi della frequenza fondamentale e si divide in 4 livelli

1. livello di frequenza (*tone tier*),
2. livello ortografico,
3. livello degli indici di pausa (*break-index tier*),
4. livello eterogeneo.

Il livelli di frequenza e degli indici di pausa rappresentano il cuore di questa analisi prosodica. Il livello di frequenza consiste nell'etichettare i movimenti distintivi della frequenza, trascrivendoli come una sequenza di toni alti (H) e bassi (L) e aggiungendo segni diacritici che indichino la loro funzione come parte di un pitch accent o di un tono fondamentale della frase intonativa. Il livello degli indici di pausa evidenzia invece i sintagmi intonativi all'interno della frase etichettando la fine di ogni parola con un indice che esprime la forza di associazione con la parola successiva in una scala da 0 a 4, in cui 0 indica la più forte sensazione di congiungimento mentre 4 indica al contrario la più forte sensazione di disgiunzione.

ToBi ha come obiettivo la descrizione della prosodia mediante dei principi guida. ToBI non si propone per la cattura di ogni possibile aspetto della prosodia, ma è orientato alla trascrizione degli aspetti prosodici di due tipi: struttura prosodica e pattern intonativo. Cioè si pone di descrivere il ritmo delle parole più o meno accentate e il loro raggruppamento in sintagmi intonativi e la sequenza dei vari eventi tonali che possono essere pitch accents, phase accents o boundary tones. ToBI dunque non etichetta ogni andamento della frequenza ma solo quelli che contengono l'informazione utile all'analisi del contorno intonativo. Come si è visto, il contorno intonativo può essere decomposto in una serie di movimenti tonali detti *pitch accents* e *edge tones* la cui interpolazione con linee di congiungimento dà vita al contorno intonativo. I pitch accents sono associati alle sillabe tonali e possono essere suddivisi in *monotonal* o *bitonal* in base al numero di target tonali. Nel caso dei bitonal pitch accents i due toni etichettati con le lettere relative (L o H) sono separati dal segno di addizione (+) ma solo al tono relativo alla sillaba tonica apponiamo anche il segno di asterisco (*). Il tono non asteriscato è detto *leading tone* se lo precede e *trailing tone* se lo segue. Abbiamo due tipi di edge tones: i *phrase accents* e i *boundary tones*. I phrase accents sono da inserirsi tra l'ultimo pitch accent e la fine del suo costituente prosodico e sono contrassegnati da un trattino che segue la relativa lettera. I boundary tones invece indicano la fine della frase intonativa e sono contrassegnati dal segno di percentuale (%) [13].

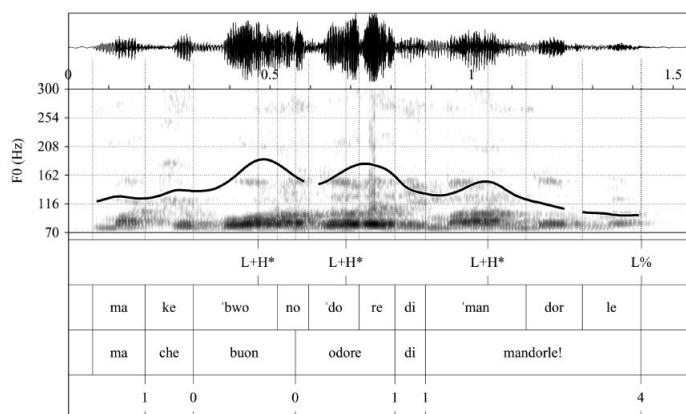
Osserviamo ora un esempio per evidenziare come vengono inseriti gli indici di pausa che, come detto, vanno da 0 a 4. In generale un 4 delimita il confine della frase intonativa, un 3 il confine di una frase intermedia, un 1 di una parola prosodica e lo 0 di un clitico. Pensiamo alla frase *'Maria mangia la mela'* se mettiamo gli indici di pausa otteniamo

Maria 3 mangia 1 la 0 mela 4 .

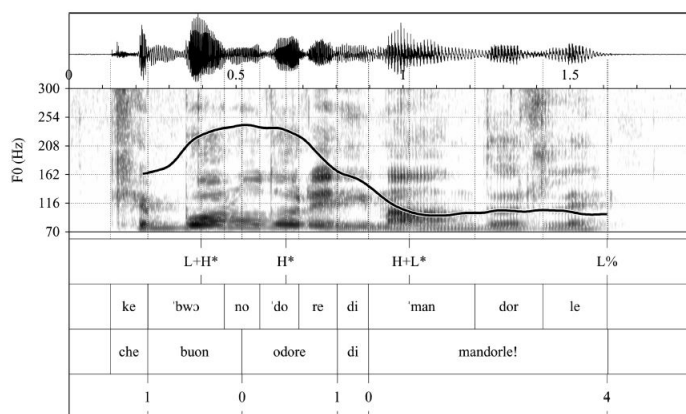
Si ha quindi una frase intonativa con 2 frasi intermedie (*'Maria'*) e (*'mangia la mela'*) e 3 parole prosodiche (*'Maria'*) (*'mangia'*) e (*'la mela'*). Supponiamo di pronunciare la frase dell'esempio in un italiano il più possibile privo di accenti e ben scandito. Allora potremmo ottenere il seguente andamento intonativo etichettato come segue secondo il sistema ToBI:

L+H* H- (L+)H* H+L* L-L%
 Maria 3 mangia 1 la 0 mela 4

Riprendendo anche quanto detto nella sezione precedente sulla differenza del profilo intonativo di una stessa frase a seconda dell'area geografica di provenienza dell'oratore, consideriamo i seguenti grafici, 2.4a e 2.4b, tratti da [12]. La frase che viene pronunciata in entrambe le situazioni è l'esclamazione 'Ma che buon odore di mandorle!' ma in 2.4a lo speaker viene da Torino mentre in 2.4b da Lucca.



(a) *Esclamativa - Torino*



(b) *Esclamativa - Lucca*

Capitolo 3

Descrizione della ricerca

Il presente capitolo si concentra sull'esposizione delle ricerche che hanno motivato il lavoro di questa tesi e sul tipo di analisi effettuate. Si presenteranno infatti parte degli studi di Marco Barone i cui dati sono quelli realmente utilizzati per la nostra sperimentazione. Successivamente verranno descritti gli obiettivi e gli strumenti del progetto di tesi, ponendo l'accento sui dati, sulle tecniche e sugli algoritmi usati. Infine verranno riportati quali tentativi sperimentali sono stati effettuati e per quali motivi.

3.1 Studi precedenti

Gli studi condotti sin ora in ambito dell'analisi dei contorni intonativi hanno portato a una classificazione dei possibili contorni per le varie modalità della frase per ogni varietà linguistica considerata. Come abbiamo detto sono state fatte molte ricerche sulle diverse varietà di italiano presenti e anche ricerche relative a una stessa area geografica per poter meglio indagare le differenze all'interno di una stessa varietà. I risultati ottenuti dalle varie ricerche confluiscono in un unico quadro generale che ci consente di osservare i contorni intonativi che maggiormente ricorrono per una data modalità o tipo di frase nella lingua italiana. Questi studi apportano contributi non solo a livello linguistico, ma concorrono anche a migliorare il sistema di annotazione ToBI per l'italiano.

Il presente progetto di tesi nasce dall'idea di Marco Barone come ampliamento al suo studio. La ricerca originariamente condotta riguarda l'analisi, la classificazione e il confronto dei profili intonativi relativi alle varietà linguistiche presenti nella zona di Pescara. Nella sua tesi di dottorato [13] si analizzano e si mettono a confronto le due varietà linguistiche presenti a cui si fa riferimento come 'Pescara Italian' e come 'Pescarese'. Gli obiettivi di ricerca dello studio citato sono molteplici tra cui il completamento degli studi precedenti fornendo un'analisi comparativa di due sistemi prosodici a contatto: si indaga infatti sulle connessioni tra le due varietà linguistiche e sulle reciproche influenze rimanendo in ambito prosodico. Tale ricerca attua un vero e proprio confronto statistico che

ha inizio con la definizione di un preciso campione di persone scelte considerandone età, estrazione sociale, livello di istruzione, esposizione a contatto prolungato con altre aree linguistiche, area di provenienza dei familiari, ecc. Si è dunque tenuto conto delle caratteristiche socio-linguistiche dei partecipanti al fine di analizzare le informazioni sulla variazione prosodica in funzione della variazione sociale. I partecipanti sono poi stati registrati mentre leggevano certe frasi o mentre rispondevano spontaneamente a situazioni sottoposte dal conduttore così da catturare un ampio numero di profili intonativi compresi quelli appartenenti in modo naturale allo speaker interrogato. Le situazioni usate hanno condotto i partecipanti ad usare determinate parole utili all'analisi più completa del profilo intonativo, ad esempio si sono portate ad usare parole con materiale pre/post tonico per riuscire in seguito a distinguere tra le varie posizioni accentuate. Le frasi sottoposte e analizzate includono diverse modalità (affermazione, domanda, esclamazione, imperativo, vocativo) manifestate in diverse forme sintattiche (domande a risposta breve sì/no, domande introdotte da pronomi interrogativi) e con diverse sfumature pragmatiche (richiesta cortese, comando, invito, offerta, ecc). Si è inoltre posta l'attenzione sul tipo di focus che lo speaker attuava sulla frase e sui possibili preconcetti eventualmente presenti (ricerca di conferma, affermazione banalmente ovvia, ecc.).

Attraverso la ricerca di M. Barone si è ottenuta una classificazione dei possibili contorni intonativi relativi alle varietà di italiano considerate in relazione a 57 diverse situazioni sottoposte ai partecipanti che hanno reso possibile l'analisi di un vasto range di tipologie di frase.

Osserviamo in particolare alcuni dei risultati così ottenuti. Ad esempio il profilo $H+L^* L\%$, realizzato graficamente come descritto dalle immagini 3.1 (in cui la zona grigia indica la posizione in cui cade l'accento),

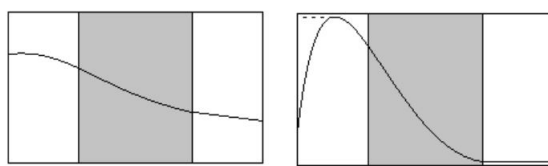


Figura 3.1: *Realizzazioni grafiche del profilo $H+L^* L\%$.*

si può ritrovare nelle seguenti tipologie di frasi:

- **Esclusivamente in assenza di H pretonico:** in ultima posizione nelle liste dichiarative, nelle liste di domande introdotte da pronomi interrogativi (wh-questions o domande parziali), in domande disgiuntive, in liste disgiuntive.
- **In assenza o presenza di H pretonico:** in dichiarative generiche (che non presentano un punto di focus, dette broad-focus statement).

- **Esclusivamente in presenza di H pretonico:** in dichiarative ovvie, esclamazioni, domande polari con ricerca di conferma o esclamative, domande parziali con ricerca di conferma, esclamative o imperative, comandi imperativi, vocative, risposte contrastive o non contrastive a domande parziali.

Il profilo $H+L^* LH\%$ invece, realizzato graficamente come descritto in 3.2

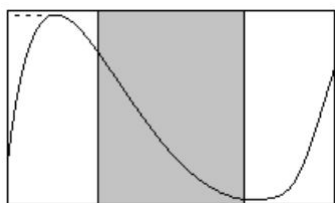


Figura 3.2: *Realizzazione grafica del profilo $H+L^* LH\%$.*

lo ritroviamo solo in presenza di H pretonico in domande polari con ricerca di conferma o di informazioni.

Inoltre viene presentato un confronto sulle diverse realizzazioni di contorni intonativi per una stessa tipologia di frase. Ad esempio viene osservato che le dichiarazioni contrastive possono attuarsi sia con un profilo $H+L^* L\%$ con H pretonico, sia con $H^*+L L\%$ indifferentemente dalla presenza o assenza di H pretonico.

Ancora, le esclamazioni a Pescara sono espresse per mezzo del contorno $H^*+L L\%$, con o senza H pretonico, oppure dal profilo $L^*+H L\%$ senza H pretonico.

Come ultimo esempio consideriamo le domande polari con ricerca di informazione. Questa tipologia di domanda a Pescara può essere realizzata in tre modi diversi, ossia attraverso $H+L^* L\%$, $L^*+H L\%$, e $L^*+H H\%$.

Possiamo concludere rimandando a [13] per consultare interamente gli interessanti risultati ottenuti che non citeremo nel presente elaborato. Nel nostro studio ci limiteremo a utilizzare frasi che tratte dalla ricerca citata di cui già conosciamo la classificazione.

3.2 Obiettivi della ricerca

Le ricerche effettuate in questo progetto di tesi mirano a mettere in relazione le tecniche di omologia persistente e il campo dell'analisi dei contorni intonativi. Si è pensato infatti che fosse possibile attuare un'indagine di tipo topologico sui profili intonativi dal momento che possiamo considerare il profilo della frequenza fondamentale f_0 come una curva o comunque come un elemento grafico che può essere analizzato geometricamente.

Questo tipo di ricerca potrebbe dare quindi un contributo allo studio linguistico dei contorni intonativi per mezzo della contaminazione tra queste due discipline. Le tecniche di omologia persistente sono tra l'altro già state utilizzate in ambito linguistico e dell'analisi di contenuti audio, il che ci suggerisce che la collaborazione tra topologia e fonetica possa portare ad interessanti risultati.

Il punto centrale di questa ricerca sarà trovare e testare delle funzioni filtranti che ci permettano di ricavare l'informazione prosodica associata al contorno intonativo esaminato. Le funzioni filtranti dovranno essere funzioni che siano in grado di evidenziare le caratteristiche comuni e rilevanti di profili facenti parte della stessa tipologia intonativa. Un buon set di funzioni filtranti consentirà cioè di distinguere tra le varie modalità e pragmaticità delle frasi ma non si dovrà curare, ad esempio, della maggiore o minore altezza di un fono (differenza tra voce maschile e femminile) o dei possibili difetti di pronuncia dello speaker. Vogliamo dunque che queste funzioni siano 'cieche' agli elementi invarianti dell'analisi dei contorni intonativi, ma che evidenzino prontamente gli aspetti cruciali. Per ottenere questo in genere sono necessarie più funzioni filtranti che agiscono sullo stesso fono. Ognuna delle funzioni create andrà a catturare un dato aspetto e componendo le varie informazioni ricavate, potremo raggiungere il nostro scopo. Questo progetto di tesi mira ad individuare alcune basilari funzioni filtranti e a osservare i risultati ottenuti attraverso esse.

3.3 Strumenti della ricerca

La ricerca condotta per questo progetto di tesi si è deciso che operasse considerando un set ristretto di dati e non tutto il campione raccolto dalle ricerche precedenti. Pertanto abbiamo lavorato su 56 tracce audio raccolte da M. Barone suddivise in tre gruppi ciascuno dei quali contiene campioni della stessa frase (e dunque della stessa modalità) emessa da speaker differenti per età, sesso, istruzione, ma appartenenti alla stessa varietà linguistica di italiano (italiano di Pescara). In particolare abbiamo analizzato le risposte a 4 contesti differenti, due dei quali portano all'emissione di una stessa tipologia intonativa e che dunque rende lecita la suddivisione in tre macrogruppi di cui parlavamo. Il primo contesto in esame richiede l'emissione da parte dell'oratore di una lista di elementi, in questo caso, di frutta. Negli audio analizzati la maggior parte degli elementi viene pronunciata come $L+H^*$ a cui segue un $H\%$ come boundary tone. L'ultimo elemento invece, in genere è pronunciato come $H+L^* L\%$ e si osserva che il penultimo elemento ha sempre una 'salita intonazionale' più evidente. Di seguito 3.3 il grafico della frequenza fondamentale (in blu) e del livello di decibel (in giallo) di un file audio appartenente a questa categoria.

Il secondo contesto osservato porta all'emissione di una domanda interrogativa parziale introdotta da un pronome interrogativo (Wh-question). Nel nostro caso la domanda

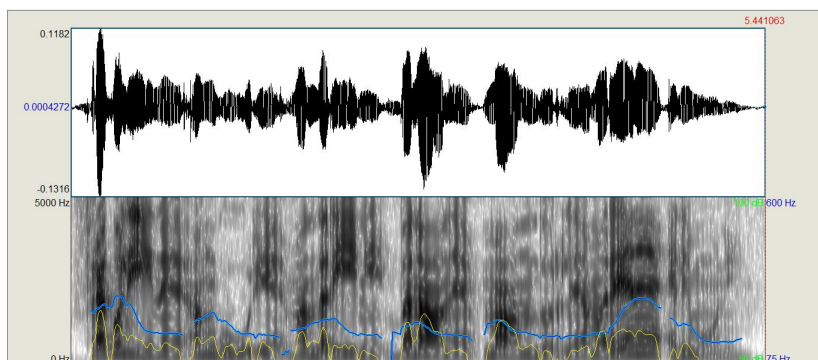


Figura 3.3: Esempio di lista relativo al file *PeIS4DicListaL80*. L'audio corrisponde alla pronuncia di 'Fragoline, angurie, mandarini, banane, mandorle, amarene e limoni'.

è 'Cosa le regalerebbero?' ed è pronunciata con un pitch accent L^*+H . È importante osservare che queste domande presentano una discesa intonativa e poi una 'intenzione di salita'. Il grafico della f_0 e del volume di un file di questa categoria è osservabile in 3.4.

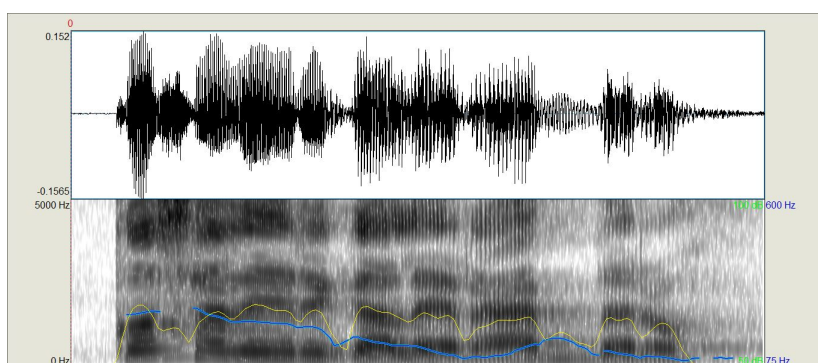


Figura 3.4: Esempio di domanda interrogativa parziale relativa al file *PeIS32IntParzL5*. L'audio corrisponde alla pronuncia di 'Cosa le regalerebbero?'.

Infine si sono esaminati due contesti in cui sono state ottenute come risposte delle domande polari (Yes - No questions) rispettivamente 'Avete dei mandarini?' e 'Avete delle mandorle?'. Negli audio utilizzati sono accentati sia il verbo 'Avete' con $L+H^* L-$, che la parola 'mandorle'/'mandarini' con $H^*+L LH\%$. In 3.5 la rappresentazione grafica della frequenza e del volume di un file audio di questa categoria.

Per completezza riportiamo in 3.1 l'elenco dei file audio con i nomi originali e le sigle che useremo in questa ricerca. Ogni file è stato etichettato originariamente secondo questo formato che permette di riconoscerne subito le caratteristiche.

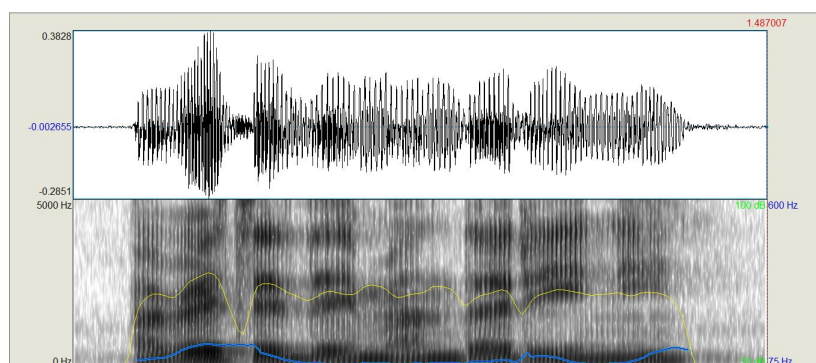


Figura 3.5: Esempio di domanda interrogativa polare relativa al file *PeIS13IntTotNeuL7*. L'audio corrisponde alla pronuncia di 'Avete dei mandarini?'.

'PE(Pescara) - IS(Inchiesta in situazione) - numero contesto - descrizione tipo pragmatico - Letta vs Spontanea - Numero d'ordine dello speaker - Eventuale sollecitazione orale (0) - eventuale seconda ripetizione (bis)'

Successivamente, per brevità, ho eliminato le informazioni ridondanti o non necessarie ai nostri fini. La nuova etichetta segue il formato 'Lista (L), Wh-question (W), Yes/No question(Y o YY)- Letta (L) o spontanea (S) - Numero dello speaker - Eventuale sollecitazione orale (0) - eventuale seconda ripetizione (bis)'

Le sigle usate ci permetteranno di comprendere a prima vista se i dati elaborati alla fine della ricerca sono stati classificati in modo coerente.

Tabella 3.1: *Elenco dati*

Tipologia	Etichetta completa	Sigla
Liste 18 elementi	PeIS4DicListaL2	LL2
	PeIS4DicListaL2bis	LL2b
	PeIS4DicListaL3	LL3
	PeIS4DicListaL6	LL6
	PeIS4DicListaL8	LL8
	PeIS4DicListaL2O	LL2O
	PeIS4DicListaL2Obis	LL2Ob
	PeIS4DicListaL4O	LL4O
	PeIS4DicListaL6O	LL6O
	PeIS4DicListaL8O	LL8O
	PeIS4DicListaS3	LS3
	PeIS4DicListaS7	LS7
	PeIS4DicListaS8	LS8

	PeIS4DicListaS10	LS10
	PeIS4DicListaS30	LS30
	PeIS4DicListaS40	LS40
	PeIS4DicListaS60	LS60
	PeIS4DicListaS80	LS80
Domande Parziali (Wh) 19 elementi	PeIS32IntParzL3	WL3
	PeIS32IntParzL3O	WL3O
	PeIS32IntParzL4	WL4
	PeIS32IntParzL4Obis	WL4Ob
	PeIS32IntParzL5	WL5
	PeIS32IntParzL5O	WL5O
	PeIS32IntParzL6	WL6
	PeIS32IntParzL6bis	WL6b
	PeIS32IntParzL6O	WL6O
	PeIS32IntParzL7	WL7
	PeIS32IntParzL7O	WL7O
	PeIS32IntParzS1	WS1
	PeIS32IntParzS1O	WS1O
	PeIS32IntParzS2	WS2
	PeIS32IntParzS4	WS4
	PeIS32IntParzS4O	WS4O
	PeIS32IntParzS5	WS5
	PeIS32IntParzS6	WS6
	PeIS32IntParzS6bis	WS6b
Domande polari (Yes/no) 19 elementi	PeIS13IntTotNeuL2O	YL2O
	PeIS13IntTotNeuL3O	YL3O
	PeIS13IntTotNeuL4O	YL4O
	PeIS13IntTotNeuL5O	YL5O
	PeIS13IntTotNeuL6O	YL6O
	PeIS13IntTotNeuL7	YL7
	PeIS13IntTotNeuL7bis	YL7b
	PeIS13IntTotNeuL7O	YL7O
	PeIS13IntTotNeuS1bis	YS1b
	PeIS13IntTotNeuS3	YS3
	PeIS13IntTotNeuS3O	YS3O
	PeIS13IntTotNeuS4O	YS4O
	PeIS13IntTotNeuS5	YS5
	PeIS13IntTotNeuS7	YS7
	PeIS13IntTotNeuS7bis	YS7b
	PeIS13IntTotNeuS7O	YS7O

	PeIS14IntTotNeuL3O	YYL3O
	PeIS14IntTotNeuS1	YYS1
	PeIS14IntTotNeuS1O	YYS1O

Per elaborare i dati ed ottenere il grafico di persistenza e la matrice delle distanze si è usato il codice creato dal Dott. Ivan Tomba che prende in input i grafi di taglia relativi alle funzioni filtranti usate e ritorna il relativo grafico di persistenza e la matrice delle distanze dei diagrammi tra loro relativamente a una data funzione filtrante. La matrice delle distanze è computata utilizzando la distanza di bottle-neck e permette di operare un confronto tra tutti i 56 file in esame.

Nel corso della ricerca otterremo dunque una matrice di distanza per ogni funzione filtrante usata e la matrice avrà dimensione $N \times N$, con $N =$ numero dei dati confrontati (nel nostro caso 56). Definita D tale matrice, l'elemento $d_{i,j}$ esprime la distanza tra il dato di posizione j e quello di posizione i ; per questo motivo la matrice D sarà simmetrica con diagonale nulla.

Una volta costruite le matrici contenenti le distanze dei vari dati tra loro, si è deciso di usare un *dendrogramma* per la visualizzazione grafica utile alla comprensione e analisi dei dati elaborati. Un dendrogramma è il risultato grafico di un processo di classificazione (*cluster analysis*) il cui obiettivo è quello di individuare la struttura aggregativa dei dati. La cluster analysis è una tecnica di analisi attraverso la quale è possibile raggruppare i singoli dati in gruppi in modo da raccogliere le unità simili tra loro in insiemi la cui unione ridà la totalità degli elementi considerati. La lontananza tra gli elementi e tra i raggruppamenti è definita attraverso le misure di distanza precedentemente scelte per osservare la disposizione dei dati. Nel nostro caso le distanze degli elementi l'uno dall'altro sono raccolte nelle matrici ottenute nei passaggi precedenti.

Il tipo di algoritmo usato per la creazione dei dendrogrammi è di tipo gerarchico aggregativo, cioè si è usato un algoritmo che giunge alla classificazione finale per passi successivi, creando gruppi aggregando le unità attorno a centri prestabiliti. Questo tipo di procedimento quindi suddivide i dati in gruppi di dati sempre più simili tra loro, fino a giungere al dato stesso. Esistono diversi metodi di classificazione gerarchica distinti secondo il tipo di distanza tra i gruppi scelta [15]:

- Metodo del legame medio (average): i gruppi si formano man mano che varia la distanza media tra le unità in essi comprese.
- Metodo del centroide (centroid): i gruppi si formano man mano che varia la distanza da un centroide ottenuto come media delle unità del gruppo.
- Metodo della mediana (median): i gruppi si formano man mano che varia la distanza dalla mediana del gruppo.

- Metodo del legame singolo (single): i gruppi si formano aggiungendo di volta in volta l'unità più vicina ad una delle unità del gruppo già formato.
- Metodo del legame completo (complete): i gruppi si formano aggiungendo di volta in volta l'unità più vicina a tutte le unità già presenti nel gruppo.
- Metodo del legame pesato (weighted) : i gruppi si formano man mano che varia la distanza media pesata tra le unità.
- Metodo di Ward (Ward): i gruppi si formano man mano che varia la distanza minima quadrata tra le unità.

Una volta terminato il processo che porta alla creazione del dendrogramma, è necessario definire un *livello di taglio* della suddivisione per poter determinare i gruppi della partizione finale.

È importante osservare che gli algoritmi di classificazione gerarchica sono sensibili all'ordine in cui immettiamo i dati, pertanto non è detto che scambiando due posizioni degli elementi nella matrice delle distanze, si ottenga esattamente lo stesso dendrogramma. Per superare questa problematica si è deciso di usare la funzione MatLab di `OptimalLeafOrder` che prende in input l'albero creato dal metodo di classificazione e la matrice delle distanze e ritorna l'ordine ottimale degli elementi con cui andremo a costruire il dendrogramma. A questo proposito si possono compiere ulteriori studi ma per l'interesse di questo studio è sufficiente avere un modo unico attraverso il quale leggere tutti i risultati in modo che essi siano confrontabili.

Come anticipato l'algoritmo usato per la creazione dei dendrogrammi è scritto su MatLab e usa alcune funzioni della **Statistics and Machine Learning Toolbox**. Si rimanda alle appendici per l'algoritmo completo.

3.4 Analisi effettuate

Durante l'analisi dei dati precedentemente descritti ho cercato di definire le funzioni filtranti che maggiormente catturavano le informazioni di nostro interesse. Uno dei concetti chiave su cui mi sono concentrata è stato il cercare di ottenere una funzione che catturasse l'*andamento* della curva di frequenza. Infatti è anche attraverso l'osservazione dell'andamento che è possibile ottenere una classificazione ToBI dell'enunciato. A tal proposito si è ritenuto utile utilizzare come prima funzione filtrante la funzione di altezza di f_0 , ossia la funzione che ad ogni istante temporale (del campionamento discreto) associa la corrispondente frequenza in Hertz. Una problematica subito riscontrata è relativa alla discontinuità del dato che andiamo a studiare. Infatti come già detto la curva f_0 di nostro interesse non è una curva continua a causa delle pause o degli effetti della pronuncia delle consonanti. Per superare questa problematica si è pensato di effettuare

un'interpolazione dei dati con lo scopo di 'chiudere' la discontinuità in modo coerente con l'andamento della curva. In questo modo però avremmo aggiunto alla curva una informazione artificiale, andando a creare una parte di curva (con un suo andamento) che non è relativa a nessuna emissione sonora. Visto che i cambiamenti della curva f_0 che caratterizzano una tipologia intonativa sono a volte molto piccoli, si è ritenuto opportuno non procedere in questa direzione. Dunque si è pensato di togliere le discontinuità semplicemente collegando in successione l'ultimo valore prima del salto con quello immediatamente successivo. In questo modo non avremmo aggiunto ulteriori informazioni artificiali al file. Per ottenere i valori di questa funzione filtranti ci siamo affidati al software Praat che è in grado di estrarre la lista dei valori della f_0 per ogni istante di tempo campionato. Praat di fronte a una discontinuità, cioè di fronte all'assenza di valori di f_0 per i motivi sopracitati, attribuisce un valore `-undefined-`. La creazione dei file da inserire per ottenere le matrici di distanza è avvenuta dunque inserendo in ordine i valori f_0 registrati da Praat eliminando, man mano che si presentavano, i valori `-undefined-`. Di seguito 3.6 vediamo la sovrapposizione di tre curve f_0 relative ai tre tipi di file audio descritti e osservati in esempio nelle sezioni precedenti.

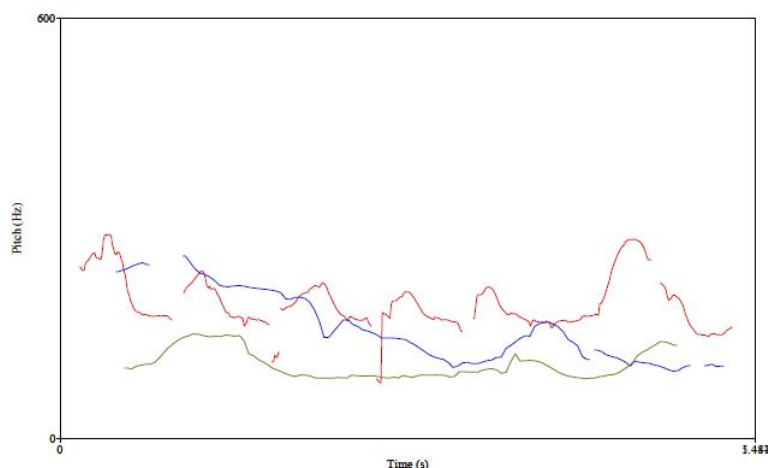


Figura 3.6: Curve f_0 . In rosso la curva della lista PeIS4DicListaL8O, in blu la curva della domanda parziale PeIS32IntParzL5 e in verde la curva della domanda polare PeIS13IntTotNeuL7.

Continuando a ragionare in termini di andamento la seconda funzione filtrante che si è ritenuto interessante utilizzare è la funzione di derivata di f_0 . Per computare la derivata della precedente funzione di f_0 mi sono avvalsa del software MatLab su cui ho inserito i dati discreti di f_0 e ho implementato l'algoritmo di derivata attraverso la funzione

$$f'(x_i) = \frac{-f(x_{i+2}) + 8f(x_{i+1}) - 8f(x_{i-1}) + f(x_{i-2})}{12}.$$

Successivamente ci si è soffermati a considerare il tipo di legame che intercorre tra intonazione e volume di emissione. Come già precedentemente citato esiste una relazione tra il modo di realizzare un accento e l'aumento del volume del segmento accentato, pertanto si è ritenuto che anche il volume di un enunciato costituisca un'interessante informazione. La terza funzione filtrante presa in considerazione è dunque l'altezza, in termini di Decibel, del livello sonoro (volume) del file in esame. Per estrarre l'informazione dai dati abbiamo nuovamente utilizzato gli strumenti che il software Praat mette a disposizione ed in particolare abbiamo potuto estrarre la lista dei valori (in Decibel) relativi alla frase selezionata. Si è proceduto a tagliare la lista dei valori del volume in corrispondenza con gli istanti di inizio e fine dell'emissione sonora così da non includere i valori di livello sonoro che sarebbero corrisposti al cosiddetto rumore bianco. Non si è invece proceduto a 'pulire' i valori di volume dalle pause del parlato a causa della natura vibratoria del suono. Infatti le discontinuità della f_0 nei file in esame non sono mai tali da causare un reale silenzio. In quei pochi istanti cioè, la voce continua a vibrare nell'aria e dunque si registra ancora un livello sonoro consistente e non assimilabile con il livello sonoro del silenzio che vale in media circa 20 Db per il luogo dove si sono eseguite le registrazioni. Giunti a questo punto si è ritenuto opportuno iniziare le fasi di indagine per mezzo dell'algoritmo per la creazione della matrici di distanza relative a queste tre funzioni filtranti per vedere se con soli questi dati basilari si è in grado di ottenere una classificazione che rispecchi la realtà.

Un secondo tipo di analisi che si è svolta consiste nel sommare i contributi ottenuti dalle singole funzioni filtranti per osservare se le informazioni combinate assieme creino un risultato più completo. In particolare per costruire questa somma si è proceduto a sommare i contributi normalizzati delle varie matrici di distanza ottenute. Avremo dunque una nuova matrice di distanza S i cui elementi $s_{i,j}$ sono dati dalla somma degli elementi normalizzati corrispondenti alla posizione i, j delle altre tre matrici di distanza relative a ciascuna funzione filtrante considerata. La normalizzazione effettuata in questo contesto serve per fare in modo che i valori delle varie matrici apportino lo stesso contributo, in termini di peso, nella somma finale. Infatti possiamo osservare che mediamente i valori della matrice di distanza relativa alla frequenza fondamentale sono mediamente più alti rispetto quelli relativi alla funzione di volume. Quindi per normalizzare i valori di una matrice si è calcolato il valore massimo e si è proceduto a dividere tutti gli elementi per quel valore, cioè:

$$dnorm_{i,j} = \frac{d_{i,j}}{\max_{i,j} d_{i,j}}, \forall i, j < 56.$$

La matrice somma è stata dunque costruita a partire da questi elementi normalizzati e l'iterazione utilizzata è consultabile in appendice.

Capitolo 4

Risultati

Nel presente capitolo si andranno ad esporre i risultati ottenuti dalle prove sperimentali ed i commenti a riguardo soffermandosi su quali siano stati i successi ottenuti e sul perché si siano osservati dei dati problematici. Verranno infine presentati gli spunti per un eventuale lavoro futuro in questo ambito.

4.1 Risultati ottenuti

Una volta raccolti i valori delle funzioni filtranti in file di testo, il programma per il calcolo dei diagrammi di persistenza ci ha fornito i diagrammi di persistenza e le matrici di distanza relative a ogni funzione filtrante. Idealmente vorremmo che inserendo le matrici di distanza ottenute nel programma di clustering, i dendrogrammi mostrassero tre cluster che suddividessero i tre tipi di dato associandoli per similarità. Per vedere se l'analisi ha riportato dei risultati corretti andremo dunque a vedere quanti gruppi di associazioni corrette ci sono nei dendrogrammi e quanti dati tali gruppi riescono a riunire: andremo cioè ad osservare quanti file dello stesso tipo sono raggruppati insieme in uno stesso cluster. In generale possiamo dire che non ci sono risultati eccessivamente accurati tali da consentire di vedere a colpo d'occhio la classificazione dei dati, in ogni caso da quanto ottenuto è comunque possibile trarre interessanti considerazioni e tracciare una profilo per future indagini.

Innanzitutto osserviamo che tra i metodi di linkage utilizzati quelli che portano a risultati migliori sono i metodi `complete` e `ward` a cui seguono, nell'ordine, i metodi `average`, `weighted` e `single`. Si osserva che non sono stati usati tutti i possibili metodi di linkage conosciuti, infatti i metodi `centroid` e `median` non sono stati ritenuti adatti dal programma per analizzare i tipi di dato in esame. Se messi in esecuzione MatLab mostra un messaggio di warning `'Warning: Non-monotonic cluster tree'`. La documentazione di Matlab [16] spiega che i metodi `centroid` e `median` possono creare un albero non

monotono quando la distanza tra l'unione di due clusters (\mathbf{r} ed \mathbf{s}) con un terzo cluster è minore della distanza dei due cluster \mathbf{r} ed \mathbf{s} . In tal caso otterremo un dendrogramma del tipo raffigurato in 4.1. Mentre gli altri metodi sono maggiormente intuitivi, ricordiamo

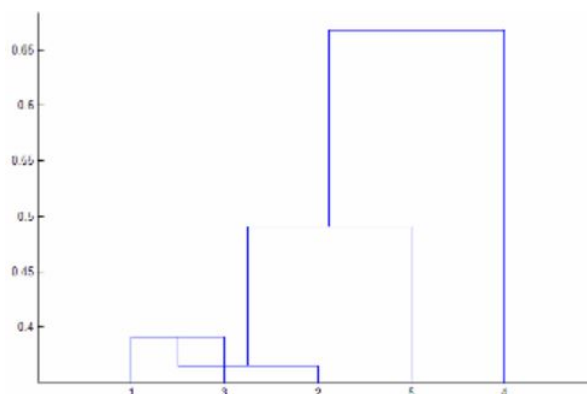


Figura 4.1: Esempio di dendrogramma creato da un albero non monotono.

che il metodo di linkage Ward (1963) considera tutte le possibili coppie di gruppi ad ogni stadio e procede alla fusione dei due che minimizzano l'incremento della devianza totale dal centroide del nuovo gruppo. All'inizio del processo, quando ogni cluster è formato da un solo elemento, la devianza interna è zero, ed infine quando due entità si fondono in un singolo cluster si introduce un grado di variabilità destinato a crescere in funzione della numerosità del gruppo stesso.

Un'altra considerazione di carattere generale che possiamo trarre è che le funzioni filtranti che contengono più informazione in termini di associazioni tra tipi di dato della stessa categoria sono la funzione di altezza del volume e la funzione della frequenza fondamentale. I dendrogrammi relativi alle sole funzioni citate infatti presentano un maggior numero di raggruppamenti di audio corretti rispetto al dendrogramma relativo alla funzione di derivata della f_0 . D'altra parte i dendrogrammi relativi alla funzione filtrante del volume sono quelli che presentano in assoluto il maggior numero di associazioni corrette. È interessante quindi notare (facendo riferimento ai soli risultati ottenuti in questo lavoro di tesi) che l'andamento del volume di emissione dell'enunciato contiene effettivamente una certa quantità di informazione relativa alla modalità della frase. La funzione filtrante che descrive la derivata di f_0 invece sembra non riuscire a raggruppare assieme la stessa quantità di enunciati delle altre funzioni filtranti ma, nonostante ciò, svolge un ruolo fondamentale nell'evidenziare file audio contenenti grosse diversità o particolari casistiche come vedremo in seguito.

La considerazione fatta non si basa solamente sull'osservazione dei risultati ottenuti relativamente alle diverse funzioni filtranti ma è intrinseca nell'informazione contenuta nei diagrammi di persistenza ricavati. L'osservazione di tali diagrammi è stata molto utile

al fine di comprendere a fondo i risultati racchiusi nei dendrogrammi e di correggere eventuali errori avvenuti del corso della ricerca. Come introdotto nel primo capitolo, i diagrammi di persistenza sono composti da un insieme di punti e linee detti *cornerpoints* e *cornerlines* che indicano la ‘nascita’ e ‘morte’ di componenti connesse. In particolare, nei diagrammi ricavati dai nostri dati, con le funzioni filtranti discusse, avremo diversi *cornerpoints* e una sola *cornerline* che indica la ‘nascita’ nella sua totalità della curva analizzata. Quando andiamo a confrontare i diagrammi di persistenza ricavati secondo una stessa funzione ma relativi a diversi dati, la distanza di *bottle-neck* usata va a calcolare la distanza tra i *matching* secondo la sua definizione e per fare ciò associa necessariamente le *cornerlines* dei diagrammi. Nel caso le due *cornerlines* fossero vicine, la loro posizione non va ad influire sull’esito della distanza; ma nel caso le due *cornerlines* fossero particolarmente distanti allora la distanza di *bottle-neck* avrà valore pari alla loro distanza. Questa seconda eventualità si è presentata nei dati analizzati dandoci modo di osservare e comprendere il perché di una clusterizzazione insolita per certi file.

Una funzione filtrante che consente la corretta associazione di un buon numero di dati significa dunque che crea diagrammi di persistenza i cui *cornerpoints* e *cornerlines* sono disposti in maniera simile per tipologie di dati simili. Dunque si è attuato un confronto grafico in questa direzione che ci permette di osservare l’assetto dei *cornerpoints* e *cornerlines* relativo a una stessa funzione filtrante sia per le tre classi di dati congiuntamente che per i singoli dati di una classe. Ecco che, osservando i grafici presenti in appendice che mostrano tale confronto, possiamo giungere alla conclusione presentata precedentemente secondo la quale la funzione filtrante relativa al volume meglio raccoglie l’informazione sulla modalità della frase in confronto alle altre funzioni. Relativamente a questa funzione filtrante osserviamo infatti che le *cornerlines* dei dati delle diverse classi rimangono abbastanza separate tra di loro (C.1) mentre quelle relative allo stesso tipo di frase sono piuttosto ravvicinate e ciò consente all’algoritmo che calcola la distanza di *bottle-neck* di non doversi occupare di distanze eccessive tra *cornerlines*. Inoltre è possibile notare che all’interno della stessa classe di dato, i diagrammi di persistenza risultano simili non solo per la vicinanza delle *cornerlines* ma anche per l’intera disposizione dei *cornerpoints* (C.7, C.8, C.9).

Entrando nel vivo della ricerca abbiamo osservato che la classe di dati meglio distinta e raggruppata, indifferentemente dal metodo di linkage usato o dalla funzione filtrante considerata, è la classe delle domande polari. Osservando infatti uno qualsiasi dei dendrogrammi creati noteremo che tali domande vengono raggruppate tra di loro con maggior frequenza rispetto agli altri dati e che in generale appartengono a cluster più grandi rispetto a quelli che si formano con le altre frasi. Il motivo per cui ciò avviene riguarda il fatto che questa tipologia di frase si presenta con elementi audio molto simili tra loro. Andando infatti a guardare le curve di frequenza fondamentale e di volume di frasi appartenenti a questa categoria riusciremo a riconoscere ad occhio una forte somiglianza. Ad esempio di seguito (4.2) vediamo coppie di curve di frequenza fondamentale

f_0 (in blu) e di volume (in rosso) relative a quattro dei 19 file estratti dal data set di indagine: le curve sono molto simili tra di loro.

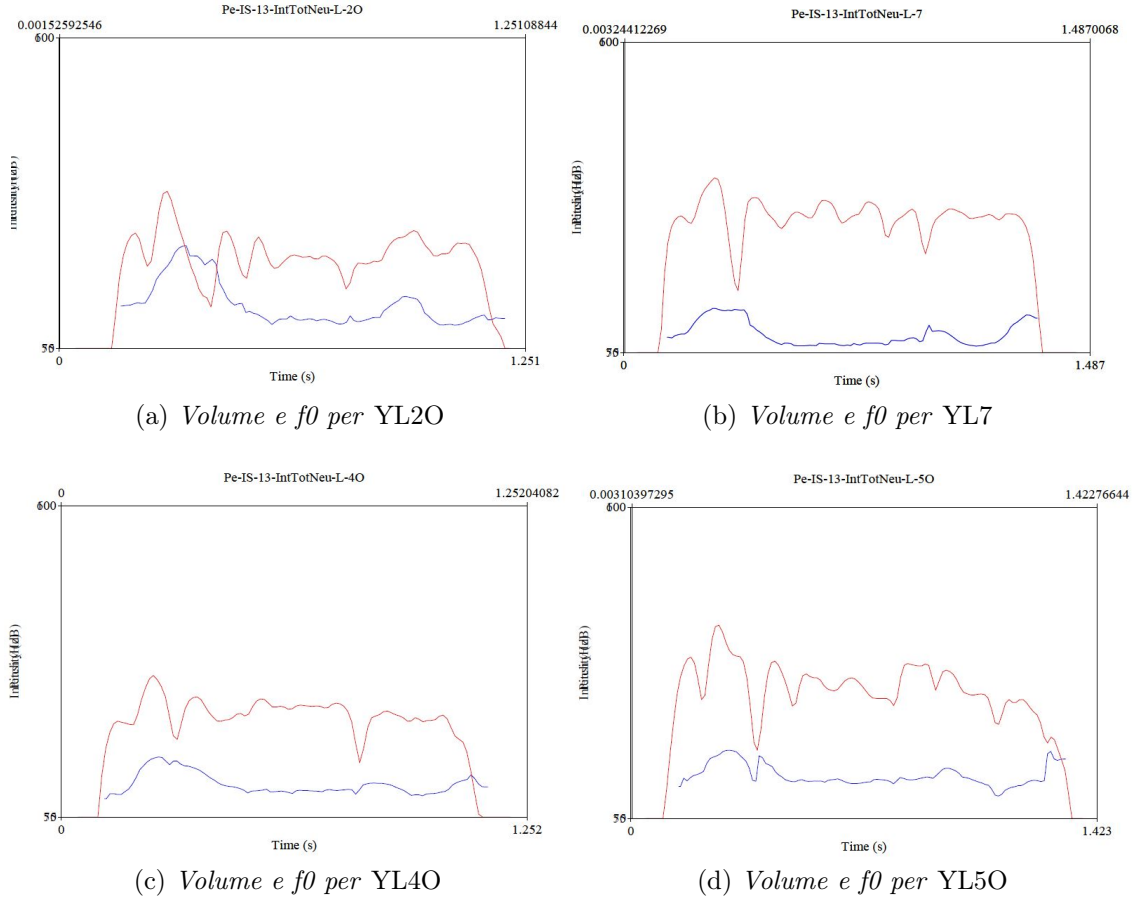


Figura 4.2: *Confronti tra i parametri per le domande polari*

Questa somiglianza è presente anche tra le curve delle altre classi di dato ma non con questa forza, per questo tale categoria è meglio distinta e associata rispetto alle altre. Dal punto di vista matematico, quanto osservato è ancora una volta supportato dall'analisi dei dati mediante i diagrammi di persistenza; d'altra parte, come sappiamo, catturare le somiglianze tra curve ottenute secondo parametri filtranti è il principale scopo della topologia persistente.

Nel senso discusso, i dati appartenenti alla classe delle domande parziali sono secondi in ordine di consistenza del raggruppamento mentre i dati della classe delle liste vengono associati con maggior difficoltà. Si osserva che le frasi registrate facenti parte della categoria delle liste, sono maggiormente variegata rispetto alle frasi delle altre classi anche in termini di pluralità del contenuto semantico e dunque questo fatto potrebbe incidere

sulle cause che hanno portato allo scarso successo di raggruppamento. Mentre le registrazioni delle domande sono per la maggior parte relative ad un'unica domanda posta con le stesse parole da parte dei vari speaker, le registrazioni delle liste raccolgono sequenze diverse di item che differiscono per lunghezza e contenuto. Nonostante ciò, le liste così registrate sono state inserite nel data set pensando che potessero venire clusterizzate in modo corretto a causa dell'andamento caratteristicamente altalenante delle curve di rifermento che si contrappone in modo piuttosto evidente agli andamenti più uniformi delle curve degli altri dati.

Un aspetto evidente dall'osservazione dei dendrogrammi è la presenza di dati che si collocano all'esterno dei cluster indipendentemente dal metodo di linkage usato. In particolare si nota che ci sono due dati che rimangono esterni per le funzioni filtranti riferite al parametro di frequenza e due dati con comportamento analogo relativamente alla funzione filtrante del volume. Andiamo ad osservare nello specifico i file introdotti. Il file audio LS7 notiamo nei dendrogrammi (ad esempio in B.6, B.1) che non riesce ad essere collocato all'interno di un raggruppamento dalle funzioni filtranti f_0 e Df_0 . Il file in analisi è una lista prodotta spontaneamente dallo speaker numero 7 il cui contenuto semantico si distacca molto dal resto delle registrazioni di questa categoria. Esso infatti corrisponde all'emissione della frase: 'Mi piacciono mandarini, fragole, banane, lamponi, succhi di frutta in genere, comunque la frutta in media tutta.' che come possiamo notare non presenta solo una lista di elementi ma contiene anche del materiale non strettamente allineato con il tipo di dato di questa categoria. In prima analisi potremmo pensare che sia questo il solo motivo per cui il file non viene in alcun modo associato ad altri file della sua categoria; andando però ad analizzare il suo diagramma di persistenza relativo alla funzione filtrante Df_0 notiamo che la cornerline del diagramma è consistentemente distaccata dal resto dei punti e, ponendola in confronto con gli altri dati del tipo liste, si trova isolata da tutte le altre cornerlines (C.5). Questa lontananza così consistente della cornerline del file LS7 dagli altri file, per i motivi spiegati sopra, causa l'elaborazione di un valore di distanza molto alto dagli altri elementi, ecco perché il dato non riesce ad essere inserito in un cluster. Ci siamo però chiesti cosa causi questa anomalia, se sia il diverso contenuto del file o piuttosto un qualche altro fattore. A tal proposito ho studiato il file osservandone la curva della frequenza fondamentale che riporto di seguito 4.3.

Come evidenziato, la curva di frequenza, dopo qualche istante, presenta un picco che raggiunge i 598.51 Hz mentre il resto dell'emissione varia tra i 75 Hz e 110 Hz. Tale picco è registrato anche dai nodi della funzione filtrante f_0 e di conseguenza dalla funzione filtrante di derivata che presenta pertanto un picco negativo in corrispondenza del valore anomalo che giustifica dunque la creazione di una cornerline come descritta. Andando a isolare il picco di frequenza con gli strumenti offerti dal software Praat è stato possibile scoprire che tale valore corrisponde all'emissione della consonante 'f' presente nella parola 'fragole' pronunciata dallo speaker con un effetto di fischio. Infatti ascoltando

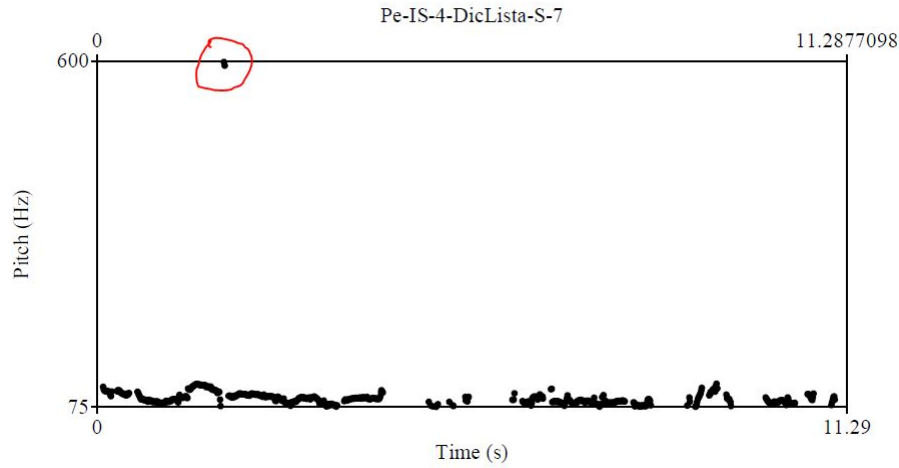


Figura 4.3: *Frequenza fondamentale del dato LS7*

l'audio associato al picco si sente un fruscio proprio della pronuncia della consonante 'f' e ascoltando la traccia negli istanti immediatamente precedenti e successivi possiamo identificare il materiale fonetico che precede e segue tale consonante nella frase esaminata. Dunque la causa dell'isolamento di questo dato nei diagrammi di cluster è dovuta non tanto alla presenza di un differente contenuto semantico ma piuttosto alla registrazione di un suono che causa una diversità consistente nelle curve delle funzioni filtranti.

In maniera analoga a quanto appena descritto, anche il file LS80 presenta un'anomalia che non ne consente la corretta clusterizzazione per mezzo delle funzioni filtranti relative al parametro di frequenza. Il dato corrisponde all'emissione della frase 'Cocco, melone, mela, anguria, arancia, mandarino, limone.' ed è infatti appartenente alla classe delle liste presentando un contenuto piuttosto standard. Come prima, andando ad osservarne il diagramma di persistenza relativamente alla funzione filtrante Df_0 notiamo che la cornerline si trova tra quella del dato LS7 e le altre rimanendo comunque molto distaccata (C.5). Di seguito (4.4) osserviamo la curva della frequenza fondamentale f_0 estratta per mezzo di Praat in cui possiamo notare la presenza di un altro picco (514.39 Hz) seppur più basso rispetto a quello del dato precedente.

In questo caso il picco corrisponde alla pronuncia della sillaba finale della parola 'cocco'. Isolando la sezione di frequenza possiamo sentire come il suono di una 'o' molto acuto come se nel pronunciare l'ultima sillaba la doppia consonante 'c' avesse causato la fonazione della vocale 'o' con una spinta aggiuntiva, sappiamo infatti che fenomeni di questo tipo sono naturali del processo di articolazione del suono. Dunque il dato LS80 rimane in genere isolato dagli altri elementi seppur capiti che venga associato con LS7 in quanto presentano la stessa anomalia.

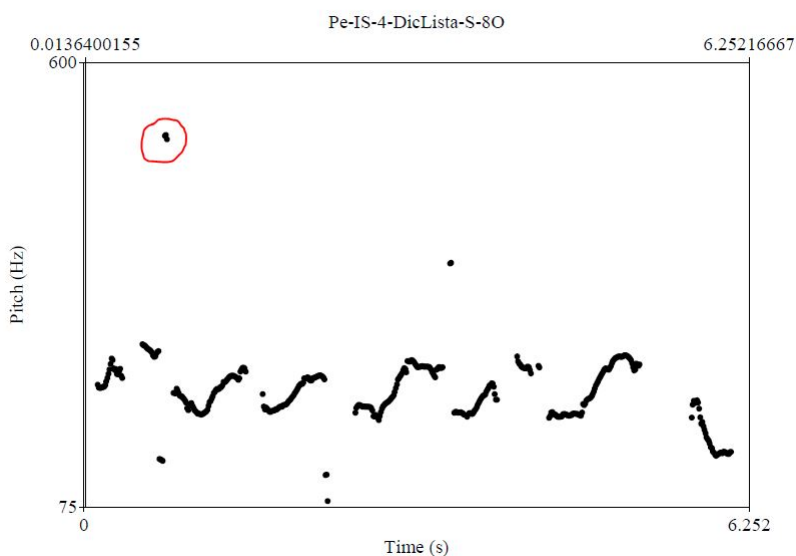
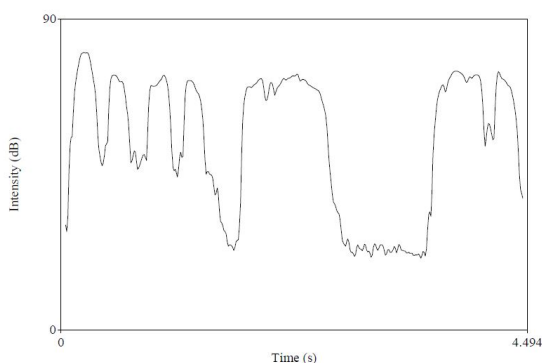
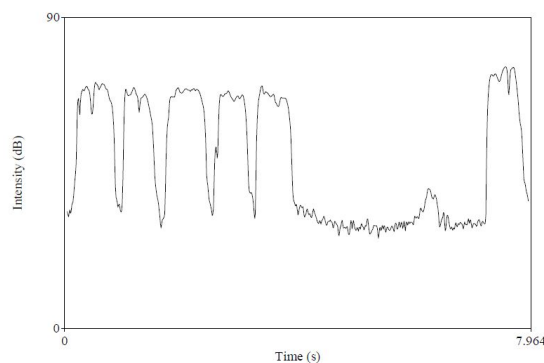


Figura 4.4: *Frequenza fondamentale del dato LS80*

Consideriamo ora i dati LS60 e LS40 appartenenti, ancora una volta, alla classe delle liste e accomunati dal fatto che entrambi non vengono raggruppati in cluster corretti per mezzo della funzione filtrante relativa al parametro di volume. Infatti osservando i dendrogrammi ottenuti con diversi metodi di linkage ma relativi al volume si può notare che questi due dati rimangono esterni alla clusterizzazione. Considerando i diagrammi di persistenza relativi alla funzione filtrante del volume per entrambi i dati vediamo come le rispettive cornerlines siano distaccate da quelle degli altri dati appartenenti a questa classe (C.4) riportandoci a un fenomeno analogo a quello già descritto. In questo caso però l'anomalia non riguarda la frequenza delle registrazioni ma il volume: entrambi i dati sono accomunati dalla presenza di pause di considerevole durata in una qualche parte dell'enunciato. Tenendo conto di quanto detto nel capitolo precedente circa la natura vibratoria continua del suono, si ricorda che i valori ottenuti da Praat relativi all'andamento del volume erano stati considerati nella loro integrità senza eliminare sezioni di discontinuità come invece fatto per i dati sulla frequenza. Questo ha portato a includere anche valori di livello sonoro corrispondenti a pause di varia durata, da quelle impercettibili a quelle più consistenti.

Come si vede dai grafici di seguito (4.5a e 4.5b) i dati LS60 e LS40 presentano pause di durata rispettivamente 1.01 e 3.29 secondi, di gran lunga superiori a quelle presenti negli altri file, in cui si registrano valori fino a 20 Db che indica una situazione di assenza di vibrazione sonora legata al parlato.

Dunque a causa di queste caratteristiche i dati osservati rimangono esterni ai cluster secondo la funzione filtrante del volume sebbene presentino il tipico andamento di livello sonoro della classe delle liste.

(a) *Livello di intensità sonora per il file LS6O*(b) *Livello di intensità sonora per il file LS4O*

Avendo analizzato e commentato gli elementi le cui caratteristiche sono causa della mancata clusterizzazione, soffermiamoci ora su quanto invece è stata catturata correttamente dalla nostra ricerca. Come detto, la tipologia di dato che meglio viene raggruppata è quella delle domande polari, che presenta raggruppamenti in media di 13 elementi sui 19 totali di cui è composta la classe per qualsiasi tipo di linkage utilizzato nella creazione del dendrogramma. Tali raggruppamenti a volte arrivano a comprendere sino a 8/9 elementi mentre in genere i cluster che si formano relativamente ad altre classi di dato comprendono fino al numero massimo di 5 elementi. Numeri come quelli appena dati possono portare a pensare che il risultato ottenuto non sia molto interessante, ma dobbiamo ricordare che si sta lavorando con un piccolo data set e che gli elementi per classe sono poco meno di due decine.

Le clusterizzazioni computate contengono interessanti considerazioni oltre a quelle fatte. Possiamo osservare infatti che per la maggior parte, i cluster corretti si formano andando ad associare da prima i dati relativi a uno stesso speaker allargandosi poi a dati prodotti da speaker dello stesso sesso fino a includere in un secondo momento, qualora avvenga, dati prodotti da altri speaker appartenenti al sesso opposto. Ciò significa che, prendendo come esempio un cluster di domande polari, è più facile trovare associati gli elementi YL7, YL7b, YL7O, YS7, YS7b, YS7O tra di loro perché prodotti tutti dallo speaker (uomo) numero 7 e in un secondo momento trovarli associati agli elementi YL5O, YS5, YL2O, YS3, YS3O perché corrispondenti a frasi della stessa classe pronunciate da speaker dello stesso sesso. Dall'analisi dei dendrogrammi vediamo come questo avvenga costantemente indifferentemente dal tipo di dato, funzione filtrante o metodo di linkage considerato. Infatti nel caso siano presenti più cluster di dati appartenenti alla stessa classe, capita spesso che un cluster sia relativo agli audio emessi da speaker di sesso femminile e un altro sia relativo solamente a dati prodotti da speaker di sesso maschile. Per fornire al lettore una chiave di analisi autonoma dei dendrogrammi riportiamo che

gli audio prodotti dagli speaker numeri 1, 3, 6 e 8 sono relativi a voci femminili mentre quelli emessi dagli speaker numeri 2, 4, 5, 7 sono relativi a voci maschili.

Questo aspetto sebbene sia interessante e giustificato dal fatto che una stessa voce (o voci dello stesso sesso) generano curve simili tra loro, è da tenere in costante studio; infatti il nostro scopo non è attuare un riconoscimento vocale, ma ottenere un riconoscimento modale e pragmatico di enunciati.

Tenendo conto di questa osservazione e di quelle fatte in precedenza, possiamo osservare che il cluster più grande lo si trova nel dendrogramma costruito con il metodo ward e relativo alla funzione filtrante del volume (B.11). Esso raggruppa 12 domande polari anche se contiene un errore, infatti uno degli elementi raggruppati non è una domanda polare ma una domanda parziale che però vediamo essere prodotta dallo speaker numero 7 ed essere inserita in un raggruppamento di emissioni prodotte dallo stesso speaker. È interessante notare inoltre che il cluster in esame contiene 4 sottoraggruppamenti dei quali uno è relativo a sole emissioni prodotte da voci femminili, uno è relativo a sole frasi prodotte da voce maschili ed i due restanti sotto gruppi posizionati al centro contengono emissioni miste. In generale nel dendrogramma citato è possibile anche distinguere a grandi linee tre zone corrispondenti ai tre tipi di dato guardando l'ordinamento delle etichette poste a lato. È evidente la presenza di errori ma si ritiene essere comunque un buon risultato.

L'analisi dei risultati non si è fermata alla sola considerazione dei dendrogrammi costruiti a partire dalle singole funzioni filtranti ma si è anche considerata una matrice somma delle distanze data dalla somma normalizzata dei contributi delle tre funzioni filtranti al fine di ricavare un dendrogramma riassuntivo. I grafici ottenuti con tale matrice (B.16, B.17, B.18, B.19, B.20) elaborano le informazioni interne ai grafici relativi alle singole funzioni e pertanto si è pensato potessero essere utili a riassumere i risultati delle ricerche. Osservandoli nel dettaglio si nota però che non riescono a riportare con la stessa efficacia le informazioni estrapolate da altri grafici. Infatti indifferentemente dal metodo di linkage utilizzato, il dendrogramma della matrice somma in genere non presenta un maggior numero di raggruppamenti rispetto ai dendrogrammi delle singole funzioni. In media infatti il dendrogramma relativo alla funzione somma raggruppa correttamente circa 25 elementi mentre i dendrogrammi relativi alla funzione filtrante del volume arrivano a raggrupparne in media 35. Questo significa che la somma diretta normalizzata riassume le caratteristiche più persistenti ma non esprime i risultati al meglio delle analisi effettuate. È dunque in linea con le aspettative osservare che i dati meglio raggruppati siano ancora una volta quelli relativi alle domande polari e che questa somma abbia continuato ad evidenziare la presenza dei dati outlier descritti precedentemente. In particolare rimangono esterni i dati LS7 e LS80 che sono distaccati dagli altri secondo le funzioni filtranti relative alla f_0 e alla sua derivata.

A seguito dell'analisi sulla funzione somma, si è pensato di creare una matrice somma che raccogliesse le informazioni provenienti dalle varie funzioni filtranti associando a ciascuna un peso in relazione al grado di efficacia di raggruppamento. Per quanto considerato i set di pesi applicati andranno a favorire l'informazione raccolta dalla funzione filtrante del volume, in questo modo si cerca di ottenere dei dendrogrammi ottimizzati per mostrare il meglio della ricerca eseguita. Inoltre creando matrici somma con tali pesi si va a correggere anche la posizione esterna dei dati discussi che, a causa delle loro caratteristiche, non vengono associati correttamente dalle due funzioni filtranti relative alla frequenza fondamentale.

Il primo set di pesi applicato prevede l'assegnazione dei coefficienti $\frac{1}{4}$, $\frac{1}{4}$ e $\frac{1}{2}$ ai contributi normalizzati provenienti rispettivamente dalle funzioni filtranti relative alla frequenza fondamentale f_0 , alla sua derivata Df_0 e al volume. Si è computato cioè

$$(Somma - pesata1)_{i,j} = \frac{1}{4}(f_0)_{i,j} + \frac{1}{4}(Df_0)_{i,j} + \frac{1}{2}(Db)_{i,j}.$$

Osservando i dendrogrammi così ottenuti (B.21, B.22, B.23) si nota un miglioramento rispetto ai dendrogrammi relativi alla matrice di somma semplice. Dalle analisi compiute (in riferimento ai soli metodi di linkage complete, ward ed average) si osserva che si passa da 24/25 elementi raggruppati a circa 31 elementi per la matrice di somma pesata.

Cercando di ottimizzare maggiormente il risultato si è proceduto alla creazione di una seconda matrice pesata utilizzando altri coefficienti che andassero a favorire la funzione del volume ma anche quella della frequenza fondamentale. È noto dalle precedenti considerazioni che anche i dendrogrammi relativi alla sola funzione filtrante f_0 colgono una certa quantità di informazione. Pertanto si è proceduto a calcolare

$$(Somma - pesata2)_{i,j} = (0.35) * (f_0)_{i,j} + (0.15) * (Df_0)_{i,j} + (0.5) * (Db)_{i,j}.$$

I dendrogrammi computati in questo modo, e relativi alle funzioni di linkage complete, ward, average, (B.24, B.25, B.26) presentano un ulteriore piccolo miglioramento rispetto a quelli precedenti: essi infatti riescono a raggruppare in media 32 elementi.

Un'osservazione necessaria relativa ad entrambi i set di dendrogrammi ottenuti con le funzioni di somma pesata è relativa al tipo di associazione che avviene nella creazione dei cluster. Sono infatti aumentati gli errori di associazione tra frasi che vengono legate assieme perché prodotte dallo stesso speaker piuttosto che per stessa modalità intonativa. Ovviamente i numeri appena forniti sono relativi al numero di associazioni corrette avvenute, ma osservando gli altri cluster 'sbagliati' quanto descritto è ciò che si nota maggiormente. Infine occorre ricordare che un'assegnazione di pesi ancora più accurata, sulla base di esperimenti effettuati su un data set così piccolo, potrebbe produrre l'indesiderato fenomeno dell'overfitting.

4.2 Conclusioni e obiettivi futuri

Come conclusione del presente progetto di tesi ritengo che il migliore risultato ottenuto sia mostrato dal dendrogramma (B.11) relativo alla funzione filtrante del volume e computato per mezzo del metodo di linkage ward in cui sono raggruppati ben 38 elementi su 56 totali. Come ampiamente discusso anche in questo caso la classe meglio riconosciuta e associata è quella delle domande polari anche se relativamente a questo dendrogramma è cospicua anche la quantità di dati appartenenti alla classe delle liste che sono state correttamente associati. Quest'ultimo dato è particolarmente interessante se ricordiamo che la classe delle liste, per le caratteristiche descritte, in genere è raramente raggruppata correttamente.

Possiamo concludere affermando che gli obiettivi di tesi sono stati in parte soddisfatti e che eventuali miglorie possono essere effettuate partendo dalle analisi e osservazioni effettuate in questo progetto. In ogni caso, come ipotizzato, la collaborazione tra le discipline coinvolte si è dimostrata proficua.

Relativamente alle eventuali future analisi in questo ambito si suggerisce di provare ad utilizzare ulteriori funzioni filtranti in grado di evidenziare nuove caratteristiche degli enunciati. Ricordiamo che il contorno intonativo di una frase contiene l'informazione relativa alla sua modalità e pragmaticità soprattutto nella sezione in cui sono situati gli accenti di frase. In questa analisi non ci siamo concentrati sulla sezione accentata della frase come fulcro informativo, ma abbiamo piuttosto considerato l'intero enunciato senza aggiungere dati che rimarcassero la zona che contiene maggiore informazione. Nelle future analisi si potrebbe considerare di inserire parametri o funzioni filtranti che svolgano questo compito. Potrebbe inoltre essere interessante continuare la ricerca di coefficienti da inserire come pesi per ottenere un'unica funzione riassuntiva delle informazioni fornite dalle funzioni filtranti ottimizzata in modo che esprima al meglio i risultati, avendo cura, però, di evitare l'overfitting. Ancora, sarebbe interessante ricostruire i diagrammi di persistenza e le matrici di distanza utilizzando una distanza differente dalla bottleneck in modo da risolvere i problemi descritti in caso di cornerlines molto distanti e per osservare se è possibile trarre ulteriori informazioni dalle sole funzioni filtranti utilizzate. Un altro interessante studio si potrebbe compiere al fine di uniformare i dati per evitare problematiche di picchi di frequenza, esistono infatti tecniche di filtro che vanno a togliere valori di frequenza (o di Decibel) al di fuori di un determinato range. Studiare come inserire tali tecniche e il loro effetto sulla disposizione finale dei dati potrebbe condurre a nuovi risultati.

Infine quanto iniziato con questa ricerca potrà condurre, attraverso progressivi miglioramenti, alla classificazione modale di un set ben più ampio di quello ad ora utilizzato che preveda molte più classi di dati. Questo potrà effettivamente portare alla soluzione di eventuali nuove problematiche ed, infine, a un contributo utile allo studio dei contorni intonativi.

Appendice A

Codici Dendrogrammi

Ho creato il seguente codice in ambiente MatLab per creare e visualizzare i dendrogrammi in base alle richieste sulla funzione filtrante da considerare e sul metodo di linkage da usare.

Matlab Code

```
1 %%__Creazione di Dendrogrammi__
2 clear all
3 %Inserire la matrice delle distanze in forma simmetrica
4 MatDistf0=importdata('Distanzef0.txt');
5 MatDistDf0=importdata('DistanzeDf0.txt');
6 MatDistdb=importdata('Distanzedb.txt');
7
8 %Normalizzazione delle matrici di distanza per creare una
   matrice somma oppure introdurne una con MatDistSomma=
   importdata('MatDistSomma_norm.mat');
9 %Determino le dimensioni, valide per tutte le matrici
10 [n,m]=size(MatDistdb);
11
12 %%%Normalizzazione matrice dei Decibel
13 MaxDb=max(max(MatDistdb))
14 MatDistdb_norm=MatDistdb/MaxDb;
15 %%%Normalizzazione matrice di f0
16 Maxf0=max(max(MatDistf0))
17 MatDistf0_norm=MatDistf0/Maxf0;
18 %%%Normalizzazione matrice Df0
19 MaxDf0=max(max(MatDistDf0))
```

```

20 MatDistDf0_norm=MatDistDf0/MaxDf0;
21
22 %%Creazione della matrice somma
23 MatDistSomma=zeros(n,n);
24 for i=1:n
25     for j=1:n
26         MatDistSomma(i,j)=MatDistf0_norm(i,j)+
                MatDistDf0_norm(i,j)+MatDistdb_norm(i,j);
27     end
28 end
29
30 %%Creazione matrice somma primo set di pesi
31 MatDistSommaPESATA1=zeros(n,n);
32 for i=1:n
33     for j=1:n
34         MatDistSommaPESATA1(i,j)=(1/4)*MatDistf0_norm(i,j)
                +(1/4)*MatDistDf0_norm(i,j)+(1/2)*MatDistdb_norm(i
                ,j);
35     end
36 end
37
38 %%Creazione matrice somma con secondo set di pesi.
39 MatDistSommaPESATA2=zeros(n,n);
40 for i=1:n
41     for j=1:n
42         MatDistSommaPESATA2(i,j)=(0.35)*MatDistf0_norm(i,j)
                +(0.15)*MatDistDf0_norm(i,j)+(0.5)*MatDistdb_norm(
                i,j);
43     end
44 end
45
46 %Inserire etichette degli elementi
47 Etichette={'LL2','LL2b','LL20','LL20b','LL3','LL40','LL6','
                LL60','LL8','LL80','LS10','LS3','LS30','LS40','LS60','LS7'
                ,'LS8','LS80','YL20','YL30','YL40','YL50','YL60','YL7','
                YL7b','YL70','YS1b','YS3','YS30','YS40','YS5','YS7','YS7b'
                ,'YS70','YYL30','YYS1','YYS10','WL3','WL30','WL4','WL40b',
                'WL5','WL50','WL6','WL6b','WL60','WL7','WL70','WS1','WS10'
                ,'WS2','WS4','WS40','WS5','WS6','WS6b'};
48 n=length(Etichette);
49

```

```
50 %%Creazione alberi attraverso linkage attraverso i metodi:
    complete, single, average, ward, weighted.
51 %Scelta del Linkage da input
52 method=input('Quale metodo di linkage vuoi usare? complete ,
    single, ward, weighted, average? \n');
53
54 %Creazione alberi attraverso linkage -- f0
55 treef0 = linkage(MatDistf0,method,'euclidean');
56 leafOrderf0 = optimalleaforder(treef0,MatDistf0); %Crea l'
    ordine ottimale
57
58 %Creazione alberi attraverso linkage -- Df0
59 treeDf0 = linkage(MatDistDf0,method,'euclidean');
60 leafOrderDf0 = optimalleaforder(treeDf0,MatDistDf0);
61
62 %Creazione alberi attraverso linkage -- db
63 treedb = linkage(MatDistdb,method,'euclidean');
64 leafOrderdb = optimalleaforder(treedb,MatDistdb);
65
66 %Creazione alberi attraverso linkage -- somma
67 treesomma = linkage(MatDistSomma,method,'euclidean');
68 leafOrderSomma = optimalleaforder(treesomma,MatDistSomma);
69
70 %Creazione alberi attraverso linkage -- somma pesata1
71 treesommaP1 = linkage(MatDistSommaPESATA1,method,'euclidean'
    );
72 leafOrderSommaP1 = optimalleaforder(treesommaP1,
    MatDistSommaPESATA1);
73
74 %Creazione alberi attraverso linkage -- somma pesata2
75 treesommaP2= linkage(MatDistSommaPESATA2,method,'euclidean')
    ;
76 leafOrderSommaP2 = optimalleaforder(treesommaP2,
    MatDistSommaPESATA2);
77
78 %Stampa dei dendrogrammi ottenuti per confrontarli
79 Ordine=menu('TIPOLOGIA DI GRAFICO','Default Leaf Order','
    Optimal Leaf Order');
80
81 switch Ordine
82     case 1
```

```

83     funzione=menu('FUNZIONE DI INTERESSE','f0','Df0','db
      ','Somma','Somma pesata 1','Somma pesata 2');
84     switch funzione
85         case 1
86             dendrogram(treef0,56,'Orientation','left','
      labels',Etichette)
87             title(['Default Leaf Order - f0 function -
      method: ' method ])
88         case 2
89             dendrogram(treeDf0,56,'Orientation','left','
      labels',Etichette)
90             title(['Default Leaf Order - Df0 function -
      method: ' method])
91         case 3
92             dendrogram(treedb,56,'Orientation','left','
      labels',Etichette)
93             title(['Default Leaf Order - db function -
      method: ',method])
94         case 4
95             dendrogram(treesomma,56,'Orientation','left'
      , 'labels',Etichette)
96             title(['Default Leaf Order - Sum function -
      method: ', method])
97         case 5
98             dendrogram(treesommaP1,56,'Orientation','
      left','labels',Etichette)
99             title(['Default Leaf Order - Somma pesata 1
      function - method: ', method])
100        case 6
101            dendrogram(treesommaP2,56,'Orientation','
      left','labels',Etichette)
102            title(['Default Leaf Order - Somma pesata 2
      function - method: ', method])
103        end
104    case 2
105        funzione2=menu('FUNZIONE DI INTERESSE','f0','Df0','
      db','Somma','Somma pesata 1','Somma pesata 2');
106        switch funzione2
107            case 1
108                dendrogram(treef0,56,'Orientation','left','
      reorder',leafOrderf0,'labels',Etichette)

```



```
109         title(['Optimal Leaf Order - f0 function -
110                method: ', method])
111     case 2
112         dendrogram(treeDf0,56,'Orientation','left','
113                reorder',leafOrderDf0,'labels',Etichette)
114         title(['Optimal Leaf Order - Df0 function -
115                method: ', method])
116     case 3
117         dendrogram(treedb,56,'Orientation','left','
118                reorder',leafOrderdb,'labels',Etichette)
119         title(['Optimal Leaf Order - db function -
120                method: ', method])
121     case 4
122         dendrogram(treesomma,56,'Orientation','left'
123                , 'reorder',leafOrderSomma,'labels',
124                Etichette)
125         title(['Optimal Leaf Order - Sum function -
126                method: ', method])
127     case 5
128         dendrogram(treesommaP1,56,'Orientation','
129                left','reorder',leafOrderSommaP1,'labels',
130                Etichette)
131         title(['Optimal Leaf Order - Somma pesata 1
132                function - method: ', method])
133     case 6
134         dendrogram(treesommaP2,56,'Orientation','
135                left','reorder',leafOrderSommaP2,'labels',
136                Etichette)
137         title(['Optimal Leaf Order - Somma pesata 2
138                function - method: ', method])
139     end
140 end
```

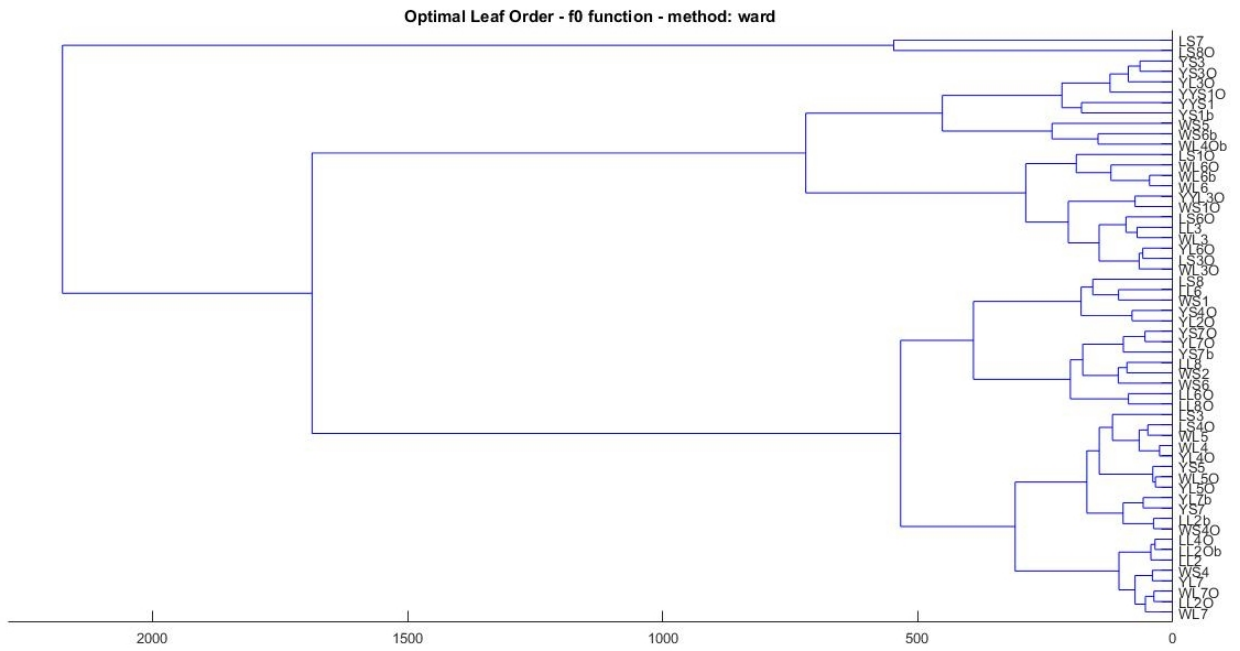
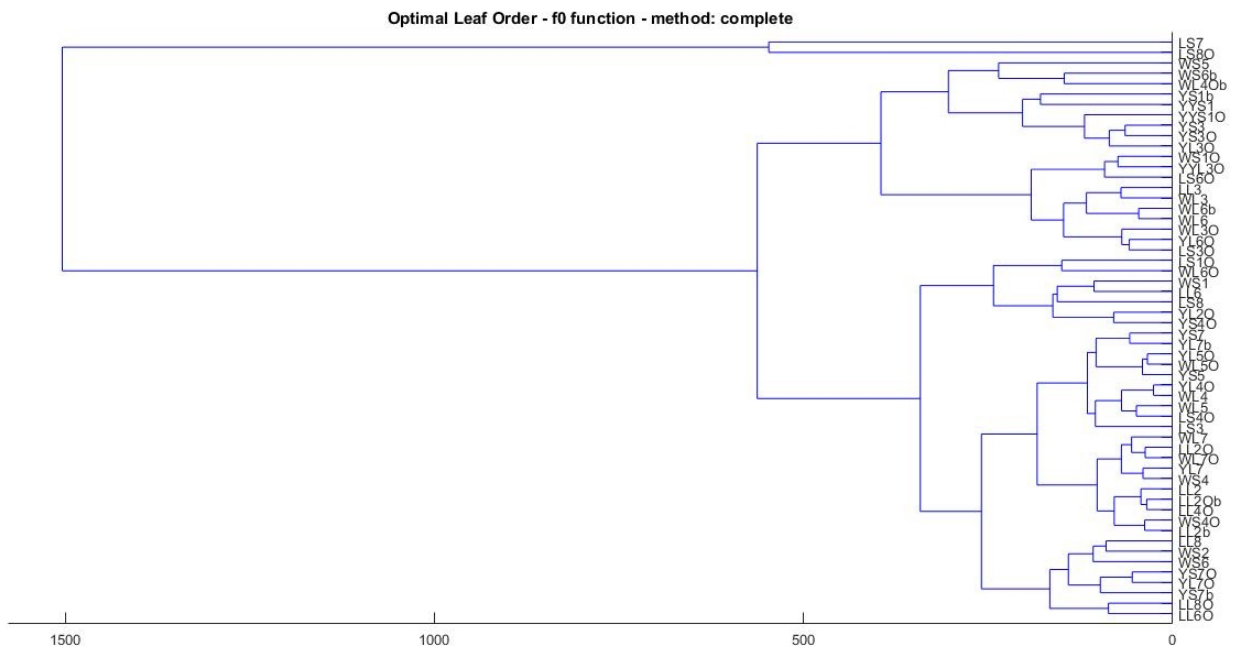

Appendice B

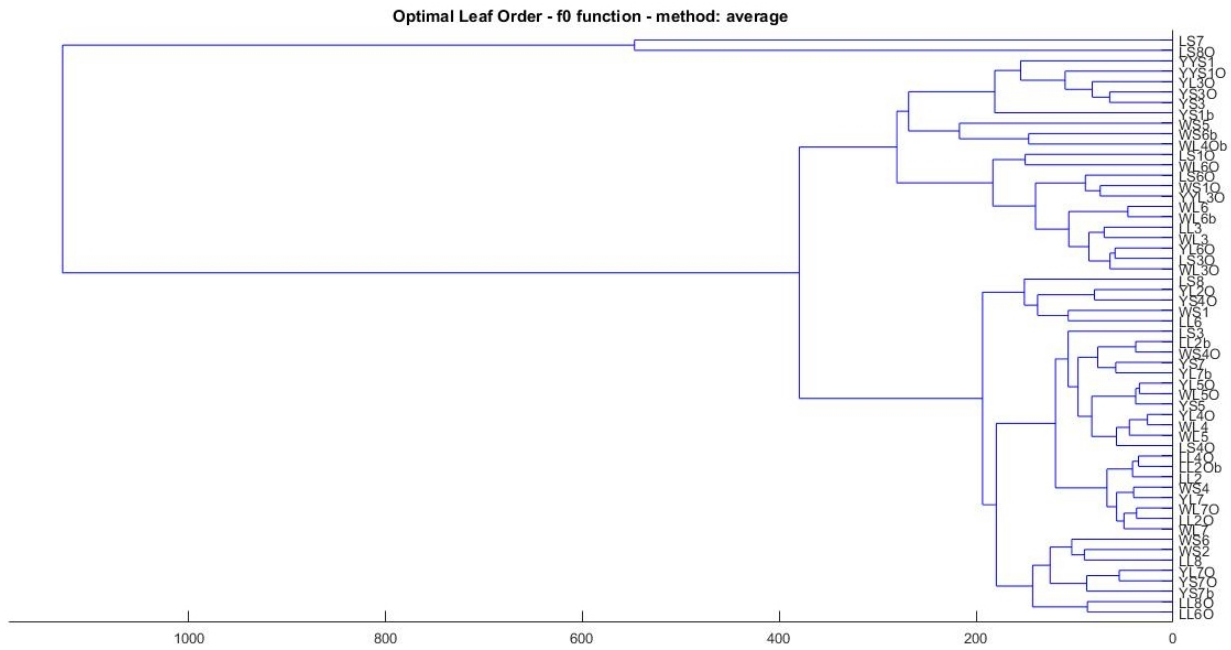
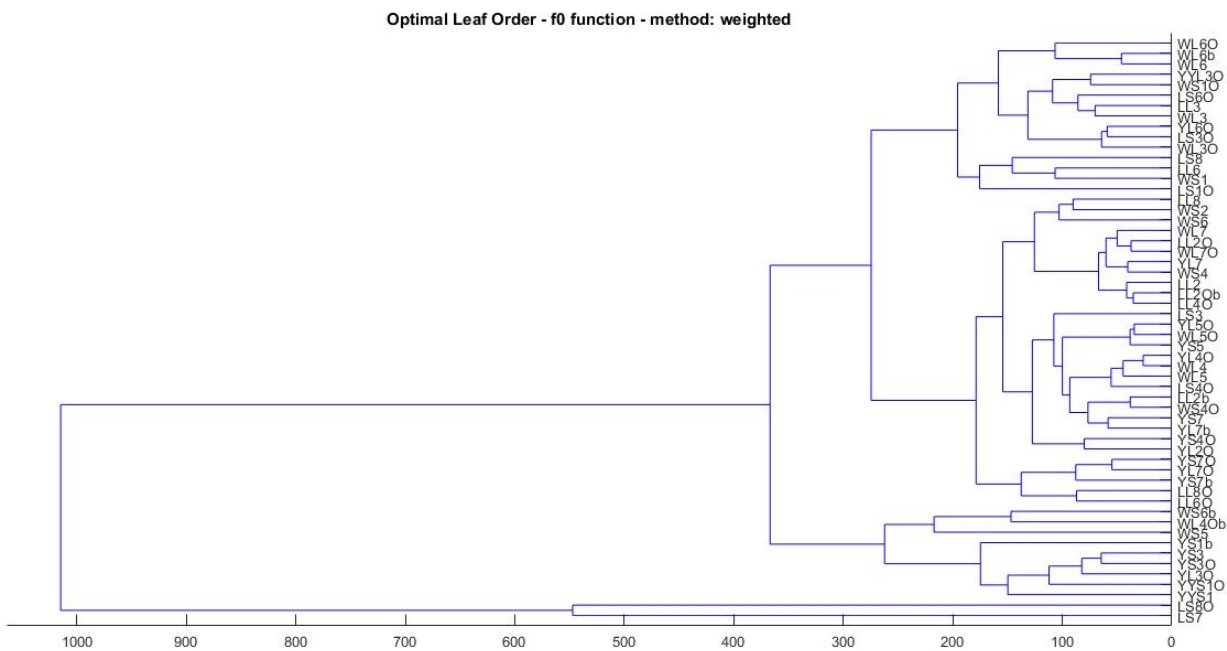
Dendrogrammi

I dendrogrammi riportati di seguito sono stati generati per mezzo del codice precedentemente riportato e rappresentano graficamente i risultati ottenuti in questo progetto di tesi. Sono stati inseriti per completezza tutti i dendrogrammi ricavati dai possibili metodi di linkage descritti per ogni funzione filtrante considerata e per le matrici somma e somma pesata.

Di seguito l'elenco delle figure.

Funzioni filtranti			Funzioni somma		
Funzione f_0	Ward	B.1	Somma semplice	Ward	B.16
	Complete	B.2		Complete	B.17
	Average	B.3		Average	B.18
	Weighted	B.4		Weighted	B.19
	Single	B.5		Single	B.20
Funzione Df_0	Ward	B.6	Somma pesata 1	Ward	B.21
	Complete	B.7		Complete	B.22
	Average	B.8		Average	B.23
	Weighted	B.9	Somma pesata 2	Ward	B.24
	Single	B.10		Complete	B.25
Funzione Volume	Ward	B.11		Average	B.26
	Complete	B.12			
	Average	B.13			
	Weighted	B.14			
	Single	B.15			

Figura B.1: *Dendrogramma: F0 - ward*Figura B.2: *Dendrogramma: F0 - complete*

Figura B.3: *Dendrogramma: F0 - average*Figura B.4: *Dendrogramma: F0 - weighted*

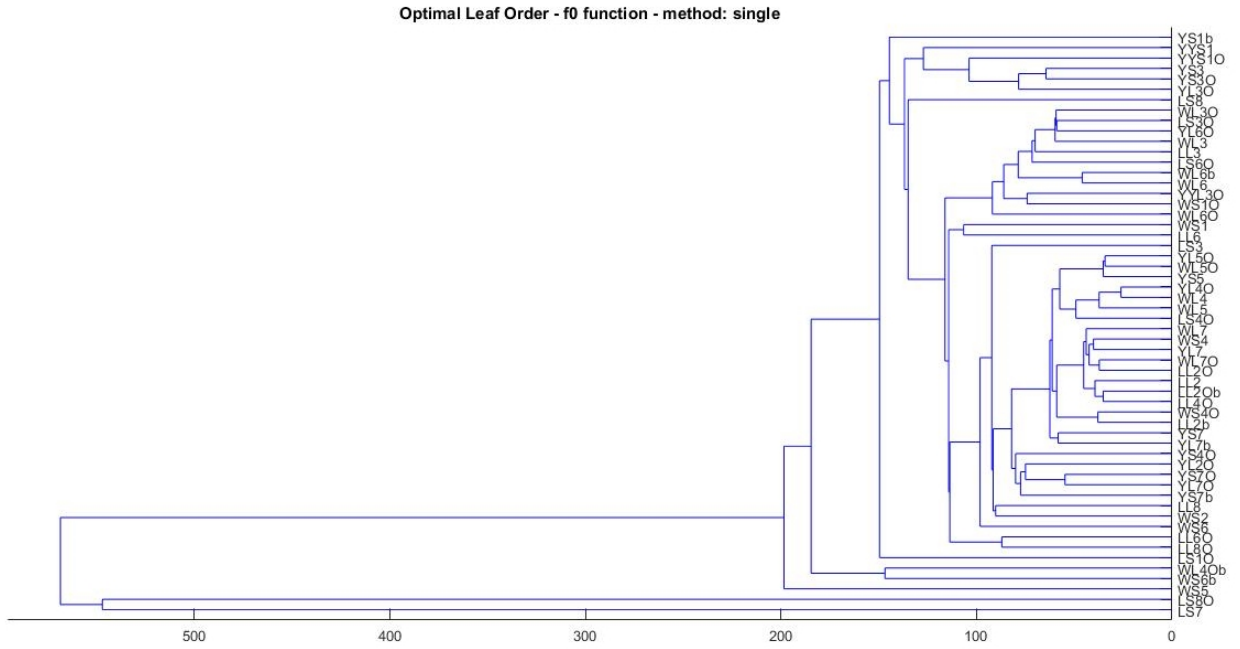


Figura B.5: *Dendrogramma: F_0 - single*

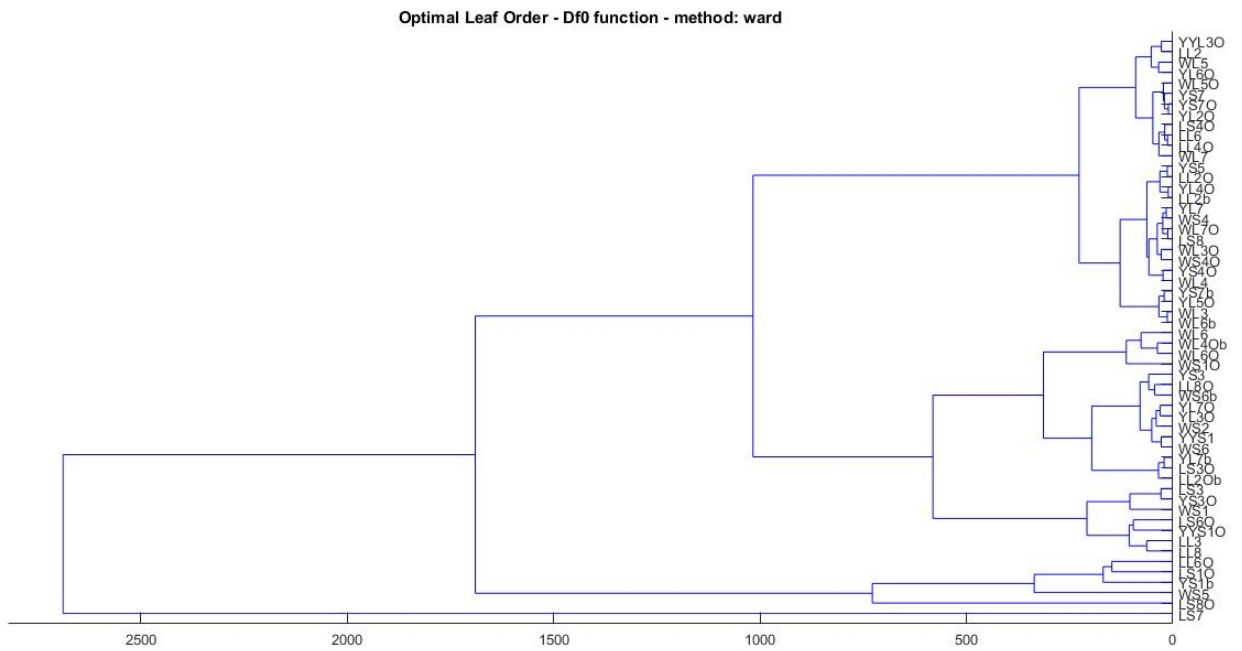


Figura B.6: *Dendrogramma: DF_0 - ward*

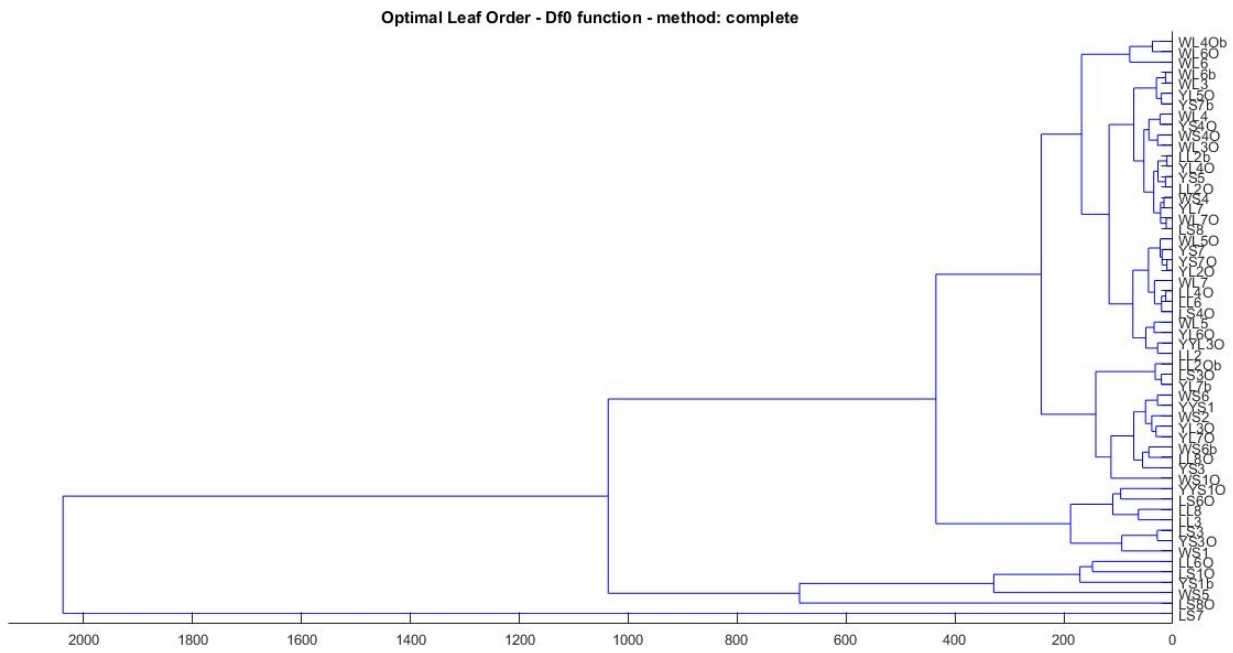


Figura B.7: *Dendrogramma: DF0 - complete*

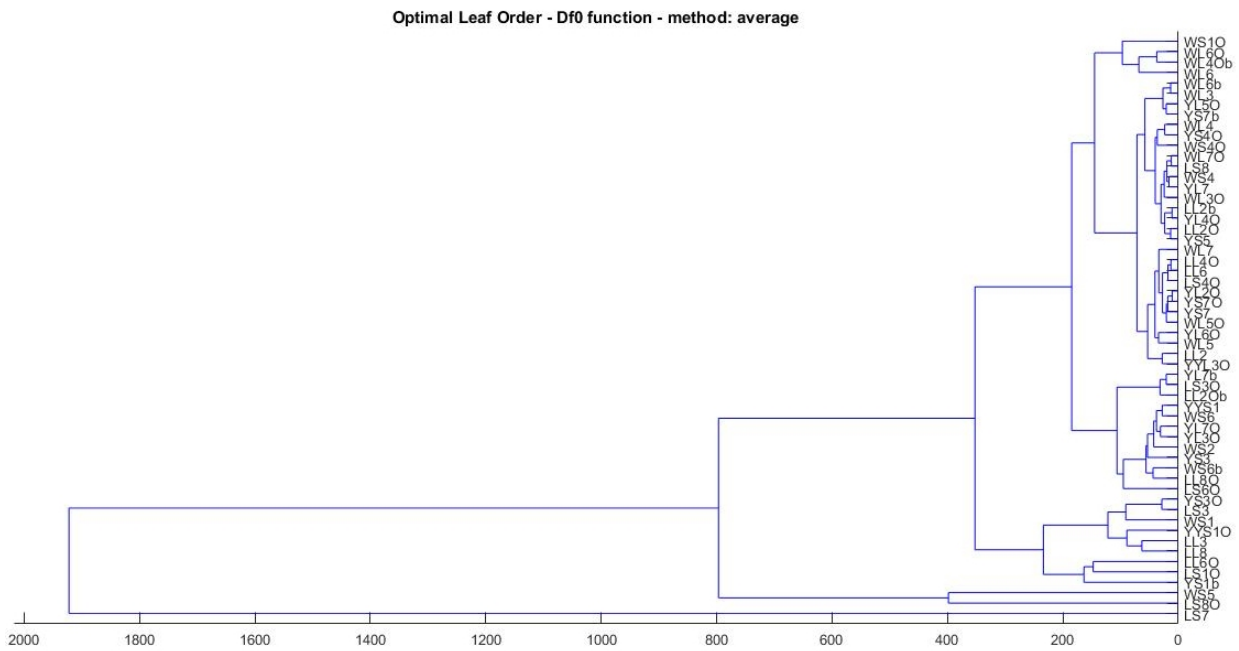
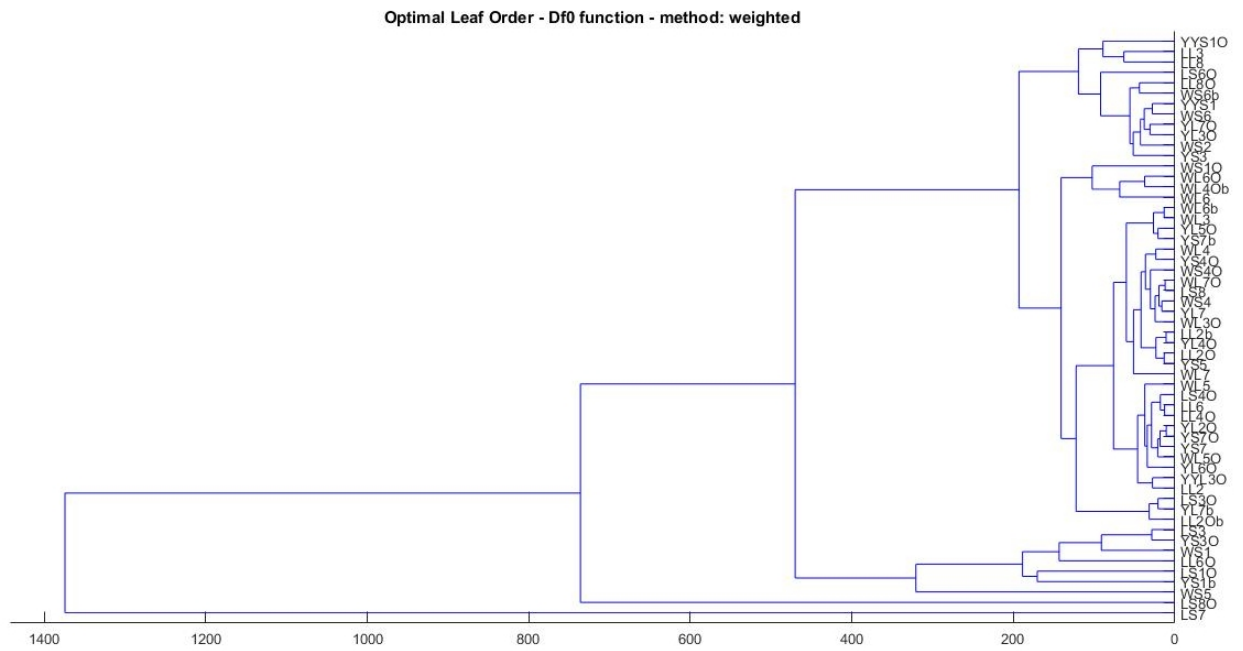
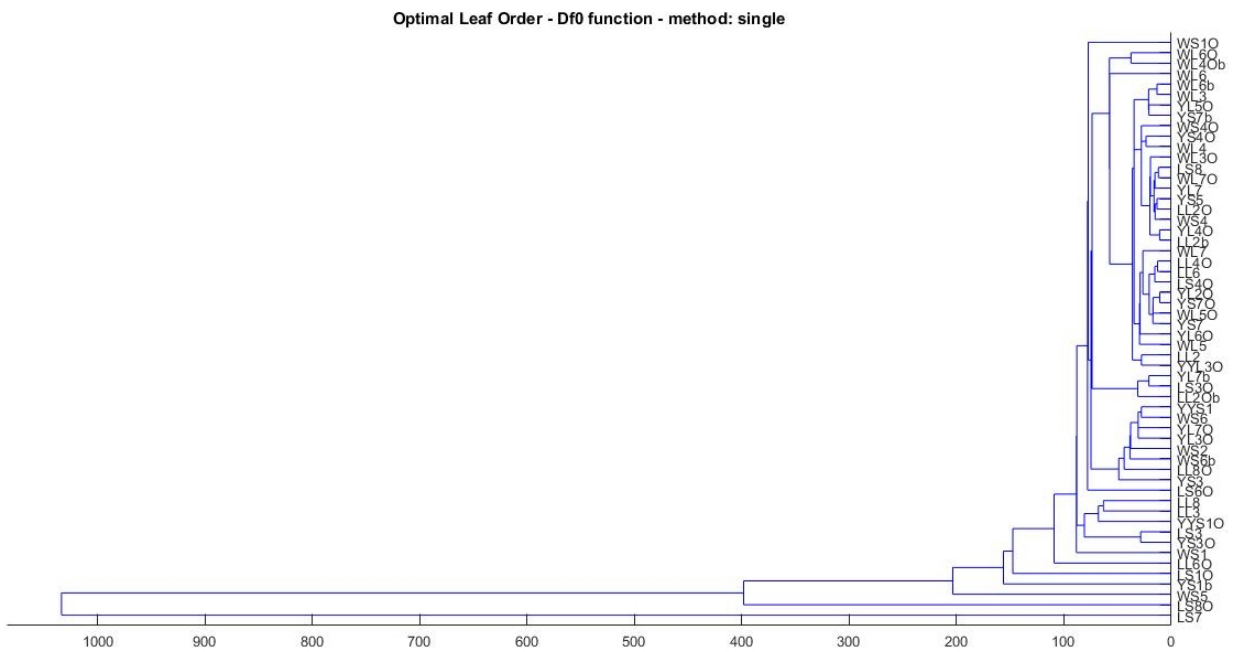
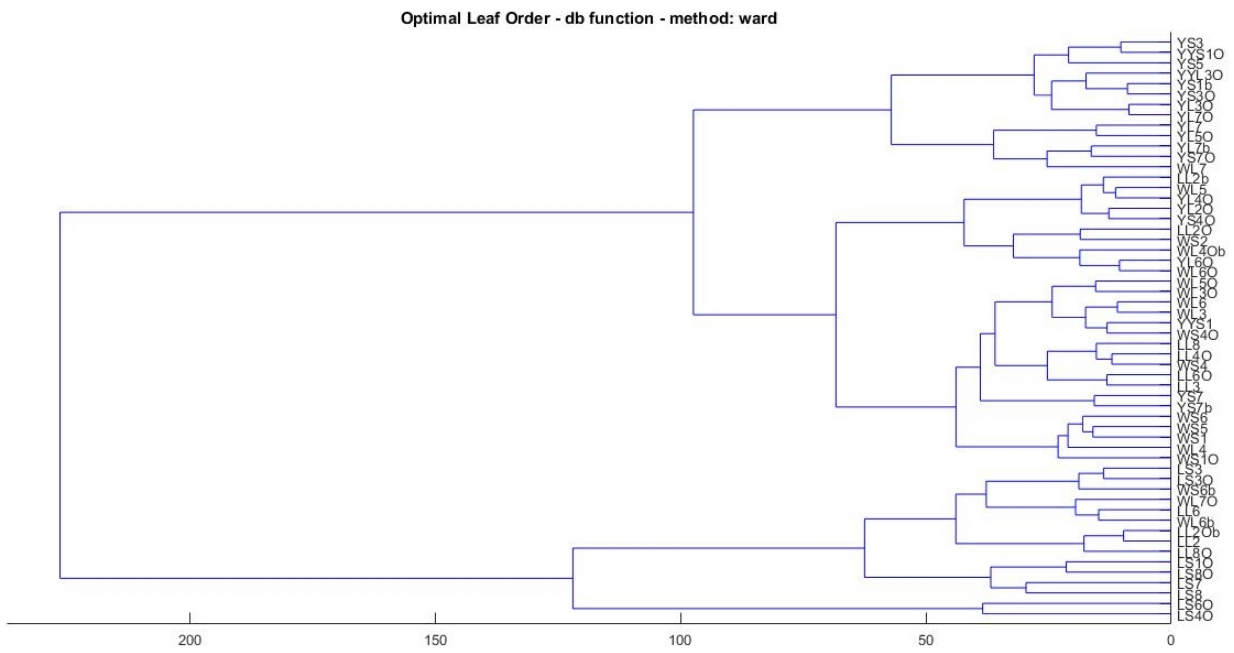
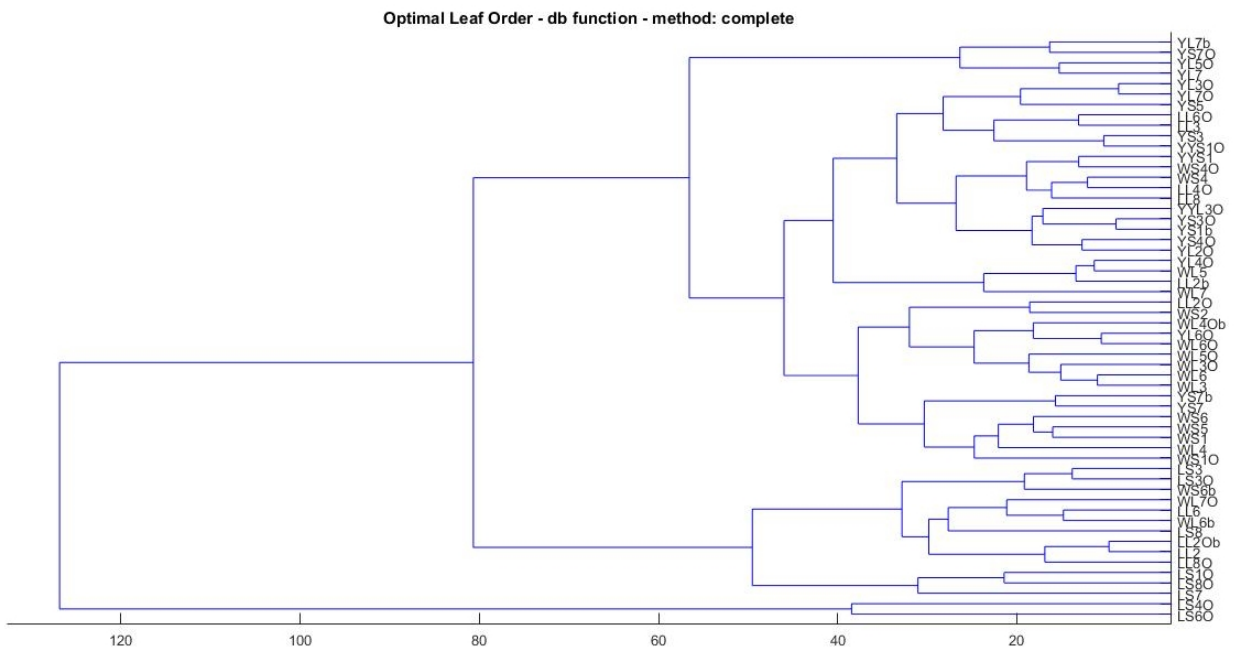
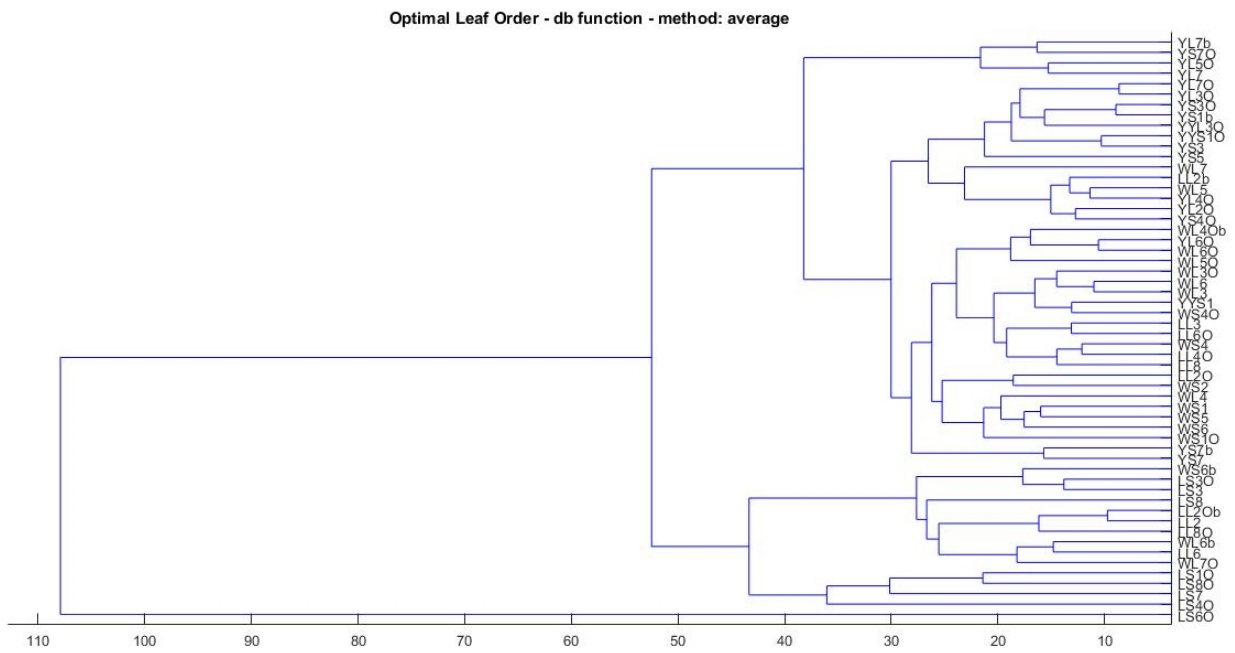
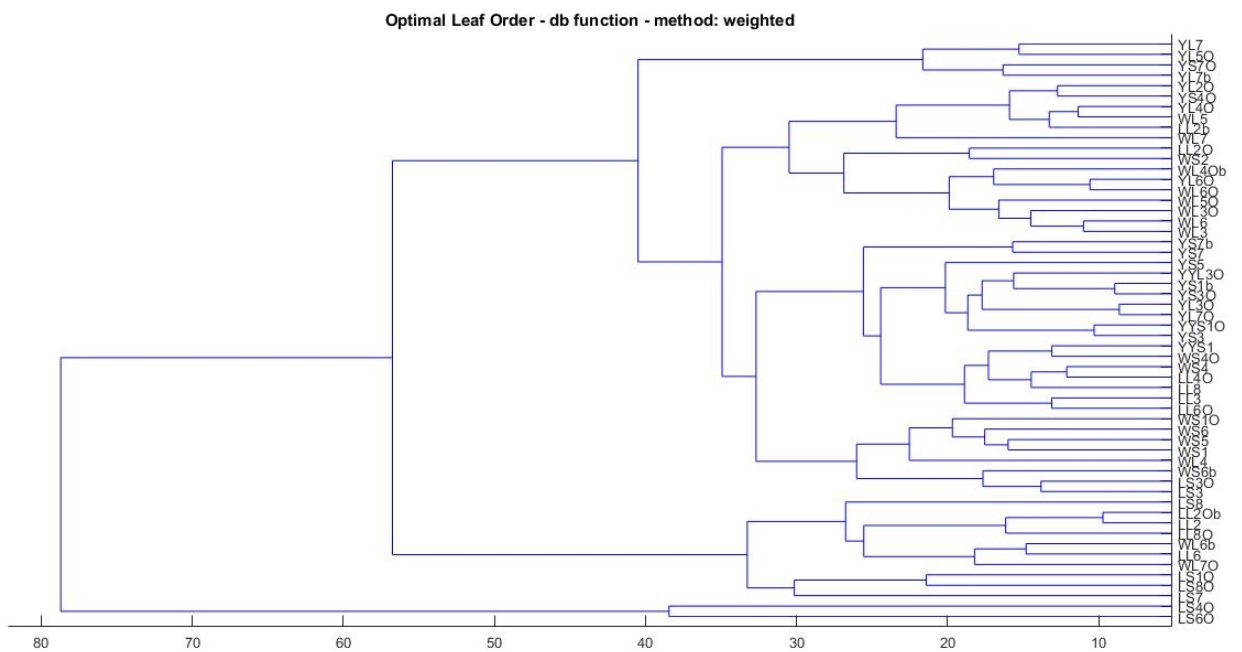
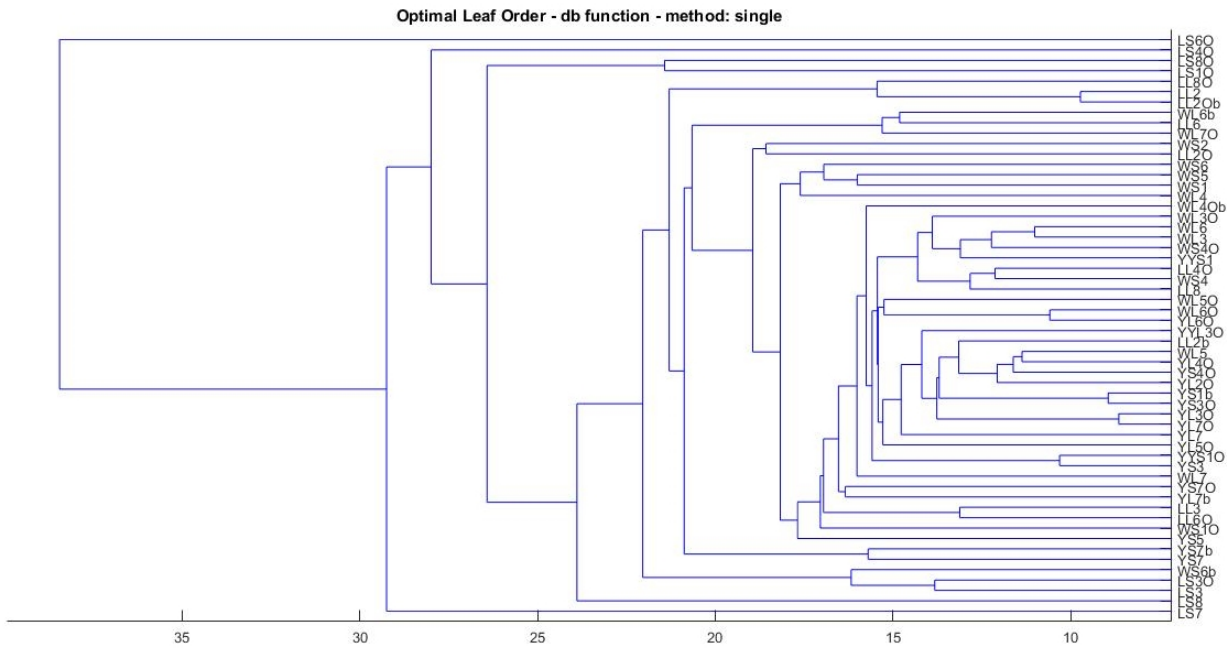
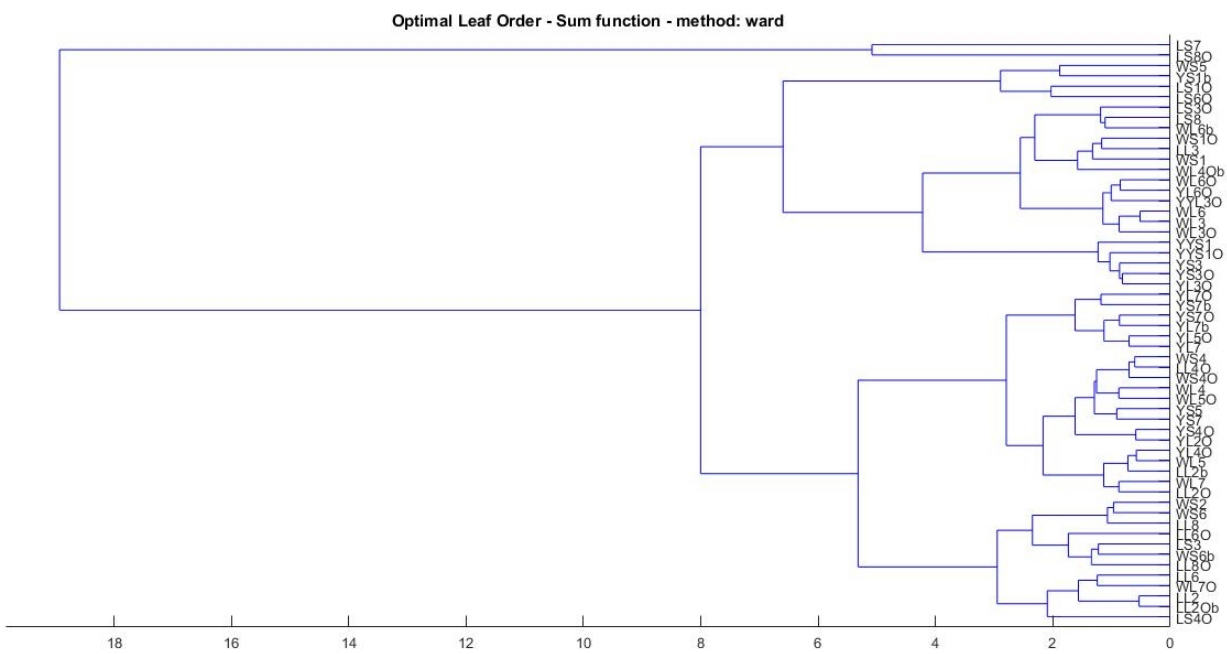


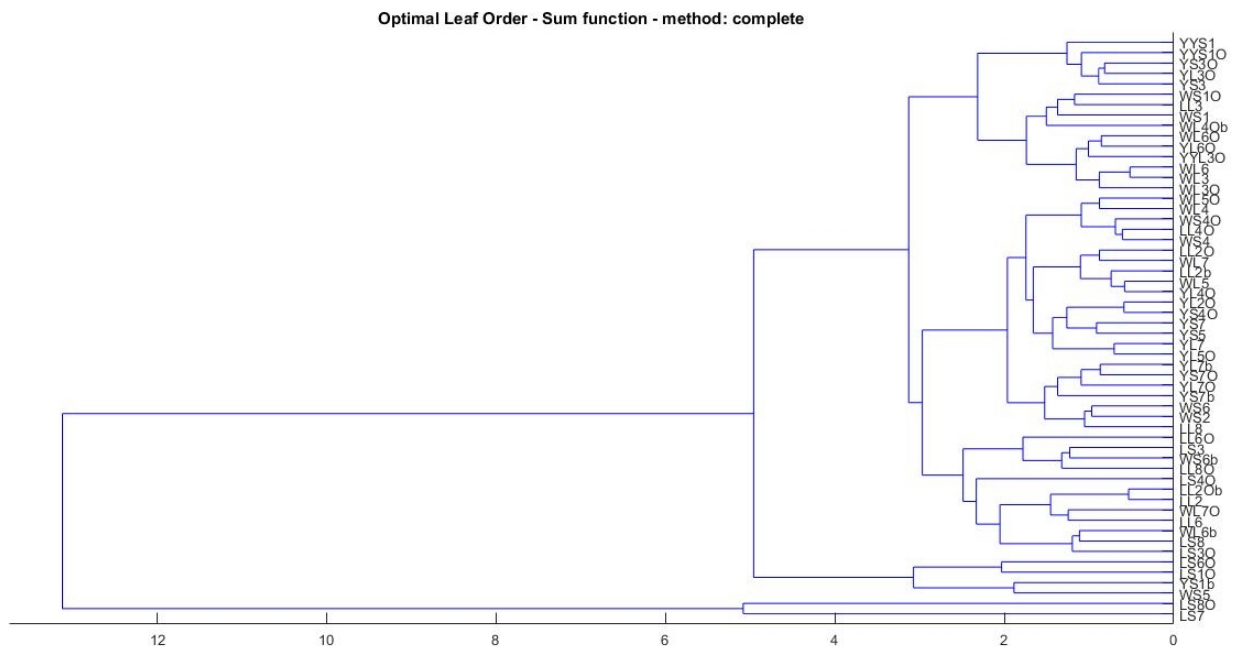
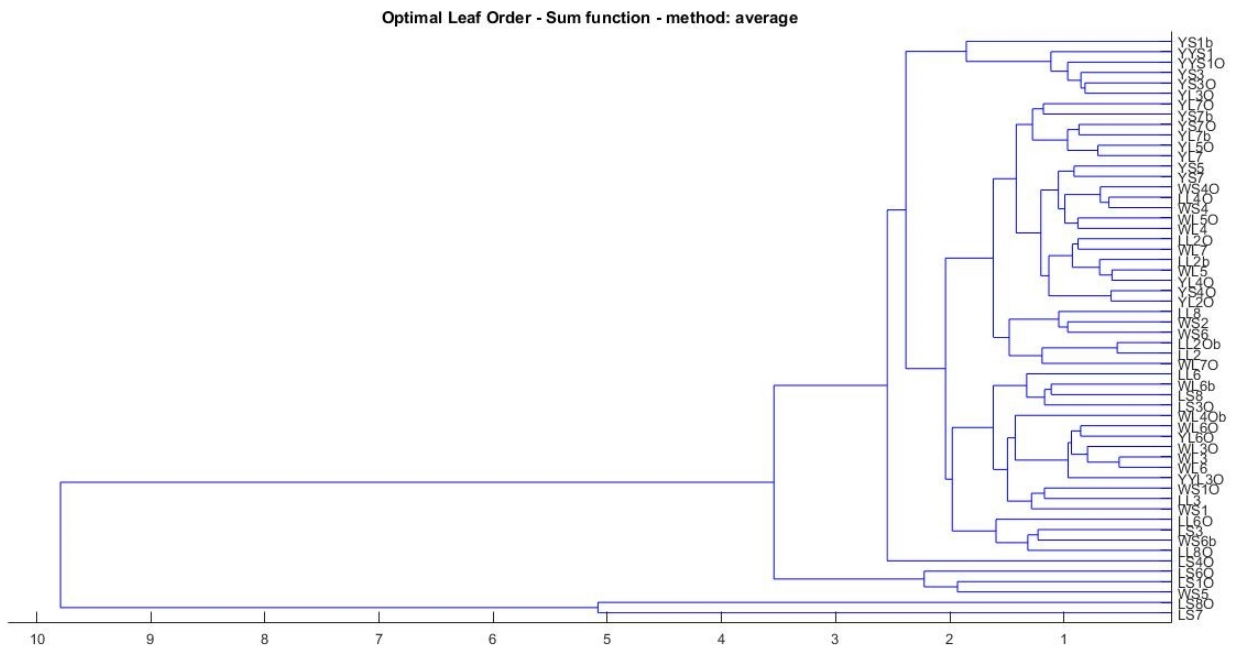
Figura B.8: *Dendrogramma: DF0 - average*

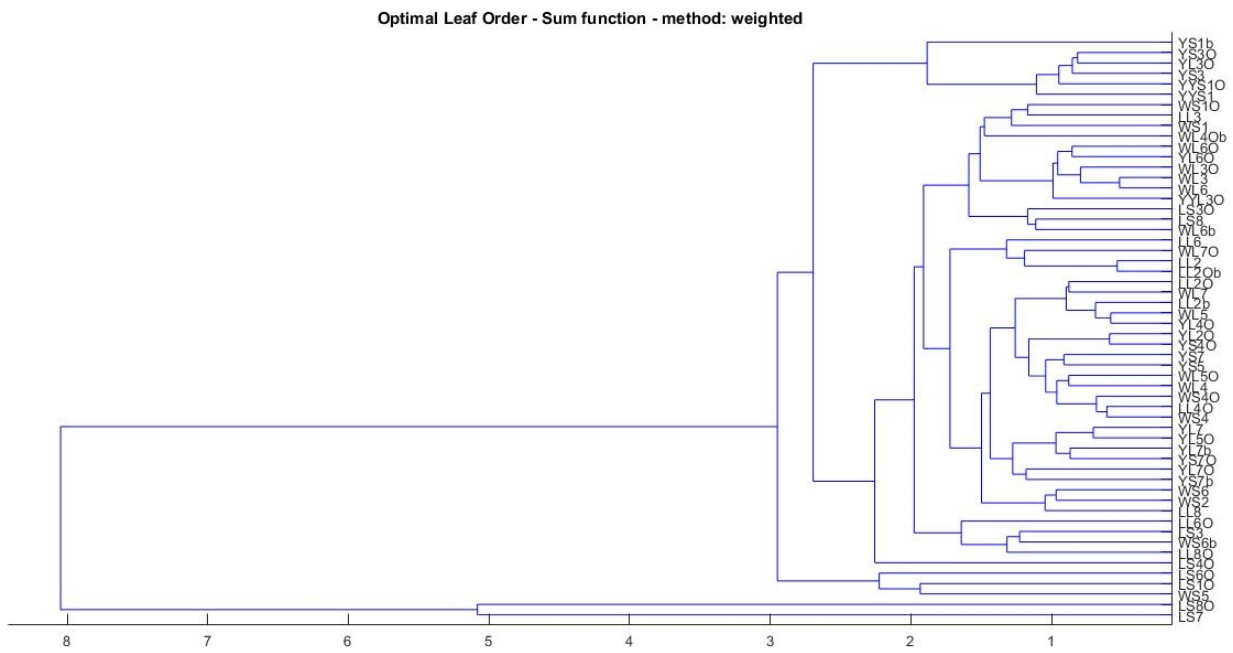
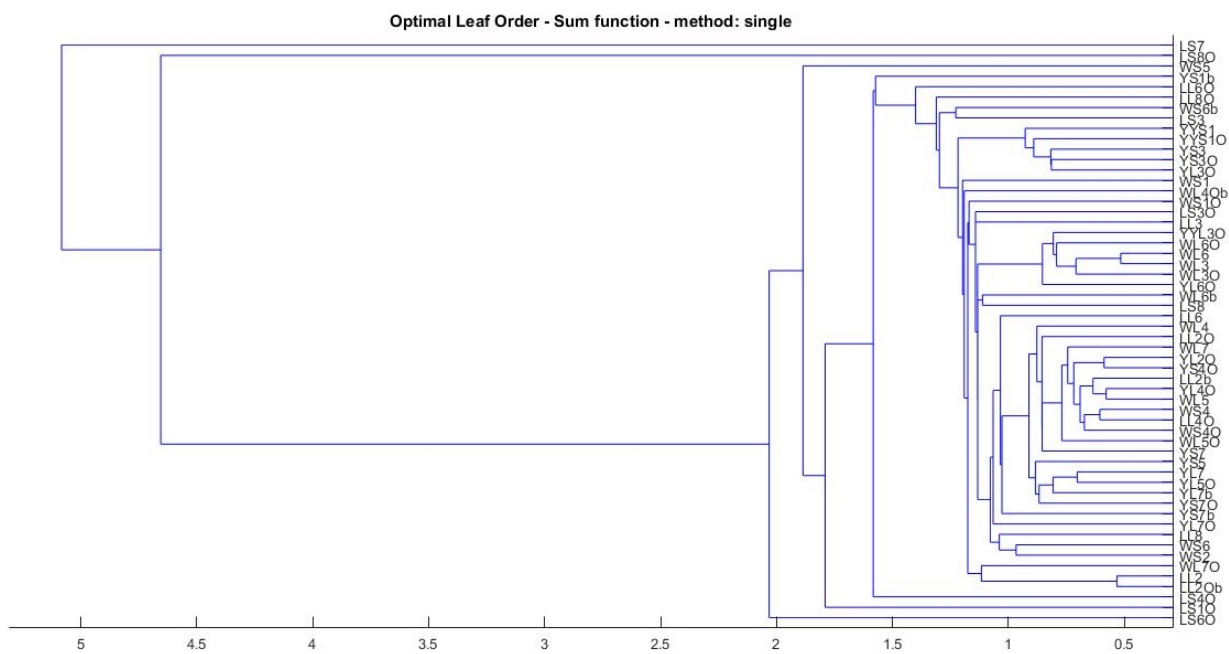
Figura B.9: *Dendrogramma: DF0 - weighted*Figura B.10: *Dendrogramma: DF0 - single*

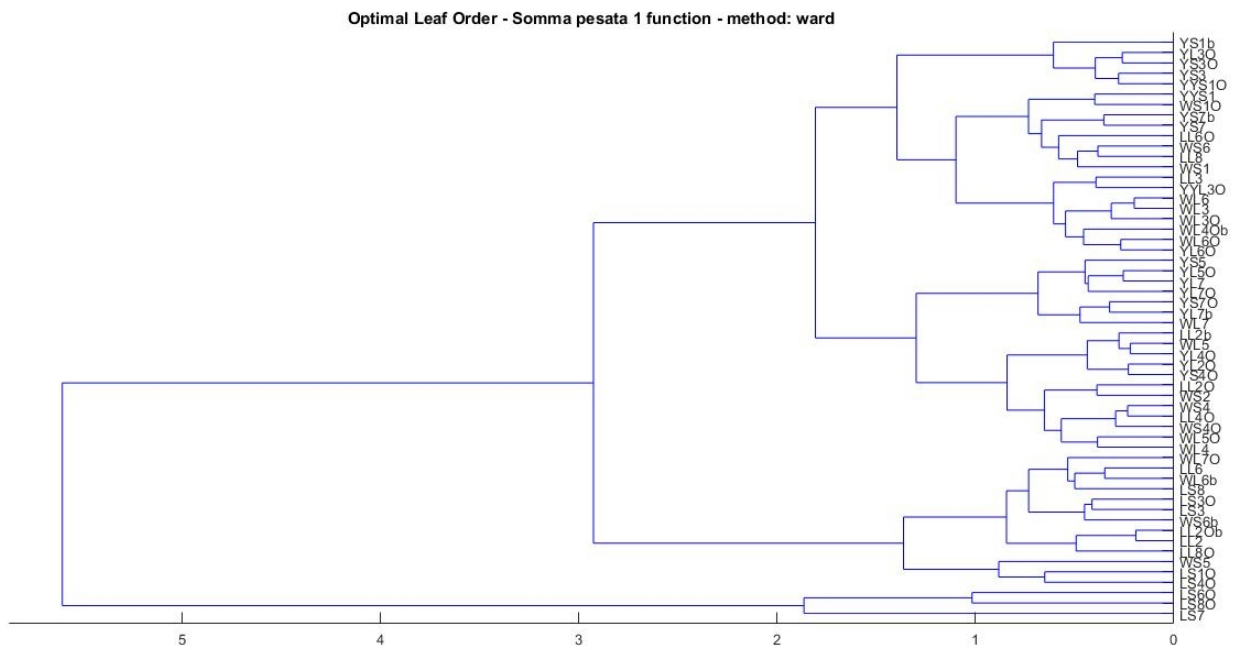
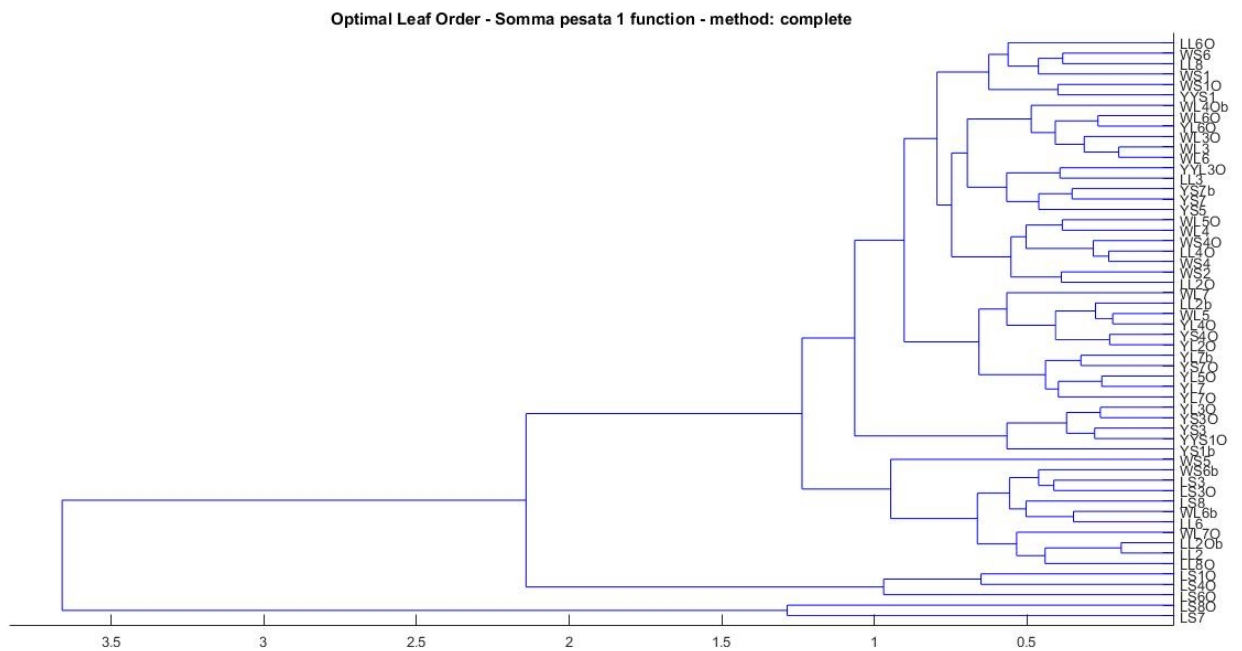
Figura B.11: *Dendrogramma: Volume - ward*Figura B.12: *Dendrogramma: Volume - complete*

Figura B.13: *Dendrogramma: Volume - average*Figura B.14: *Dendrogramma: Volume - weighted*

Figura B.15: *Dendrogramma: Volume - single*Figura B.16: *Dendrogramma: Somma - ward*

Figura B.17: *Dendrogramma: Somma - complete*Figura B.18: *Dendrogramma: Somma - average*

Figura B.19: *Dendrogramma: Somma - weighted*Figura B.20: *Dendrogramma: Somma - single*

Figura B.21: *Dendrogramma: Somma pesata 1 - ward*Figura B.22: *Dendrogramma: Somma pesata 1 - complete*

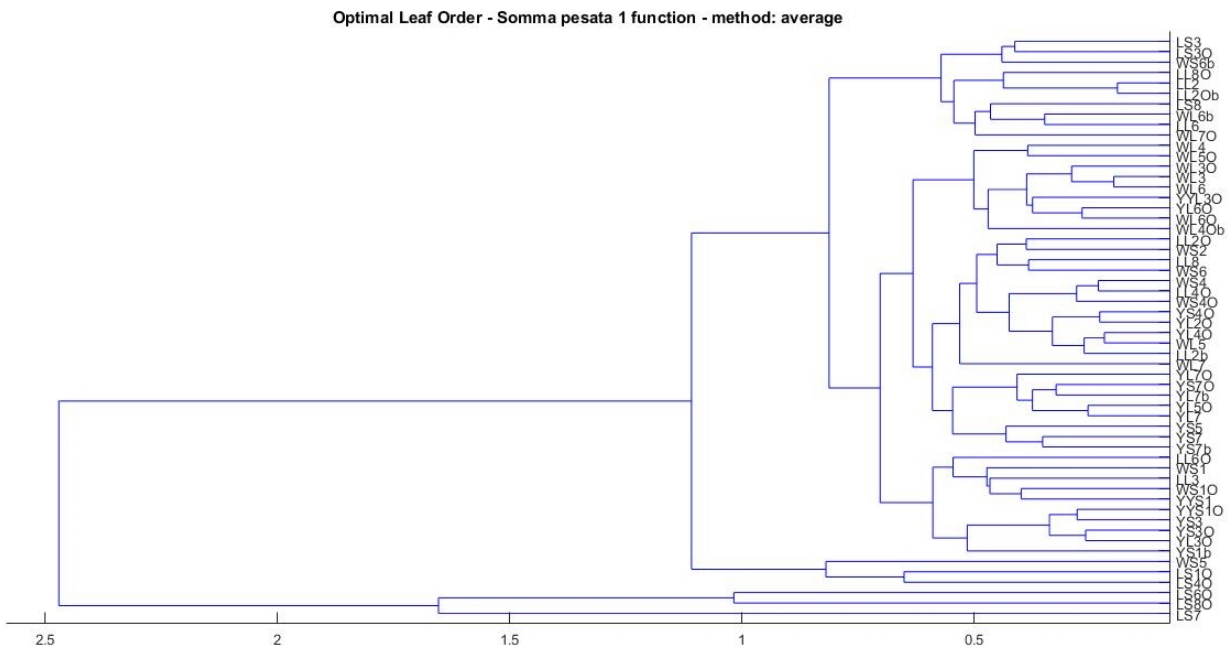


Figura B.23: *Dendrogramma: Somma pesata 1 - average*

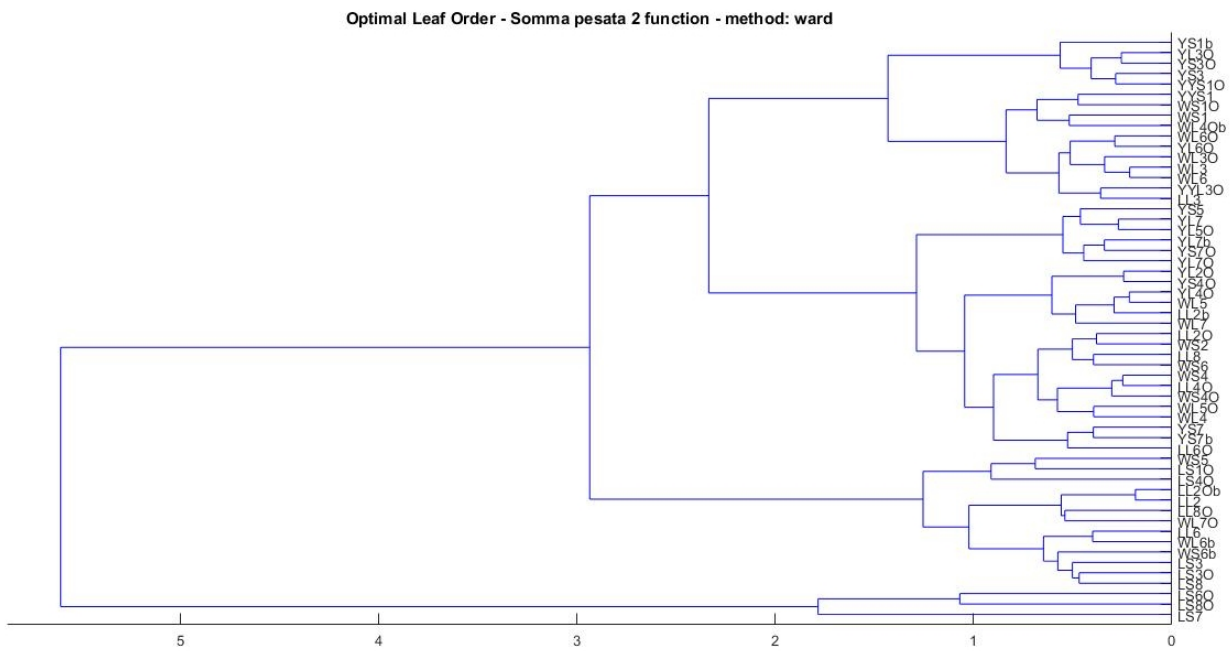
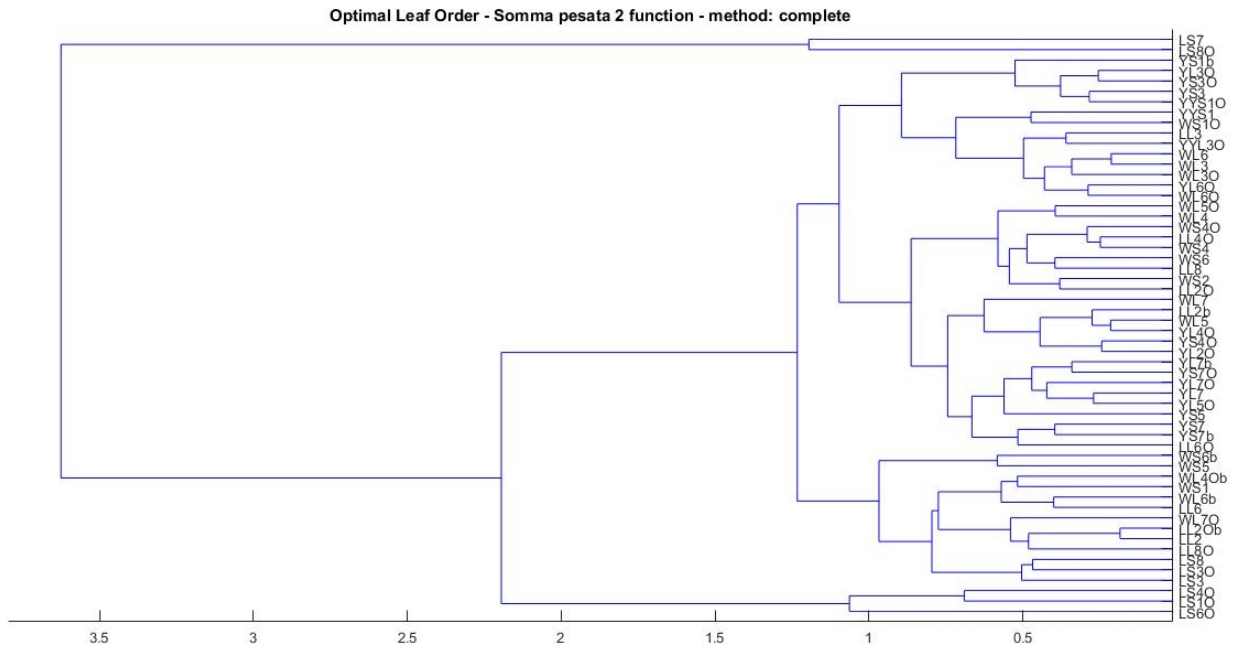
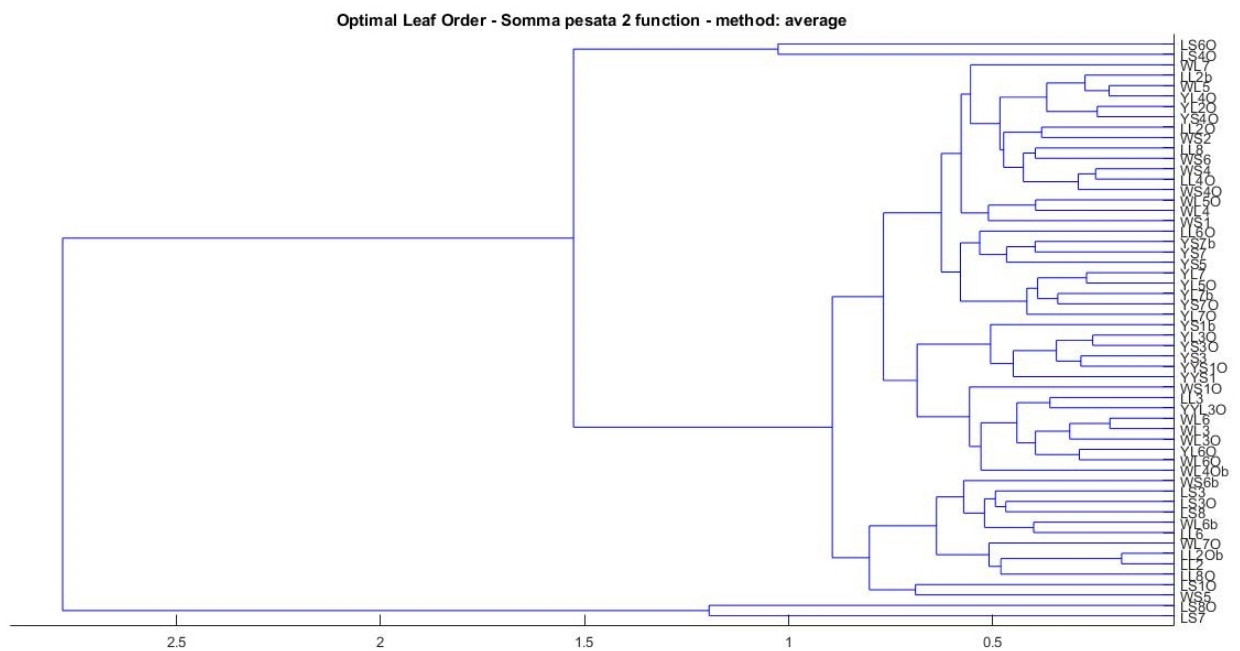


Figura B.24: *Dendrogramma: Somma pesata 2 - ward*

Figura B.25: *Dendrogramma: Somma pesata 2 - complete*Figura B.26: *Dendrogramma: Somma pesata 2 - average*

Appendice C

Confronto diagrammi di persistenza

Innanzitutto esponiamo grafici in cui vedremo raffigurato il confronto tra le posizioni delle cornerlines dei dati delle tre classi per ciascuna delle tre funzioni filtranti. Di seguito presenteremo i grafici di confronto delle posizioni delle cornerlines specificatamente per i dati della classe delle liste per ciascuna delle tre funzioni filtranti ed infine elenchiamo i grafici che attuano un confronto tra le disposizioni dei cornerpoints degli elementi di ogni classe di dato per ogni funzione filtrante.

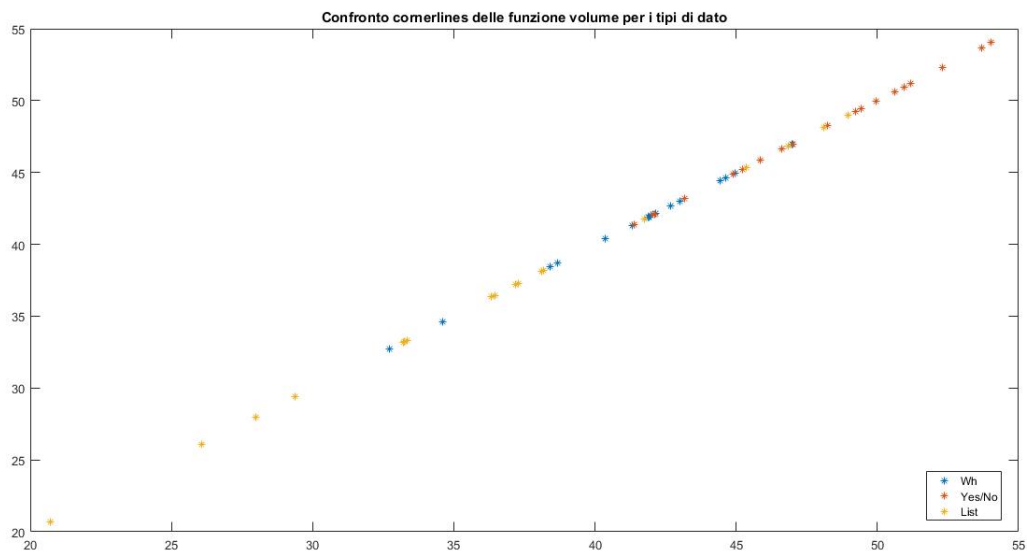


Figura C.1: *Cornerlines secondo la funzione filtrante del volume per tutte le classi di dato*

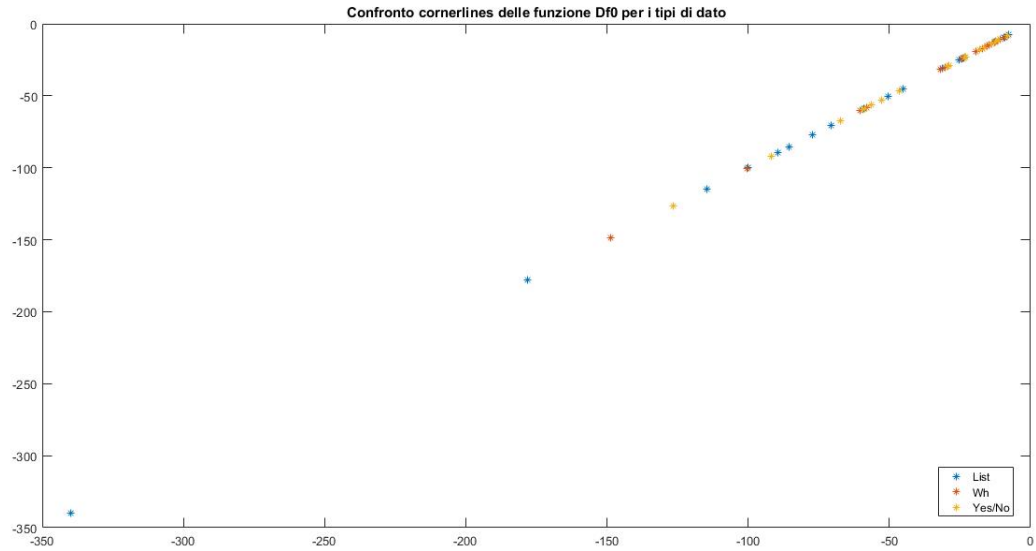


Figura C.2: *Cornerlines secondo la funzione filtrante di derivata di f_0 per tutte le classi di dato*

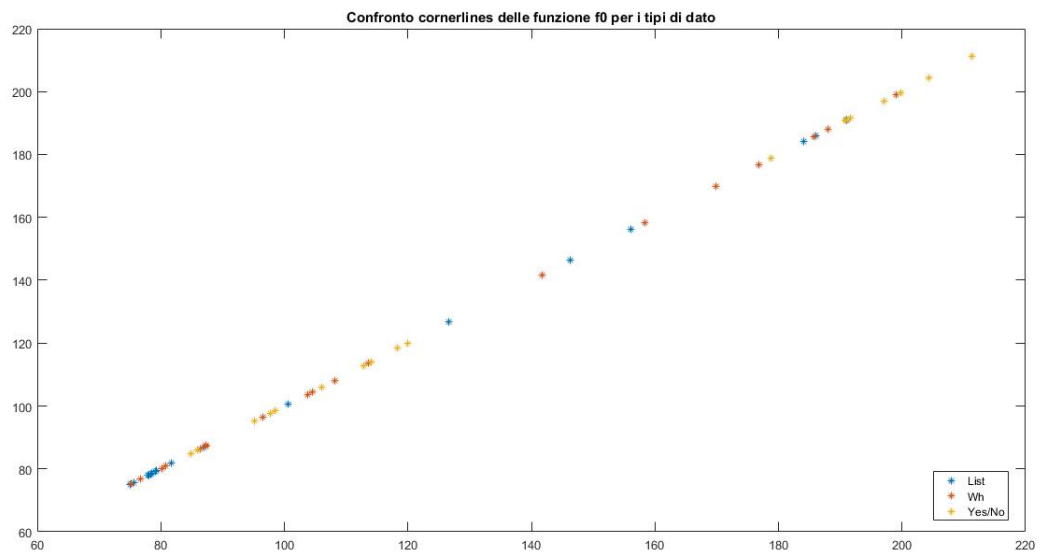


Figura C.3: *Cornerlines secondo la funzione filtrante di f_0 per tutte le classi di dato*

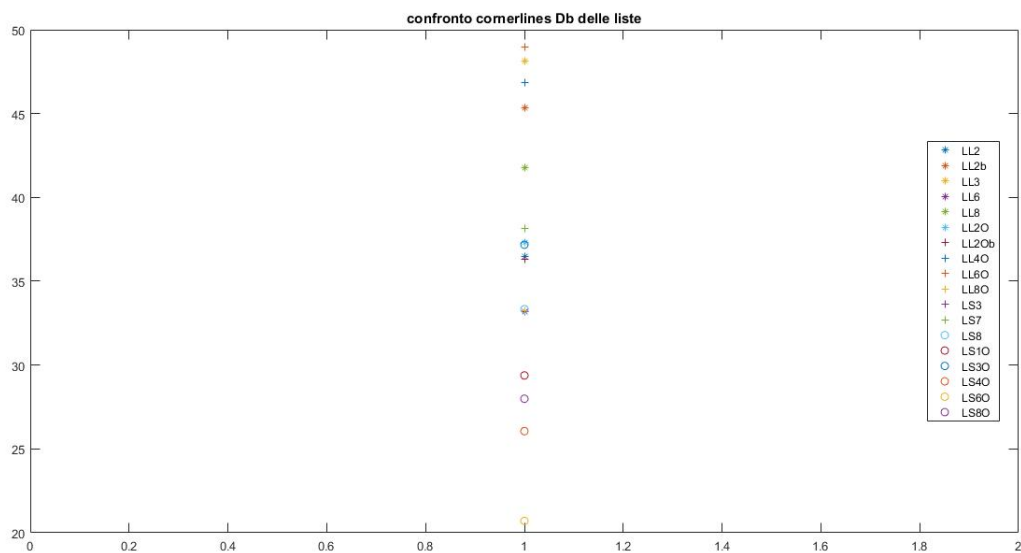


Figura C.4: Cornerlines secondo la funzione filtrante del volume per liste

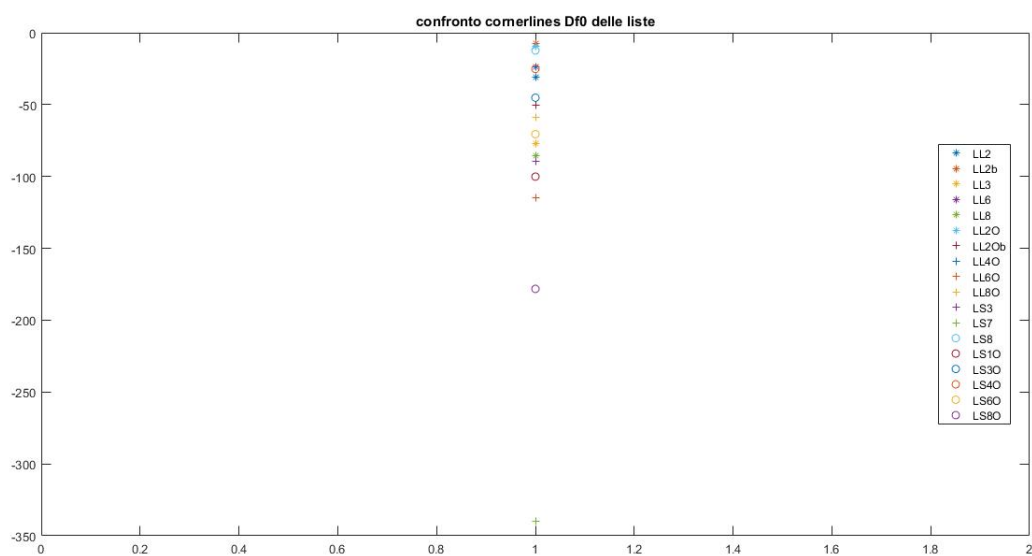


Figura C.5: Cornerlines secondo la funzione filtrante di derivata di f_0 per liste

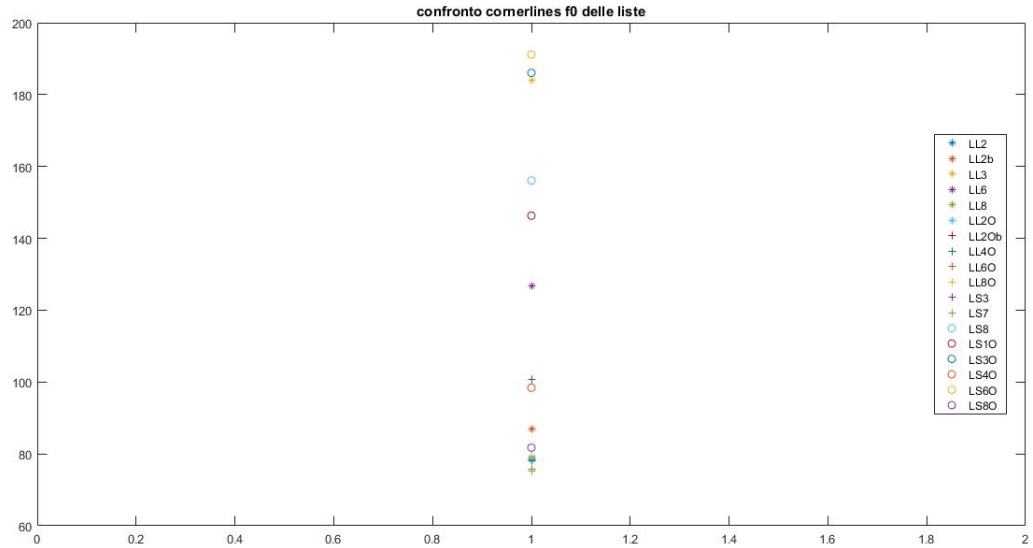


Figura C.6: *Cornerlines secondo la funzione filtrante di f_0 per liste*

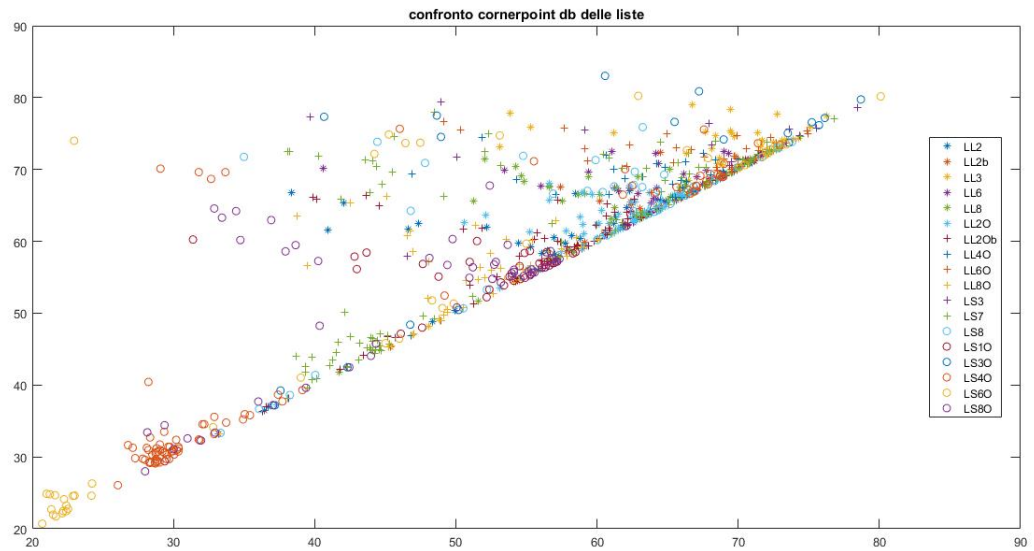


Figura C.7: *Cornerpoints secondo la funzione filtrante del volume per liste*

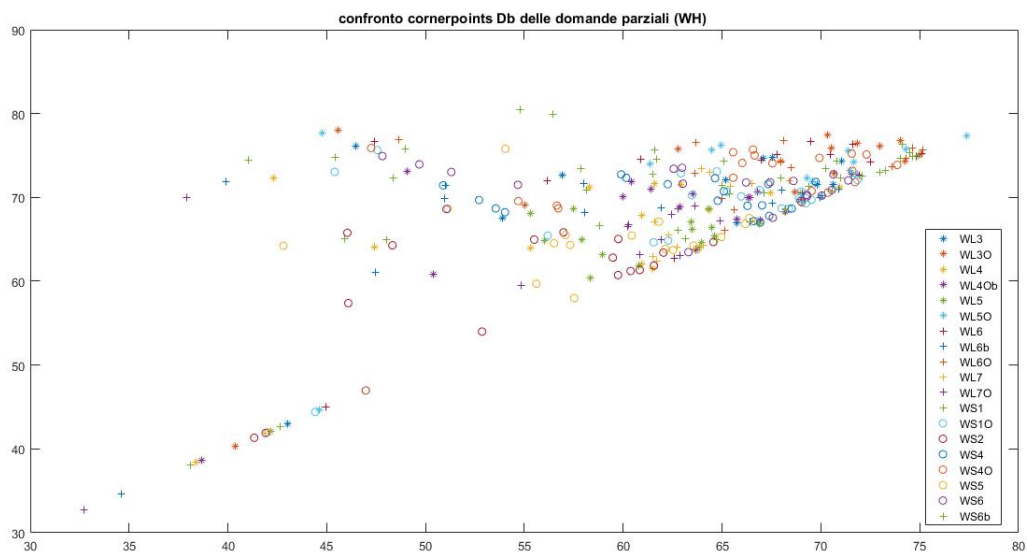


Figura C.8: Cornerpoints secondo la funzione filtrante del volume per domande parziali

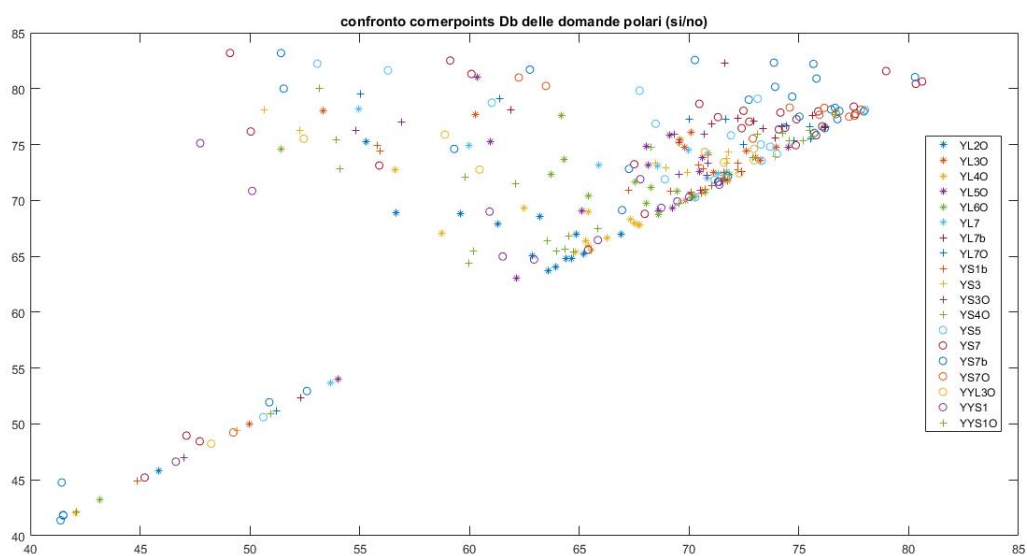


Figura C.9: Cornerpoints secondo la funzione filtrante del volume per domande polari

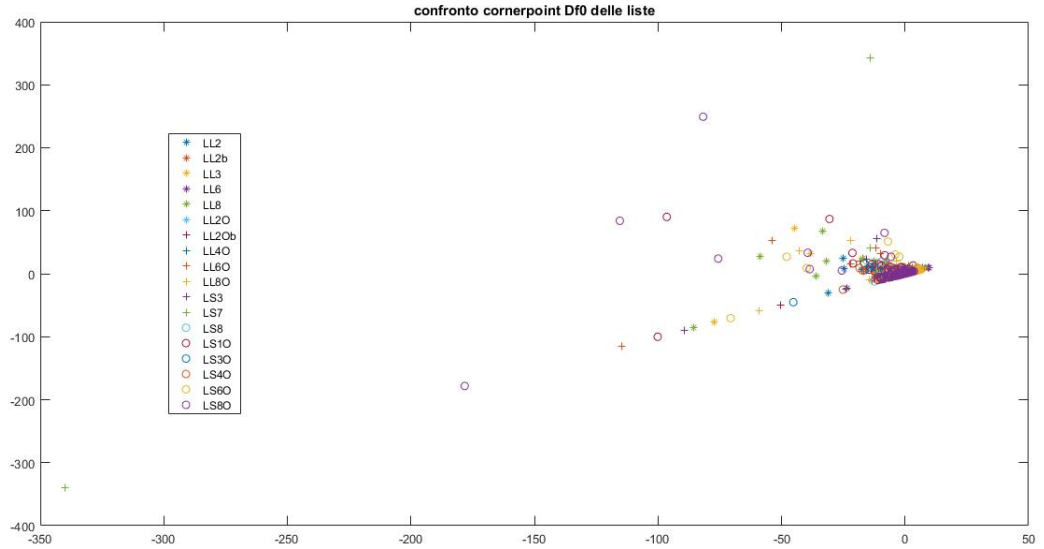


Figura C.10: *Cornerpoints secondo la funzione filtrante di derivata di f_0 per liste*

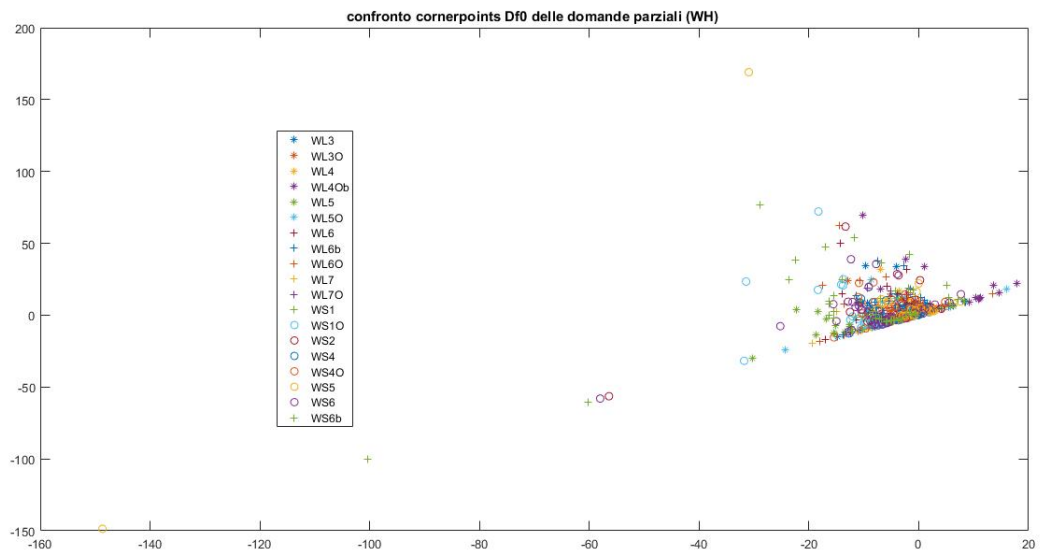


Figura C.11: *Cornerpoints secondo la funzione filtrante di derivata di f_0 per domande parziali*

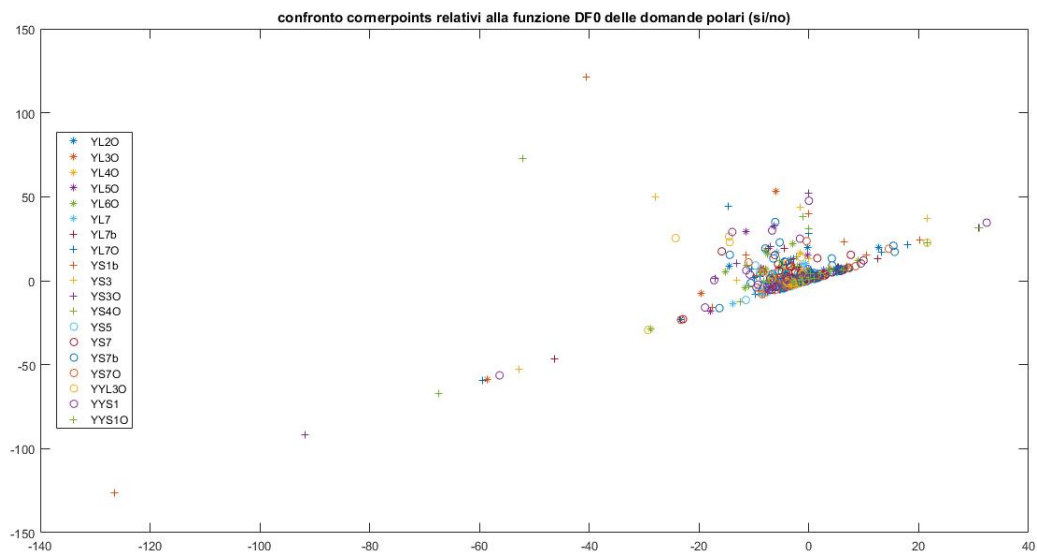


Figura C.12: *Cornerpoints secondo la funzione filtrante di derivata di f_0 per domande polari*

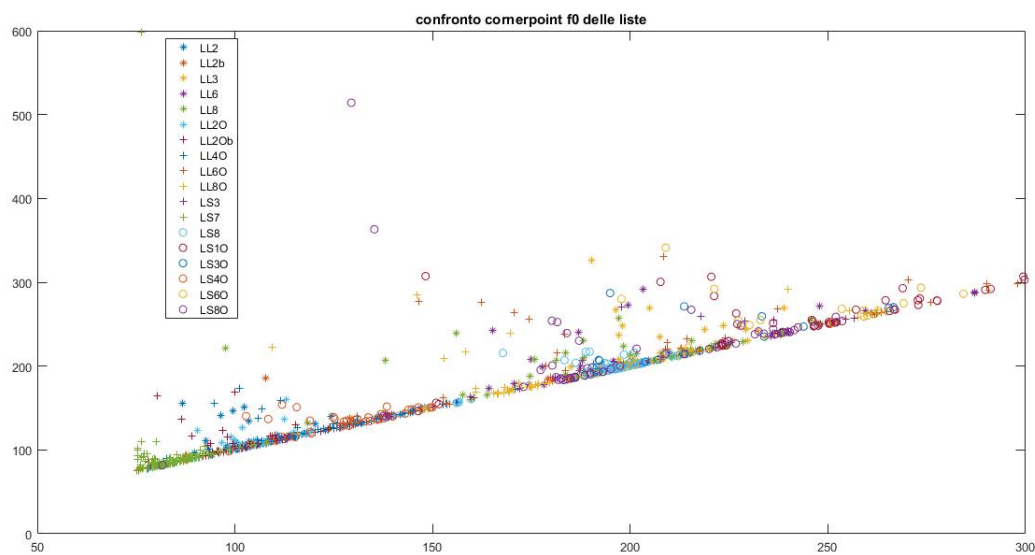


Figura C.13: *Cornerpoints secondo la funzione filtrante di f_0 per liste*

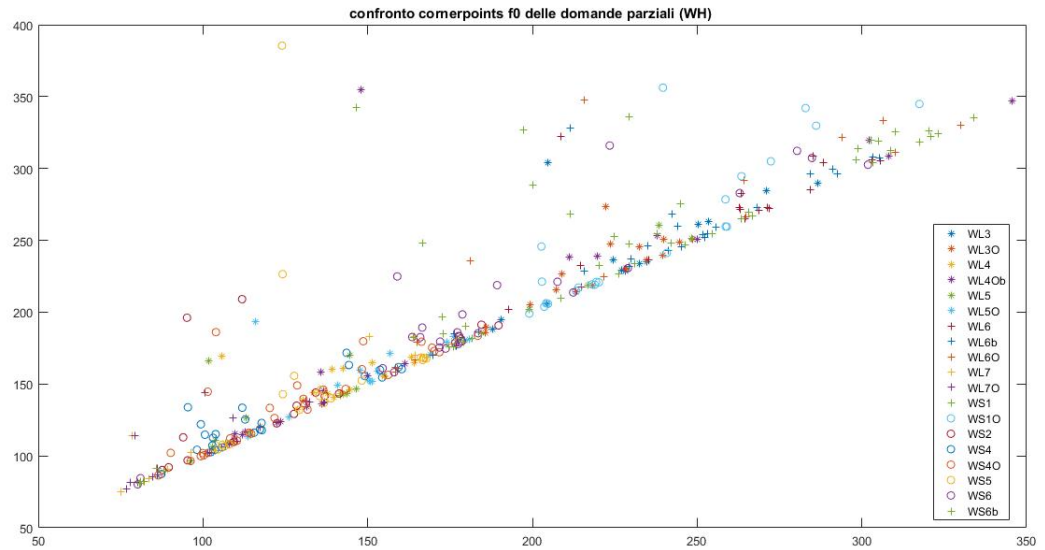


Figura C.14: *Cornerpoints secondo la funzione filtrante di f_0 per domande parziali*

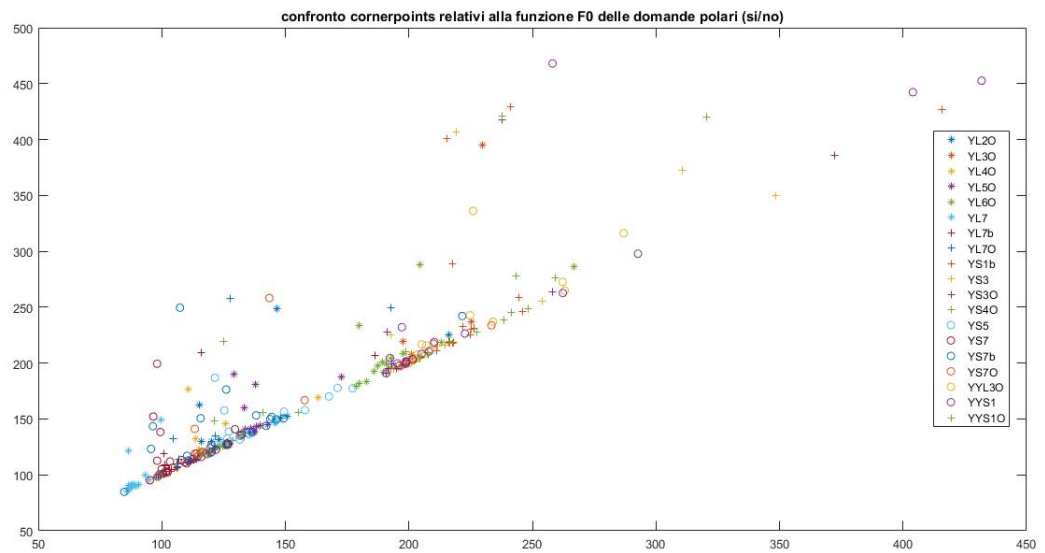


Figura C.15: *Cornerpoints secondo la funzione filtrante di f_0 per domande polari*

Bibliografia

- [1] M. Ferri, ‘*Persistent topology for natural data analysis - A survey*’, (2017). In: Holzinger A., Goebel R., Ferri M., Palade V. (eds) *Towards Integrative Machine Learning and Knowledge Extraction*, LNAI, 10344 (2017), 117-133.
- [2] M. Ferri, ‘*L’incredibile ubiquità della topologia persistente*’, <http://maddmaths.simai.eu/divulgazione/focus/lincredibile-ubiquita-della-topologia-persistente> , (2015-2016).
- [3] M. d’Amico, P. Frosini, and C. Landi. ‘*Natural pseudo-distance and optimal matching between reduced size functions*’, *Acta Applicandae Mathematicae*, 109(2) (2010), 527-554.
- [4] Herbert Edelsbrunner, John Harer, ‘*Computational Topology - An Introduction*’, Departments of Computer Science and Mathematics, Duke University, 2008.
- [5] [http://www.treccani.it/enciclopedia/tratti-soprasegmentali\(Enciclopedia-dell’Italiano\)](http://www.treccani.it/enciclopedia/tratti-soprasegmentali(Enciclopedia-dell’Italiano)).
- [6] Andrea Marini, ‘*Elementi di psicolinguistica generale*’, Springer Verlag, 2001.
- [7] <https://it.wikipedia.org/wiki/Fonazione>.
- [8] <https://it.wikipedia.org/wiki/Suono>.
- [9] [http://www.treccani.it/enciclopedia/fonetica-acustica-nozioni-e-termini-di\(Enciclopedia-dell’Italiano\)](http://www.treccani.it/enciclopedia/fonetica-acustica-nozioni-e-termini-di(Enciclopedia-dell’Italiano)).
- [10] Wladimiro De Colle, ‘*PRAAT Fondamenti teorico-pratici di utilizzo clinico*’, Università di Padova, 2017.
- [11] Marina Nespor, ‘*Le strutture del linguaggio - Fonologia*’, Il Mulino, Bologna, 1994.
- [12] Gili Fivela Barbara, Avesani Cinzia, Barone Marco, Bocci Giuliano, Crocco Claudia, D’Imperio Mariapaola, Giordano Rosa, Marotta Giovanna, Savino Michelina and Sorianello Patrizia (in press), ‘*Varieties of Italian and their intonational phonology*’

in S. Frota and P. Prieto (eds), *Intonational phonology of the regional varieties of Italian*, Oxford, Oxford University Press.

- [13] Marco Barone, 'Intonation in Pescara: the case of two languages in contact', Tesi di dottorato in andamento, Univerisità del Salento - Universitat Pompeu Fabra , 2016.
- [14] Mary E. Beckman, Gayle M. Ayers, *Guidelines for ToBI Labelling*, 1994, <http://www.speech.cs.cmu.edu/tobi/ToBI.0.html>
- [15] Loredana Cerbara, *Metodi di classificazione*, UniRoma.
- [16] <https://it.mathworks.com/help/stats/linkage.html> (Documentazione online MatLab).

Ringraziamenti

A conclusione di questo percorso di laurea magistrale sono doverosi alcuni ringraziamenti.

In primis un ringraziamento sentito al mio relatore, il Professore Massimo Ferri per la pazienza mostrata e per l'opportunità datami di lavorare ad un progetto così interessante da cui ho imparato moltissimo. Ringrazio anche il Professore Marco Barone, correlatore di questo lavoro di tesi, per le puntuali ed interessanti spiegazioni che mi ha fornito sul mondo della linguistica. Inoltre ringrazio il Dottore Ivan Tomba che pazientemente ha seguito il lato computazionale del progetto con infinita gentilezza. Infine sono grata a tutti gli altri professori con cui sono venuta a contatto durante questo secondo ciclo di laurea magistrale; da ognuno di loro ho tratto insegnamenti non solo accademici ma anche personali.

Un grandissimo ringraziamento va poi ai miei genitori che mi hanno consentito di percorrere questa strada continuando a sostenermi nonostante le difficoltà. Ringrazio Giacomo, mio fratello, perché è una persona meravigliosa e mi ha sostenuto in questi anni in ogni momento con consigli personali e tecnici. Un'altra persona meravigliosa che mi ha sostenuto e capito è la mia amica Jessica, che ringrazio per gli infiniti momenti di pazienza nel sopportare i miei deliri. Ringrazio anche i compagni di corso e di studio che mi hanno accompagnato in questi anni con particolare menzione a Ilaria, la quale, nonostante le distanze è sempre stata punto di riferimento e di confronto nonché di sostegno. Infine un ringraziamento agli amici e amiche che hanno fatto parte dei necessari (e divertentissimi) momenti di svago di questi anni, dalle serate fuori alle serate di canto con amici nuovi e di vecchia data.

La conclusione di questo percorso è motivo per me di grande soddisfazione e orgoglio che voglio condividere con tutti voi.
Grazie.