Università degli Studi di Padova

Dipartimento di Neuroscienze - DNS

Corso di Laurea in Tecniche Audioprotesiche

Presidente Prof. Gino Marioni

TESI DI LAUREA

Analisi accurata nel dominio della frequenza di alcuni fonemi rappresentati nel banana speech

Relatore: Ing. Antonio Franco Selmo

Laureando: Massimiliano Gatto

ANNO ACCADEMICO 2021/22

Indice

Riassunto (Abstract)

Lo studio proposto prende come punto di riflessione il banana speech, uno dei metodi utilizzati per la rappresentazione dei fonemi sull'audiogramma che viene descritto come "regione del parlato".

Osservando un qualsiasi grafico rappresentante il BS, si può notare come ogni suono sia indicato da un singolo punto. Tenendo conto che il BS è un diagramma bidimensionale, dove sull'asse X sono indicate I valori di frequenza e sull'asse Y sono indicati i valori di intensità, un singolo punto sembra indicare che il particolare suono a cui si riferisce sia costituito dalla sola componente frequenziale indicate dalla posizione del punto sull'asse X. In realtà, il punto indicato corrisponde, da un punto di vista frequenziale, al valore di frequenza corrispondente alla sola componente di ampiezza maggiore rispetto a tutte le altre, tra tutte le componenti che, nel loro insieme, realizzano il fonema.

In realtà, un fonema è ottenuto dalla somma di tante componenti armoniche (teorema di Fourier) in aggiunta alla componente fondamentale di maggior ampiezza. Se il fonema fosse generato solamente dall'unica componente indicata dal corrispondente punto posizionato sul BS, risulterebbe un suono corrispondente ad un tono puro, quindi assolutamente non interpretabile.

A questo punto era necessario capire fino a che intensità, al di sotto di quella relative alla componente di maggior ampiezza, sia necessario prendere in considerazioni le diverse formanti che, nel loro insieme, realizzano il fonema. Sono stati fatti dei test su alcuni fonemi e si è visto che prendendo in considerazione le componenti con un'ampiezza fino a -24 dB rispetto alla componente fondamentale, il fonema risulta comprensibile in maniera accettabile (test effettuati su soggetti normoacusici) anche se la sua riproduzione non risulta perfetta. Prendere in considerazione un range dell'ampiezza di soli 12 dB, porta ad un suono che risulta difficilmente interpretabile, se non per pochi fonemi.

L'analisi svolta, limitata nel dominio della frequenza, per una serie di fonemi (non tutti quelli mostrati nel BS) è stata effettuata su fonemi prodotti da due soggetti diversi (un maschio e una femmina adulta di nazionalità italiana, con una buona pronuncia, operatori nell'insegnamento a livello di scuola superior/università), prendendo in considerazione proprio un range di 24 dB tra la componente di ampiezza massima e quella di ampiezza minima.

La metodologia utilizzata

I fonemi analizzati sono stati ottenuti da una registrazione con apparecchiature di qualità elevate, utilizzando microfoni professionali da studio, preamplificatori microfonici a bassissimo rumore di fondo con una banda passante di oltre 200 kHz, un sistema di conversione analogico/digitale costituito da un registratore Tascam DV RA1000 HD impostato su 192000 campioni/s e 24 bit di risoluzione. L'analisi frequenziale è stata realizzata con adeguato software in grado di effettuare la trasformata di Fourier su un numero finito di campioni, fino ad un massimo di oltre 65000 campioni, consentendo una analisi con una dinamica fino a 80 – 90 dB.

Conclusioni

Analizzando diversi fonemi, con la metodologia descritta, si è visto come alcuni fonemi si estendano nel dominio della frequenza per oltre tre ottave. Ciò porta a rivedere l'indicazione di un fonema, nel dominio della frequenza, non come un singolo punto, bensì come un segmento. Interpretare un fonema o un suono come un'unica componente frequenziale (che corrisponde ad un tono puro) costituisce un'informazione assolutamente fuorviante. La diretta conseguenza sta nella struttura stessa del BS, dove ciascun fonema andrebbe indicato da una linea orizzontale, con una conseguente estensione del BS, lungo l'asse delle frequenze, molto maggiore di quella indicata di solito.

Introduzione

Nell'applicazione protesica, la fase che precede la scelta dell'apparecchio acustico da parte dell'audioprotesista è la raccolta dati, che nel suo insieme vede come uno dei pilastri fondamentali l'esame audiometrico. Determinata la soglia dell'udito residuo otteniamo un'informazione fondamentale ovvero l'impairment dato dall'ipoacusia, e quindi del livello di disabilità che questa porta nella comunicazione.

Audiometria

L'audiogramma è il grafico bidimensionale in cui riportiamo le misure effettuate con l'audiometro, il sistema composto da scheda audio e cuffie/casse tarate secondo norme specifiche CEI EN 60645. L'audiogramma è un grafico bidimensionale, in ascissa troviamo frequenza da 125 Hz a 8000 Hz e in ordinata l'intensità percepita in HL (Hearing Level).

Nel test di conduzione aerea, l'audiologo presenta toni puri utilizzando l'audiometro con una gamma di frequenze da 125 Hz a 8000 Hz attraverso le cuffie. È importante notare che alcuni suoni vocali non rientrano in questa gamma, come ad esempio la frequenza fondamentale di un oratore maschile che potrebbe scendere al di sotto dei 60 Hz e il sibilo turbolento delle fricative del suono /s/ con frequenze superiori a 16.000 Hz.[2][3]

Sull'audiogramma, ogni risposta (per via aerea) è sistematicamente contrassegnata da simboli di croce (per l'orecchio sinistro) e a cerchio (per l'orecchio destro). Le soglie uditive di ciascun orecchio sono collegate da una linea retta.[2][3]

Lo studio proposto pone in esame il "banana speech", lo strumento noto nel campo dell'audiologia come "rappresentazione della regione del parlato". Questo strumento dovrebbe mostrare come il dominio in frequenza ed in intensità dei fonemi prende posto in tutto il mondo dei suoni, disegnando, all'interno dell'audiogramma, un riquadro che ricorda la forma di una "banana". Nella fig.1 di possiamo osservare un esempio di come viene inserito nell'audiogramma.

Banana speech

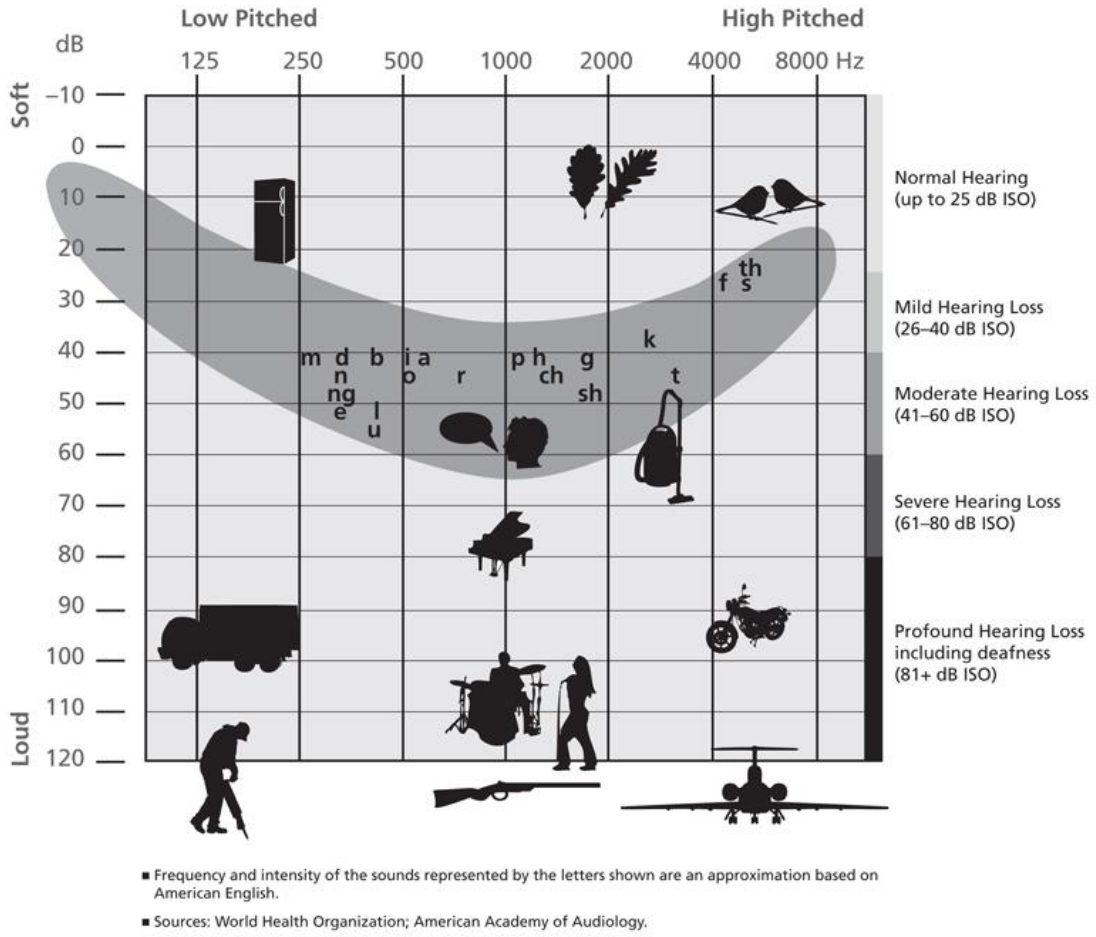


Fig.1 comune rappresentazione banana speech, autore e lingua non specificata

Il banana speech è un grafico a forma di banana che ipotizza la distribuzione delle caratteristiche del parlato, dove l'ascissa e l'ordinata rappresentano rispettivamente la frequenza (Hz) e l'intensità (dB). Questa rappresentazione mostra i suoni del parlato (consonanti e vocali) pronunciati con un'intensità normale in termini di frequenza e livello di intensità. Il parlato, effettivamente, è un flusso di suoni, con un suono che segue l'altro in modo rapido, ogni suono del parlato sembra occupare determinate posizioni sulla banana del parlato (ad esempio, le vocali tendono ad avere una frequenza più bassa e un livello di intensità più alto rispetto alle consonanti). Nonostante il banana speech sia stato ampiamente accettato e citato nei campi dell'audiologia e delle scienze dell'udito, le tecniche e le fasi impiegate nella costruzione di ciascuna banana vocale non sono ben documentate.

Aspetto fondamentale nel BS sta nel fatto che un suono è indicato da un punto, situazione non del tutto corretta, come si andrà a dimostrare con qualche esempio riportato di seguito

Nella letteratura esistono molte versioni di *banana speech* per l'inglese, una delle quali è stata proposta da Northern e Downs [fig.3]. Appreso che le lingue di tutto il mondo hanno sistemi fonemici unici, non dovrebbe sorprendere che la banana di ciascuna lingua differisca in qualche misura l'una dall'altra. È ragionevole aspettarsi che le aree figurative (nel complesso) siano simili, ma che le posizioni specifiche dei suoni vocalici possano variare. [1]

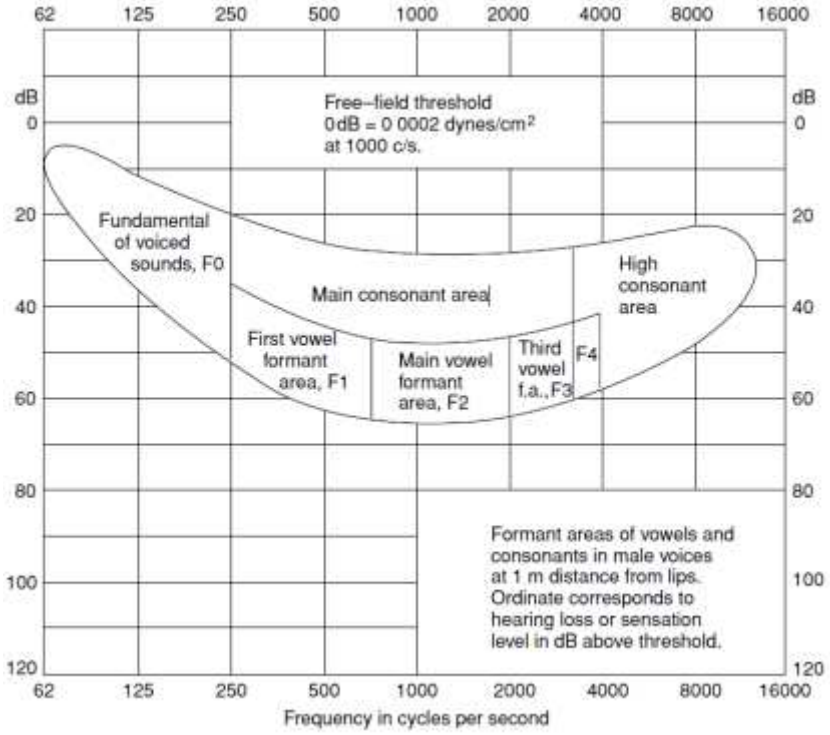


Fig.2 banana speech schematizzato per la lingua svedese [1]

L'audiogramma riportato qui sopra [Fig.2] entra più nello specifico, sempre considerando la lingua svedese come lingua analizzata, seziona in banana speech in differenti aree. Nella prima "fundamental of voice sounds" tra 60 e 250 Hz indica appunto la frequenza fondamentale F0 della voce, tutte le seguenti armoniche saranno multiple a F0. [1]

Nell'area centrale troviamo due aree, nella parte superiore le diverse formanti delle vocali e nella parte inferiore la zona acustica in cui si sviluppano le armoniche della maggior parte delle consonanti. L'ultima area viene indicata come dominio delle consonanti più acute, appunto ricche di alte frequenze, tra 4000 Hz e 15-16000 Hz, in questa area troveremo suoni come /f/ /s/.[1]

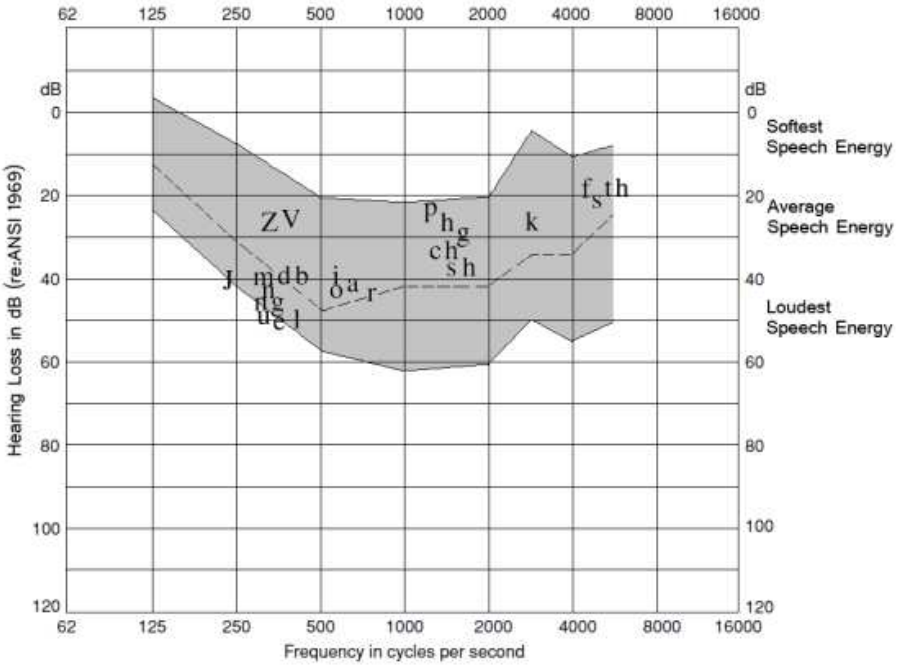


Fig.3 Banana speech Northern & Downs [2].

Nella fig.3 vediamo invece la rappresentazione proposta da Northern e Downs pubblicata in “Hearing in Children” [2]. Anche qui notiamo come il grafico sia ridotto in frequenza, dando un'idea semplicistica della rappresentazione dei suoni per parlato.

Nella fig.1 introduttiva vediamo che all’interno della banana vengono riportate, ad una certa coordinata, delle lettere. Ipotizziamo che servano ad indicare la caratteristica in frequenza di maggiore ampiezza del fonema (di norma quello maggiormente usato si basa sui fonemi di lingua inglese).

Ipotizzando lo scopo di questo grafico (Fig. 4), la sovrapposizione del banana speech su di un audiogramma possiamo teoricamente prevedere quali fonemi saranno percepiti con più facilità e quelli in cui ci sarà bisogno di un'eventuale amplificazione per essere sentiti. Da quanto descritto precedentemente capiamo che il risultato di questo confronto non sarà fedele alla realtà.

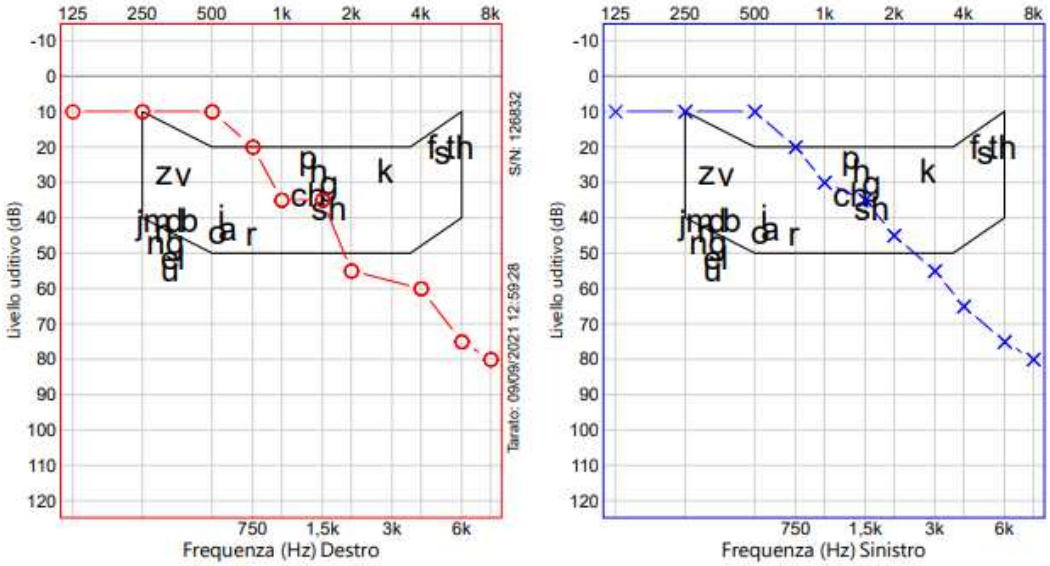
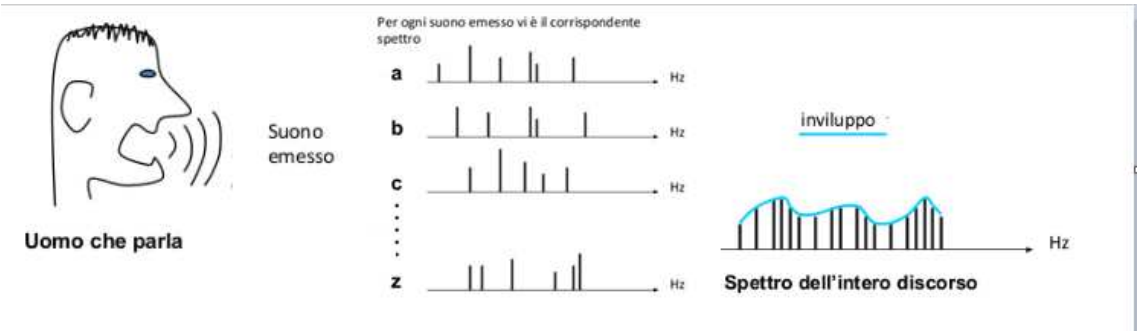


Fig.4 estrazione da Otosuite (Otosuite | Natus), ipoacusia simulata

Il punto dell'osservazione si basa sulla tipologia di rappresentazione scelta, infatti, riportando i fonemi come meri punti, l'interpretazione può essere fuorviante e dare l'idea che il suono non abbia ulteriori componenti ma solo quella rappresentata sull'audiogramma, probabilmente la componente fra tante che ha maggiore ampiezza. Ricordo che se provassimo a emulare un suono avente un'unica componente armonica otterremo un tono puro, nient'altro.

Spettro statistico cumulativo della voce umana

La premessa del nostro studio come già detto è L'estensione in frequenza della voce umana, questa è dimostrabile attraverso lo spettro statistico cumulativo della voce umana, è questo l'ingrediente che ci porta ad approfondire lo studio del banana speech.



L'analisi viene fatta su una registrazione di una conversazione di 4 minuti fra tre persone, viene fatta un'analisi in frequenza sulla sua intera durata dando come risultato una rappresentazione di tutte le componenti formanti dell'intero segnale. Le registrazioni sono state effettuate con apparecchiature di elevata qualità in condizioni prive di rumore di fondo. Sono stati analizzati i segnali relativi alla emissione acustica di più persone con lo scopo di ottenere uno spettro statistico cumulativo che si potesse considerare come una sorta di informazione abbastanza completa, relativamente alla voce umana, sia maschile che femminile. Le figure(fig5;fig6,fig7) mostrate di seguito mostrano la somma o unione di tutte queste formanti sullo spettrogramma [4].

Fig.5 spettro cumulativo relativo ad un discorso di circa 4 minuti tenuto prima da un soggetto femminile e poi maschile (soggetti adulti) con l'analisi estesa sulla prima decade tra 20 Hz e 200 Hz [4].

Fig.6 analisi estesa sulla seconda decade tra 200 Hz e 2000 Hz [4].

Fig.7 analisi estesa sulla terza decade tra 2000 Hz e 20000 [4].

Le componenti considerate dall'analisi non sono limitate in brevi intervalli di tempo ma sono quelle che si possono trovare durante l'intera evoluzione temporale del tempo.

Da un punto di vista insiemistico si tratta dell'unione di tutte le componenti sinusoidali nelle quali, in base al teorema di Fourier, può essere scomposto il segnale durante tutta la sua evoluzione. In certi intervalli di tempo si troveranno alcune componenti, in altri intervalli di tempo se ne troveranno delle altre. Al termine dell'evoluzione completa del segnale, un po' alla volta saranno state individuate tutte le possibili componenti armoniche del segnale, un po' come una specie di accumulo di tutte le componenti che costituiscono l'intero segnale. Analizzando i grafici che rappresentano un possibile spettro statistico cumulativo della voce umana, risulta evidente come vi sia una estensione frequenziale elevata, di gran lunga superiore agli 8 kHz. [5] [6]

Ora che siamo a conoscenza di quanto sia ampio in realtà lo spettro del parlato entriamo nello specifico e mostriamo come i fonemi presi singolarmente si sviluppano nel dominio della frequenza. Lo studio proposto vede in analisi alcuni dei fonemi del banana speech e ne consiglia una rappresentazione di diverso tipo che sia in grado di raffigurare le componenti frequenziali indispensabili per una riproduzione discreta del fonema. Nello studio proposto analizzeremo solo le caratteristiche in frequenza dei fonemi analizzati, non di intensità. Non viene proposto uno studio statistico per la produzione di valori medi, ma viene solo evidenziata una caratteristica fondamentale del suono, la frequenza e quanto ne sia caratterizzato ogni fonema.

Non terremo conto di fattori psicoacustica come le curve isofoniche, sulle analisi fatte non sono stati applicate conversioni SPL -> HL o altri fattori non descritti in questo studio.

Materiali e metodi

I fonemi analizzati sono stati estrapolati da diverse registrazioni di una voce maschile e una voce femminile, persone di età adulta, madrelingua e con buona dizione, entrambi insegnanti scuole medie superiori / università.

Le tracce registrate contengono logotomi e parole pronunciate ad un'intensità ipotetica di colloquio in situazione di quiete.

Le apparecchiature di registrazione usate sono:

- Microfono Rode NTA2A
- Preamplificatore Custom con banda frequenziale 2Hz-200 kHz
- Rumore equivalente in input < 0,5 micro Volt
- Registratore Tascam DV RA1000 HD, impostato con Campionamento 192k Hz,
- Quantizzazione 24 bit

L'analisi frequenziale è stata effettuata utilizzando:

- Cool edit pro versione 2.0, impostando la FFT (Fast Fourier Trasform) con finestra gaussiana

L'ascolto è stato realizzato con:

- Cuffie AKG K77 (per la riproduzione)

Campioni analizzati

- Frammenti di 80-85 ms, 16k campioni analizzati
- Frammenti di 40-45 ms, 8k campioni analizzati

L’analisi dei fonemi è suddivisa in diverse fasi. La prima fase è stata la registrazione delle voci, in ambiente silente, con un ridottissimo tasso di riverberazione. La seconda fase consiste nella scelta del frammento di registrazione che riproduce il fonema desiderato e la sua durata, questo può essere trovato solo attraverso l’oscillogramma.

Prima di tutto è necessario trovare una certa ripetitività dell’evoluzione temporale nell’andamento temporale della grandezza, una periodicità. Il periodo dà indicazione dell’intervallo di tempo nel quale evoluzione del segnale è completa [5].

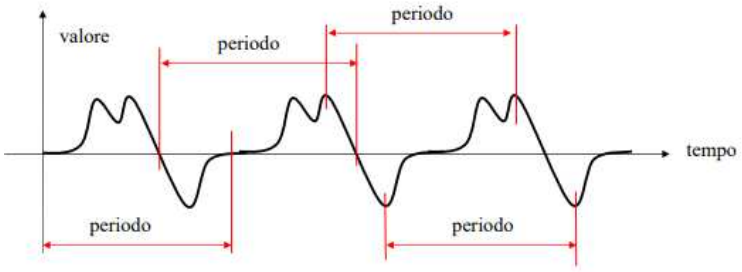


Fig.8 individuazione del periodo del segnale nell’evoluzione temporale [4].

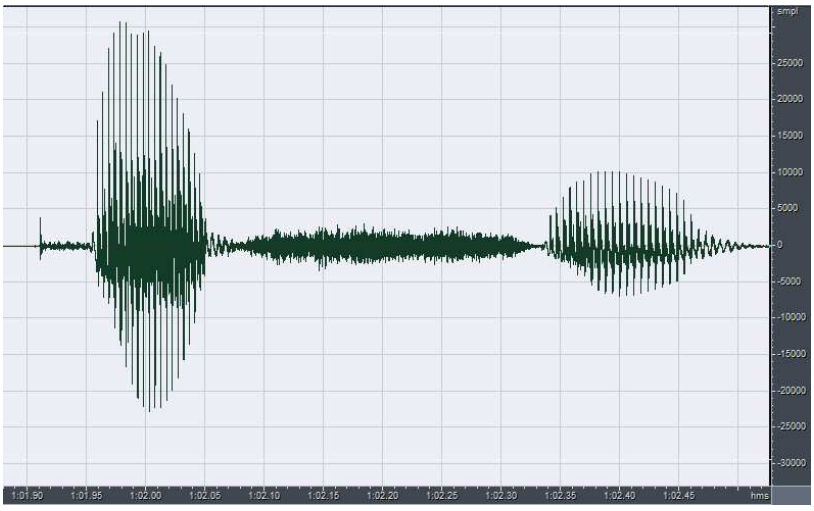


Fig.9 Oscillogramma, voce donna, parola “cassa”. Immagine estratta dal lavoro effettuato per questo studio.

Nella fig.6 possiamo apprezzare l’evoluzione temporale rappresentato nell’oscillogramma fornito dal programma in uso della parola “cassa”. L’oscillogramma ci dà una rappresentazione bidimensionale dove in ordinata c’è l’intensità definita in intervalli e in ascissa in tempo. Da questa parola abbiamo la possibilità di estrapolare quattro fonemi differenti, nella fig.9 vediamo l’identificazione del fonema “a” in un intervallo temporale di 85 ms.

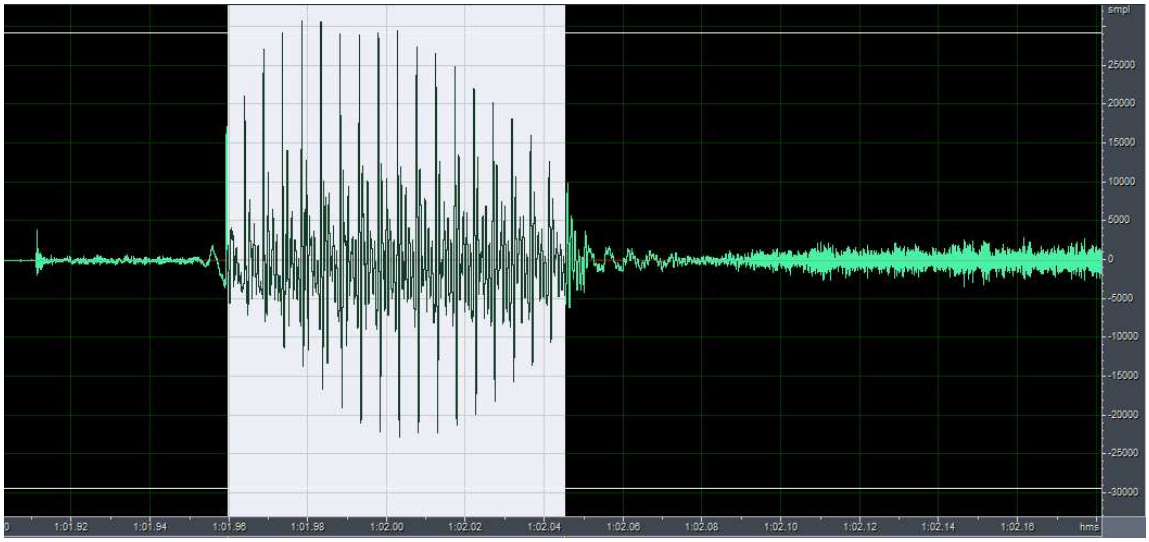


Fig.10 selezione intervallo 85 ms in cui possiamo trovare una periodicità costante per l'intera durata.

Per tutte le vocali e quasi la totalità delle consonanti analizzate è stato analizzato un frammento di 80-85 ms in cui sono stati considerati 16000 campioni al suo interno. Per i fonemi la cui durata all'interno della registrazione era inferiore a 85 ms è stato scelto di analizzare un frammento di 43 ms. Formula imposta per determinare la durata dell'intervallo in corrispondenza al numero di campioni analizzati dalla Fast Fourier Transformat.

1 secondo : 192000 Hz = T(durata intervallo) : numero campioni FFT

1 : 192000 = x : 16384 -> x = 85 ms

1 : 192000 = x : 8192 -> x = 43 ms

Nella fig.10a possiamo apprezzare meglio lo sviluppo temporale del fonema /a/

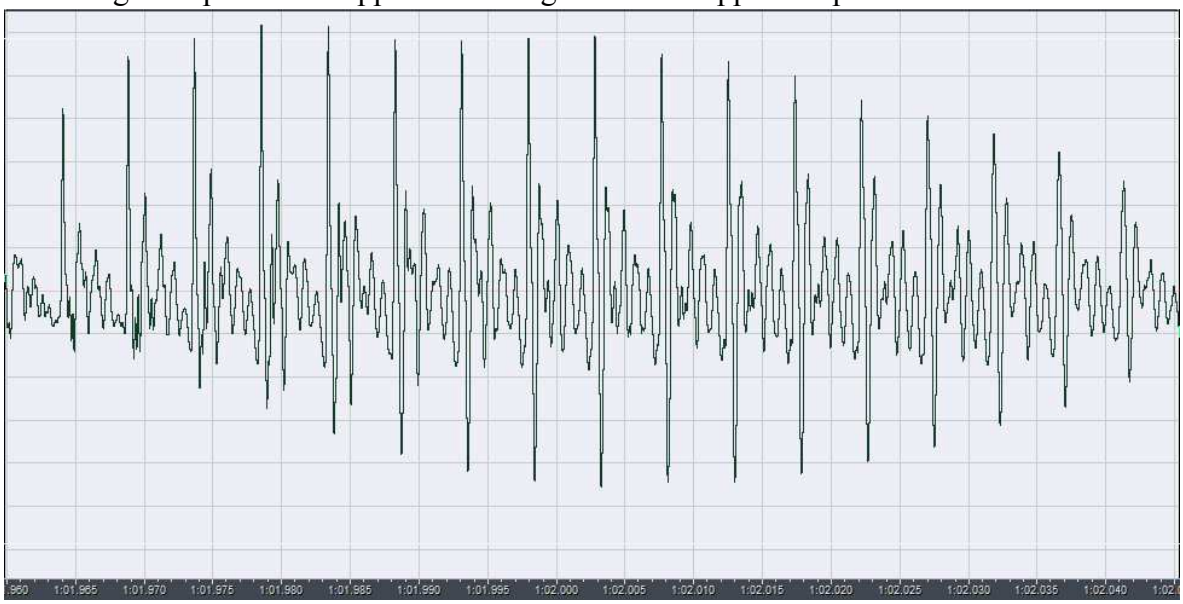


Fig.10a oscillazione periodica ripetuta durante l'intervallo scelto.

Terza fase è l'analisi del frammento e quindi l'utilizzo della FFT secondo le regole che ci siamo imposti secondo la seguente formula:

1 secondo : 192000 Hz = T(durata intervallo) : numero campioni FFT

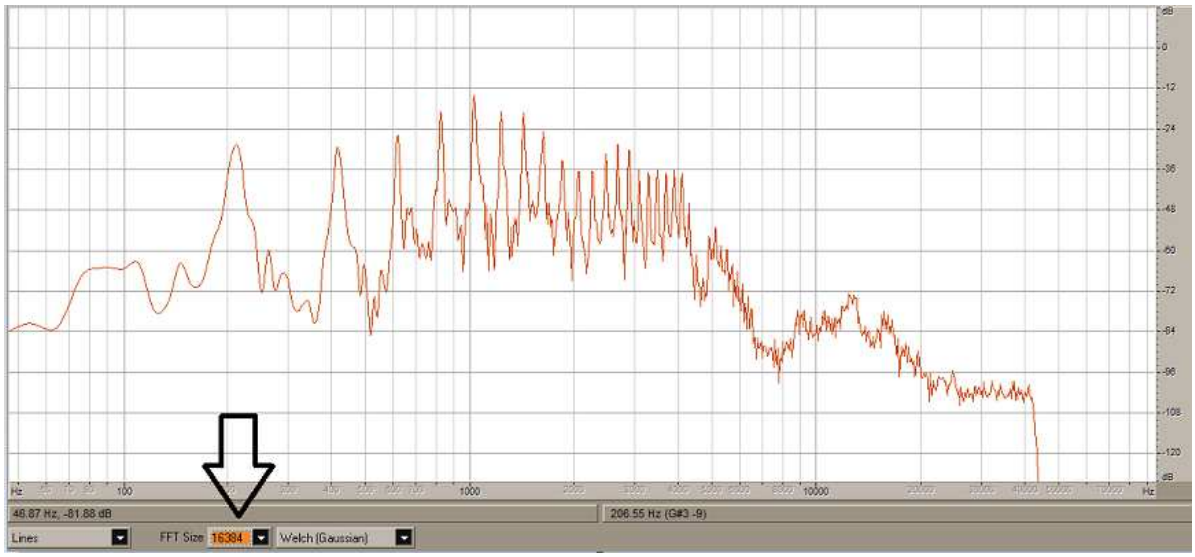


Fig.11 Inseriamo il numero di campioni da considerare nell'intervallo, vediamo inoltre l'analisi in frequenza ottenuta. Nella fig.11 possiamo apprezzare il risultato dell'estrazione armonica data dalla trasformata di Fourier. Notiamo anche la banda frequenziale utilizzata per la registrazione 20 Hz-20kHz

Discussione

In seguito all'analisi di diversi fonemi presenti nelle varie registrazioni eseguite con l'apparecchiatura prima descritta, abbiamo visto come questi si estendano nel dominio della frequenza e non si limitino semplicemente ad un punto unico e quindi ad una sola frequenza, ma invece posseggono caratteristiche armoniche che si evolvono su una banda più o meno grande (anche più ottave) nel dominio della frequenza.

Di seguito vedremo l'analisi effettuata su alcuni dei fonemi analizzati, così da evidenziare le caratteristiche frequenziali predominanti, considerando un intervallo in intensità di 24 dB, definito come intervallo minimo per riprodurre il suono con qualità discreta, comunque ancora lontana dalla vera estensione del fonema analizzato.

Fig.12 fonema /a/ uomo

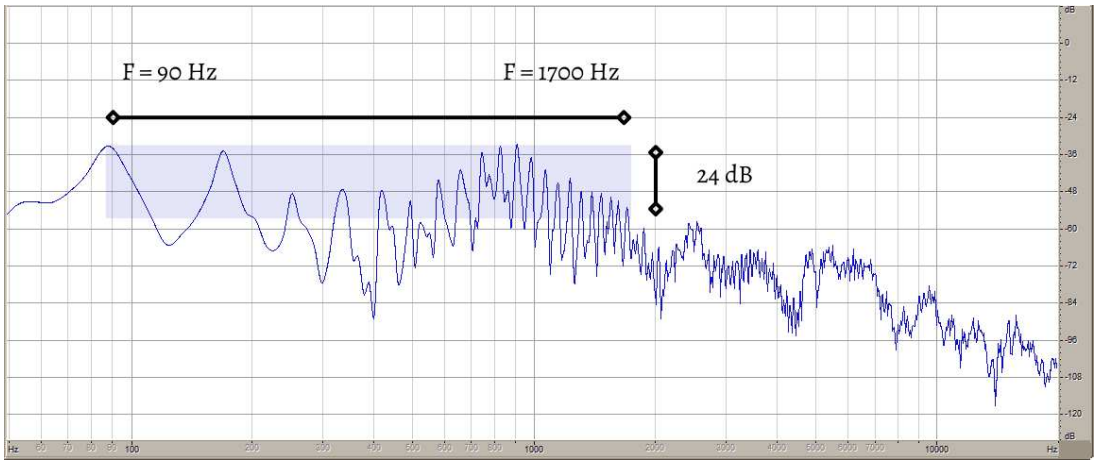
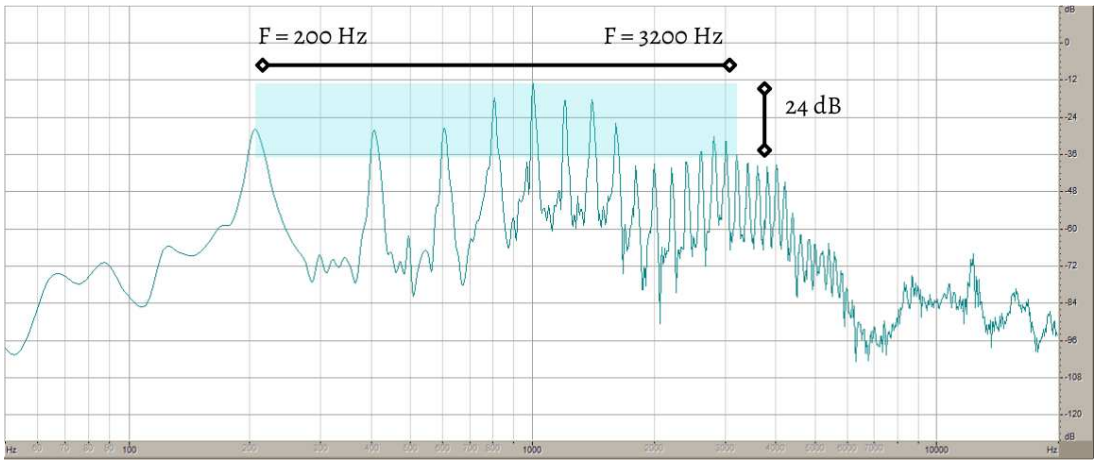


Fig.13 fonema /a/ donna.



I due grafici (fig.12, fig.13) raffigurano l'analisi in frequenza del fonema /a/ riprodotto da una voce maschile e una voce femminile registrata dalle apparecchiature ad alta fedeltà in nostra dotazione. Notiamo il primo picco corrispondente alla frequenza fondamentale caratteristica della voce maschile a 100 Hz e quella femminile a circa 200 Hz e di seguito le loro armoniche (ovvero i picchi successivi al primo e multipli di esso). Considerando l'intervallo di 24db, il fonema /a/ si estende per più di 4 ottave, dalla frequenza di 100 Hz alla 1700 Hz per la voce maschile e 200 Hz alla frequenza di 3200k Hz per la voce femminile. Chiaramente questi dati non devono essere indicazione assoluta di voce maschile o femminile ma solo una dimostrazione dell'estensione in frequenza avuto in questo caso. Se l'intervallo in intensità fosse maggiore (es 36, 48 dB) vedremo come il fonema sia ricco di ulteriori armoniche e quindi di quanto si estenda ulteriormente in frequenza, fino ad occupare più di 6 ottave.

Misure
Di seguito vengono riportate alcune delle analisi effettuate.

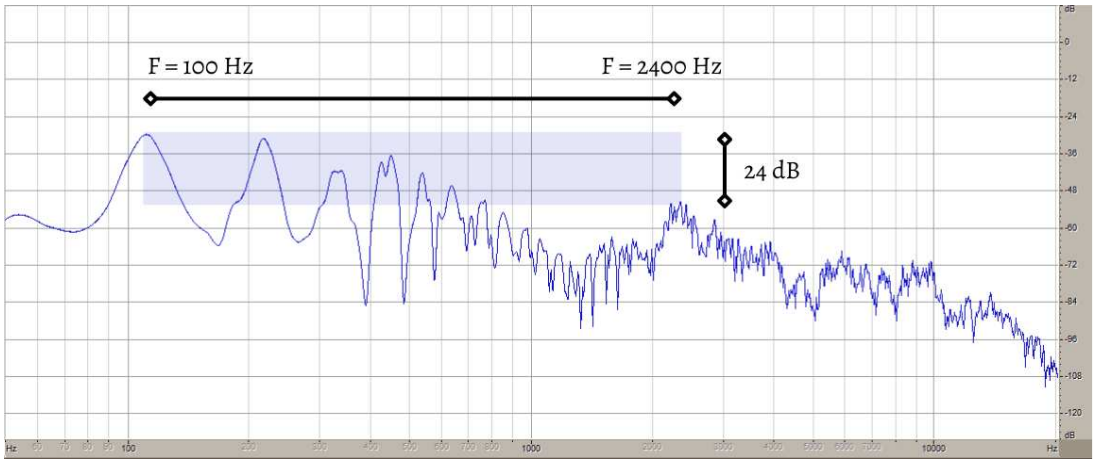


Fig.14 fonema /e/ uomo

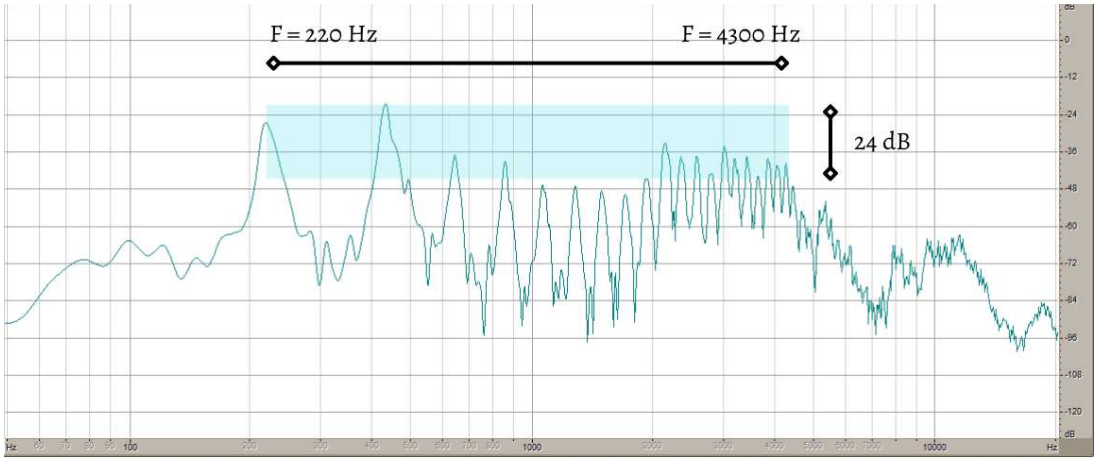


Fig.15 fonema /e/ donna

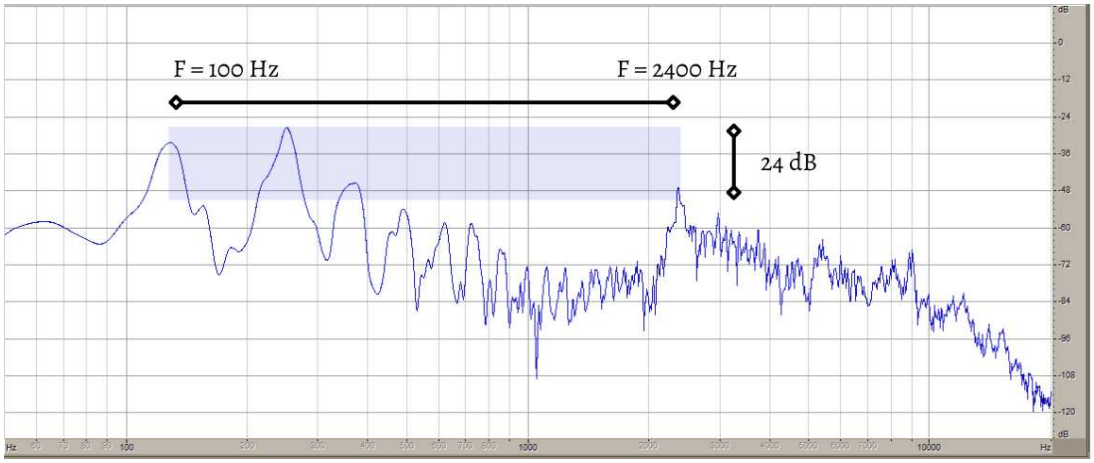


Fig. 16 fonema /i/ uomo

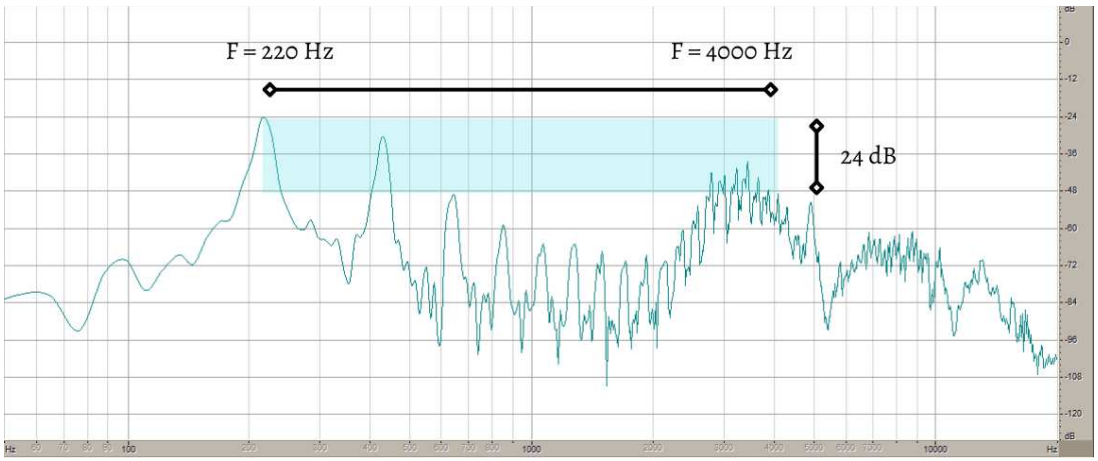


Fig. 17 fonema /i/ donna

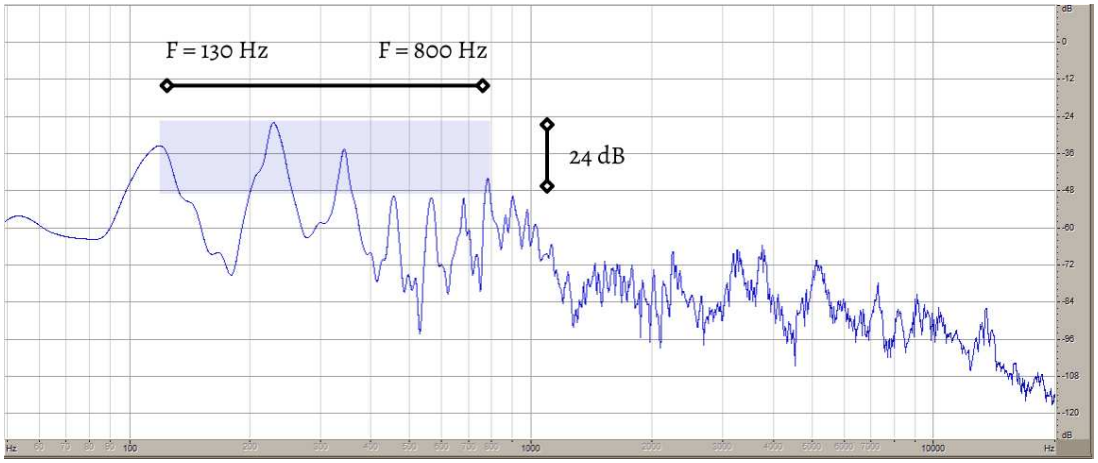


Fig. 18 fonema /u/ uomo

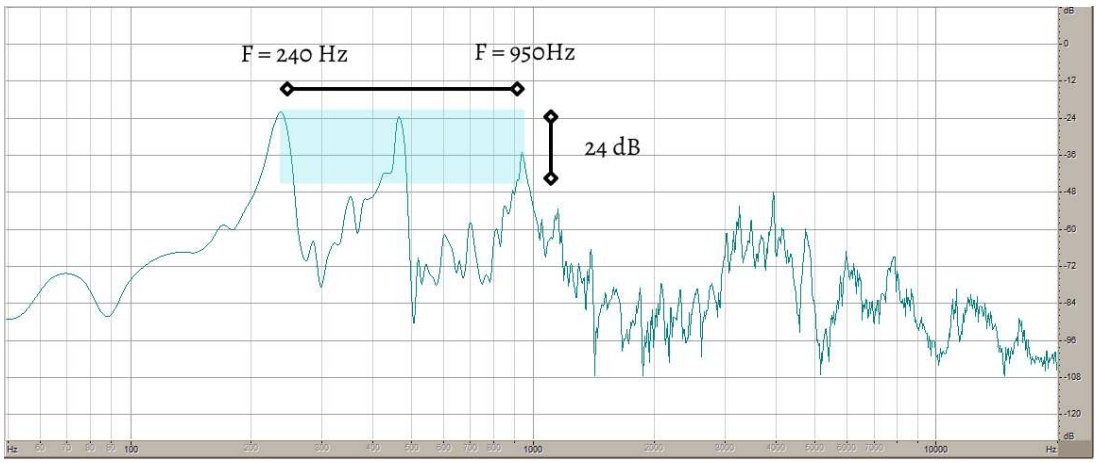


Fig. 19 fonema /u/ donna

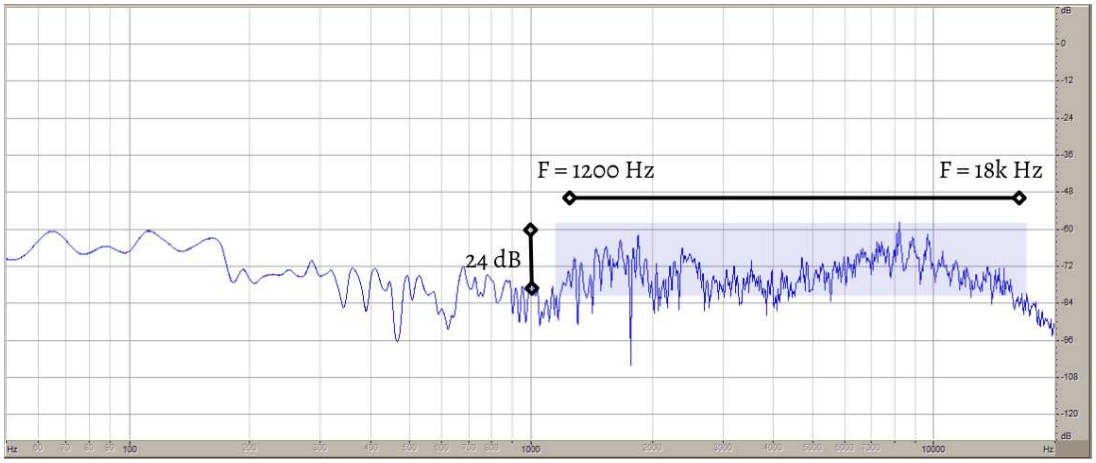


Fig.20 Fonema /f/ uomo

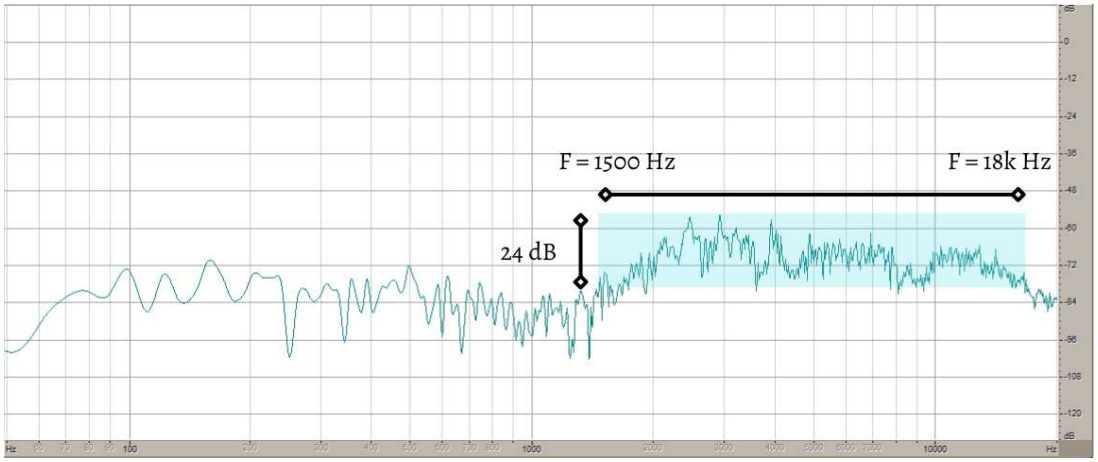


Fig.21 Fonema /f/ donna

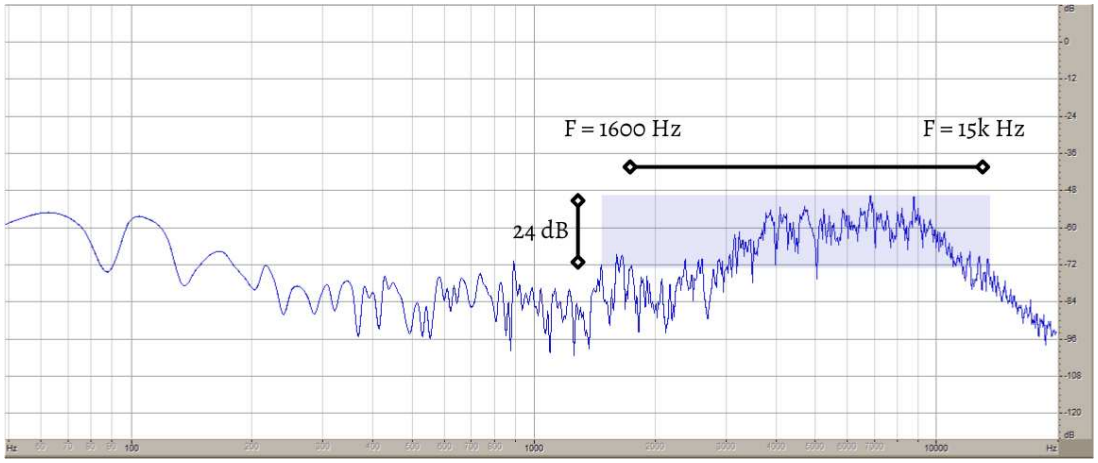


Fig.22 Fonema /s/ uomo

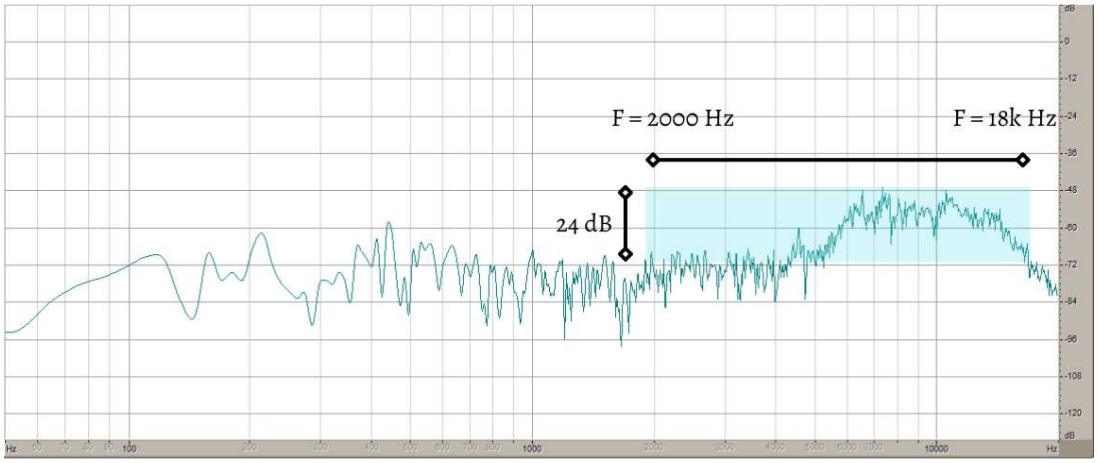


Fig.23 Fonema /s/ donna

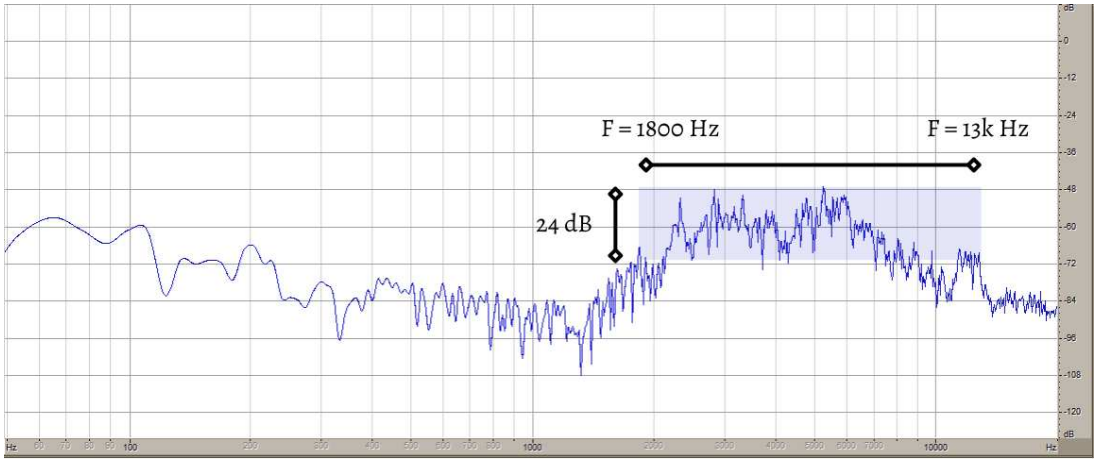


Fig.24 Fonema /tʃ/ uomo

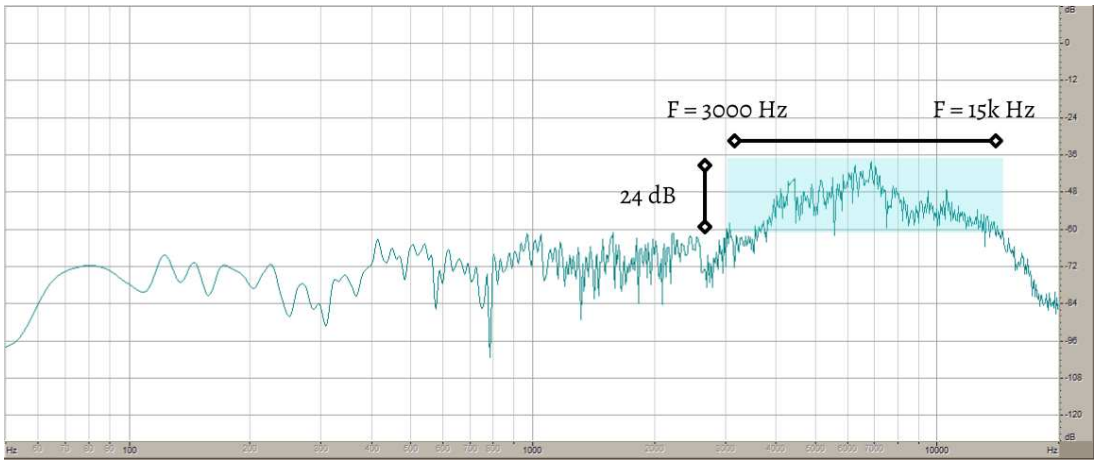


Fig.25 Fonema /tʃ/ donna

Conclusioni

Confrontando i risultati ottenuti con il presente lavoro, con la rappresentazione grafica fornita dai più comuni ed utilizzati BS, si può affermare che la descrizione fornita da questi ultimi è ben lontana da una descrizione frequenziale realistica e necessaria a comprendere le caratteristiche del parlato. Interpretare un suono o un fonema come una singola componente frequenziale è sicuramente riduttivo, oltre che essere concettualmente errato

Una caratteristica necessaria ad una lettura abbastanza corretta nel dominio frequenziale dei vari fonemi potrebbe essere una rappresentazione grafica dove, al posto del singolo punto è indicato un segmento che copre l'estensione frequenziale delle componenti formanti effettivamente significative, per una corretta interpretazione del fonema. Un punto può essere indicato in corrispondenza della frequenza della componente di massima ampiezza

Fig.20 esempio rappresentazione con segmento, fonema /a/ donna

Come già detto in precedenza, da semplici test effettuati, si è visto come, per una corretta ed agevole interpretazione di un fonema, occorra prendere in considerazione tutte le componenti formanti con un'ampiezza fino a -24dB rispetto quella di ampiezza maggiore. Se si pretende una riproduzione di fonema quasi perfetta il range come ampiezza delle componenti formanti da prendere in considerazione dovrebbe essere di almeno 30-36 dB, ciò comporterebbe un allargamento frequenziale dello spettro con una conseguente espansione grafica, in orizzontale, del BS conseguente.

Bibliografia

[1] F. Gunnar, "Speech Perception," in Speech Acoustics and Phonetics, vol. 24, Dordrecht, the Netherlands, Kluwer Academic Publishers, 2004.

[2] J. L. Northern and M. P. Downs, Hearing in Children, Pennsylvania: Lippincott Williams & Wilkins, 1984.

[3] D. Ling, Foundations of Spoken Language for Hearing-impaired Children, Washington, DC: Alexander Graham Bell Association for the Deaf, 1989.

[4] Ing.A.F.Selmo, Dispense "risposta in frequenza", Misure: ing. Selmo, Strumentazione usata: Analizzatore di spettro hp 3585 A spettro cumulativo statistico. Abilitazione della costruzione dello spettro statistico cumulativo, con la memorizzazione della massima ampiezza di ciascuna delle componenti formanti individuate durante l'analisi del segnale.