

UNIVERSITÀ DEGLI STUDI DI MILANO  
FACOLTÀ DI SCIENZE E TECNOLOGIE



CORSO DI LAUREA TRIENNALE IN INFORMATICA MUSICALE

ANALISI DEL COMPORTAMENTO DELLA DOPPIA  
INCISIONE DI CHITARRE E VOCI IN UNA  
PRODUZIONE MUSICALE

Relatore: Prof. Goffredo Haus  
Correlatore Interno: Dott. Giorgio Presti  
Correlatore Esterno: Mr. Disi Melotti

Tesi di Laurea di:  
Albanese Mirko  
Matr. Nr. 828180

ANNO ACCADEMICO 2017-2018



# Prefazione

hkjafgyruet.

## Organizzazione della tesi

La tesi è organizzata come segue:

- nel Capitolo 1 ...
- nel Capitolo 2 ...
- nel Capitolo 3 ...
- nel Capitolo 4 ...

# Ringraziamenti

asdjhgtry.

# Indice

Prefazione

Ringraziamenti

<b>1</b>	<b>Introduzione</b>	<b>1</b>
<b>2</b>	<b>Fondamenti Teorici</b>	<b>2</b>
2.1	Natura del suono . . . . .	2
2.1.1	Il suono . . . . .	2
	Inviluppo . . . . .	3
	Altezza . . . . .	4
	Intensità . . . . .	5
	Timbro . . . . .	6
2.2	Analisi segnali nel dominio della frequenza . . . . .	7
2.2.1	Segnali Periodici . . . . .	7
2.2.2	Serie di Fourier . . . . .	8
2.2.3	Trasformata di Fourier . . . . .	8
	DFT . . . . .	10
	STFT . . . . .	11
<b>3</b>	<b>Modello di Analisi</b>	<b>15</b>
3.1	Analisi della dinamica . . . . .	15
3.2	Analisi armonica - Harmonic tracking . . . . .	17
3.2.1	Interpolazione parabolica . . . . .	17
3.2.2	Differenza di fase . . . . .	20
3.2.3	Validazione dei modelli . . . . .	22
	Chirp . . . . .	22
	Riff di chitarra . . . . .	24
3.3	Analisi timbrica - Rilevamento delle formanti . . . . .	24
3.4	Analisi attacco delle note - Onset Detection . . . . .	25

## *INDICE*

<b>4</b>	<b>Test effettuati e relativi risultati</b>	<b>26</b>
4.1	Strumento : Chitarre . . . . .	26
4.2	Strumento : Voci . . . . .	26
<b>5</b>	<b>Conclusioni e sviluppi futuri</b>	<b>27</b>

# Capitolo 1

## Introduzione

...

L'obiettivo principale è quello di estrarre le differenze di un segnale stereofonico prodotto mediante la tecnica della doppia incisione. Quando si produce un disco, una tecnica usata da maggior parte dei produttori è quella di effettuare due incisioni della medesima parte musicale e posizionarle all'interno dell'immagine stereofonica (una incisione a sinistra e una a destra), questa tecnica è molto usata nelle incisioni di chitarra elettrica e chitarra acustica garantendo una notevole apertura e profondità stereofonica, ovviamente questa tecnica non vieta al produttore di effettuare quattro, sei o addirittura otto incisioni delle medesime parti suonate e ripartirle a proprio piacimento. In questo elaborato ci concentreremo a estrarre le caratteristiche identificative dei due segnali e calcolarne le differenze.

# Capitolo 2

## Fondamenti Teorici

### 2.1 Natura del suono

La percezione sonora è normalmente legata alle vibrazioni del timpano nell'orecchio. Queste vibrazioni sono provocate da piccole variazioni di pressione nell'aria. La variazione di pressione dell'aria è quindi l'equivalente fisico del suono.

#### 2.1.1 Il suono

Il suono, come abbiamo accennato, corrisponde a variazioni di pressione nell'aria. Molti suoni musicali presentano variazioni regolari di pressione, in particolare la regolarità implica che un determinato andamento della pressione si ripeta nel tempo, definendo così il termine *forma d'onda* la ripetizione di tale andamento. In questo caso il suono è detto *periodico* e la durata della singola forma d'onda è detta *periodo*. Nel caso opposto, in cui l'andamento della pressione è privo di qualsiasi regolarità, il segnale associato viene percepito come *rumore*. Il rumore può essere diviso in due classi principali:

- *rumore impulsivo* : determinato da rapide variazioni di pressione circoscritte nell'arco di pochi millisecondi. Un tipico esempio di rumore impulsivo si ha quando un corpo rigido viene percosso. Va notato che il rumore impulsivo viene regolarmente generato durante la produzione di suoni musicali,
- *rumore stazionario* : ha generalmente un'elevata estensione temporale ma è comunque privo di regolarità. Tipici esempi di rumore stazionario sono il rumore prodotto dal vento o quello proveniente da uno schermo televisivo in assenza di segnale (*effetto neve*).



In Figura 1 vengono riportati gli andamenti delle forme d'onda rispettivamente associate a una sinusoide, ad un segnale periodico costituito da una somma di 16 sinusoidi in rapporto armonico tra loro e ad un segnale rumoroso stazionario.

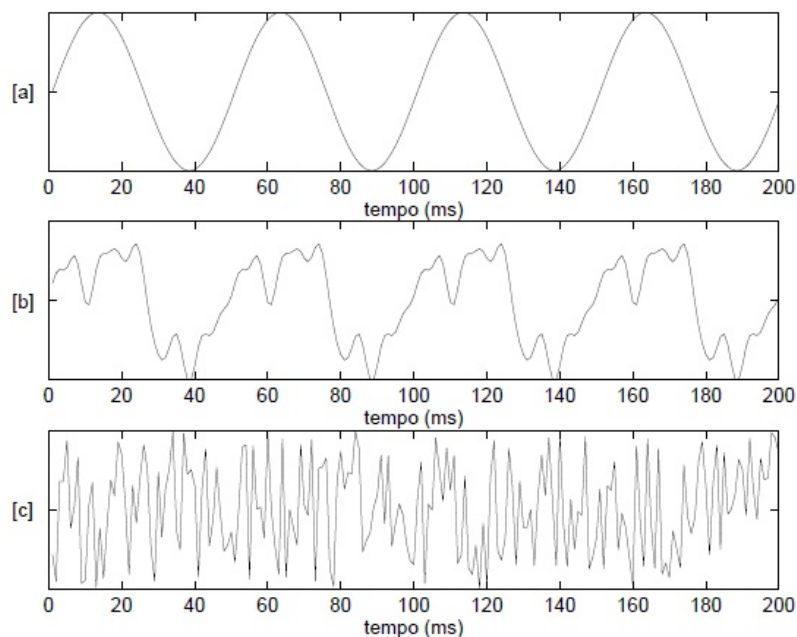


Figura 1: Andamento nel tempo di tre segnali rispettivamente: [a] sinusoidale, [b] periodico (somma di 16 sinusoidi), [c] aperiodico.

## Inviluppo

L'andamento dell'ampiezza di un suono dal momento in cui viene generato fino a quando si estingue viene chiamato inviluppo, in particolare ogni suono ha un inizio e una fine.

Prendiamo come esempio la generazione di un suono da una corda di violino eccitata con l'archetto. In condizioni di riposo la corda ha ovviamente vibrazione nulla, e quindi non produce alcun suono. Quando il violinista inizia a sfregare l'archetto sulla corda, questa inizia a vibrare abbandonando la situazione di riposo. Esiste un periodo di tempo nel quale le oscillazioni della corda, da nulle, si fanno sempre più ampie. Questa viene definita fase di attacco e solitamente indicata con il corrispondente termine inglese *attack*. Questa fase dura solitamente pochi centesimi di secondo, in relazione al tipo di strumento musicale. La fase successiva è definita con

il termine inglese *decay*: corrisponde ad un rapido assestarsi dell'ampiezza ad un valore stabile dopo una sovraelongazione a cui è stata portata dalla fase di attacco. A questo punto, esaurito il transitorio di attacco, si è realizzato un accoppiamento tra lo sfregamento dell'archetto e le oscillazioni della corda. Questo corrisponde alla fase di *sustain*, che può durare anche parecchi secondi, nella quale il suono viene appunto sostenuto dal musicista, che continua a fornire l'energia necessaria per mantenere le vibrazioni. L'ultima fase, che ha inizio nel momento in cui il musicista smette di mantenere eccitato il sistema di vibrazione, viene denominata *release* (ovvero rilascio) e corrisponde al tempo in cui il corpo vibrante (nel nostro esempio la corda di violino) smorza l'entità delle vibrazioni, fino a portarsi nuovamente nello stato di quiete.

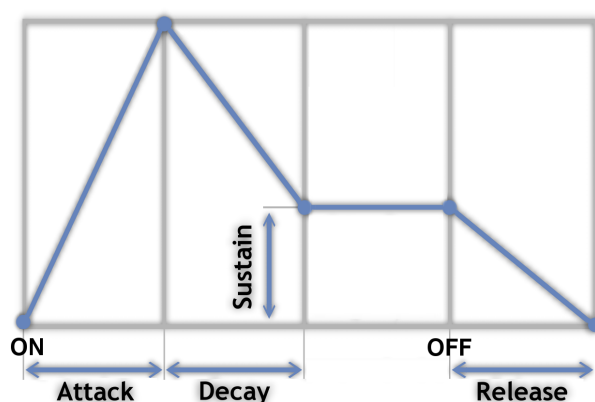


Figura 2: Inviluppo - ADSR

### Altezza

Il primo parametro percettivo che andremo a descrivere è l'altezza di un suono, ovvero la frequenza. La frequenza è una grandezza che indica il numero di oscillazioni compiute in un secondo e riguarda fenomeni periodici o fenomeni ripetitivi.

In campo musicale si è soliti descrivere un suono periodico in termini di frequenza, usualmente indicata con il simbolo  $f$  e misurata in Hertz(Hz). Il legame tra periodo  $T$  e frequenza  $f$  è descritto dalla formula:

$$f = \frac{1}{T}$$

Come vedremo di seguito un suono periodico di frequenza  $f$  può essere scomposto in forme d'onda elementari con frequenze rispettivamente  $f, 2f, 3f, 4f, \dots$ . La sinusoide di frequenza  $f$ , pari alla frequenza del suono periodico di partenza è detta

*fondamentale*, mentre le sinusoidi di frequenza multipla intera di  $f$  vengono dette parziali o *armoniche*.

Le caratteristiche frequenziali inducono una differenziazione dei suoni in suoni puri e complessi.

Un suono puro (detto anche tono) è costituito da una sola frequenza ed è quindi descritto da un'onda sinusoidale semplice;

Un suono complesso consiste invece di più frequenze sommate in un'onda dall'andamento articolato; in un singolo periodo possono essere comprese più alternanze di compressioni e rarefazioni intermedie. In generale in natura i suoni sono di tipo complesso, e lo specifico andamento deriva dal metodo di produzione del suono da parte della sorgente.

### Intensità

Si è detto che l'equivalente fisico del suono è la variazione di pressione nell'aria. L'entità delle variazioni di pressione è legata alla percezione di volume sonoro (*loudness*): maggiore è la variazione di pressione, maggiore sarà il volume sonoro percepito.

I valori di pressione, potenza e intensità acustica dei suoni si distribuiscono in un intervallo di valori molto esteso. Per questa ragione queste grandezze sono comunemente espresse in scala logaritmica. Va inoltre osservato che la scala logaritmica ha un andamento più vicino a quello delle scale percettive del nostro apparato uditivo. L'orecchio umano è in grado di percepire intensità acustiche che variano in un intervallo molto grande (12 ordini di grandezza): si definisce soglia di udibilità il valore  $I = 10^{-12} \text{W/m}^2$  al di sotto del quale non è più possibile percepire alcun rumore, mentre si chiama soglia del dolore il valore  $I = 1 \text{W/m}^2$  al di sopra del quale si inizia a provare dolore fisico.

Viene definito come livello di pressione acustica ( $PL$ ), misurata in decibel (dB) il logaritmo del rapporto tra la pressione misurata e una pressione di riferimento. In formule:

$$PL = 20 \cdot \log_{10} \frac{p}{p_{ref}}$$

dove  $p$  è un valore di pressione e  $p_{ref}$  è il valore di pressione di riferimento.

Può risultare comunque conveniente utilizzare come riferimento la minima pressione efficace udibile  $p_0 = 0,00002 \text{Pa}$ ; in questo caso si parla di Sound Pressure Level (SPL).

Indicazione	Sorgente sonora	Intensità (dB)
	Silenzio	0
	Spillo che cade	10
	Sussurro a 1 m	20
	Sala vuota	30
<i>ppp</i>	Libreria	40
<i>pp</i>	Interno auto silenziosa	50
<i>p</i>	Conversazione pacata	60
<i>mp</i>	Traffico	70
<i>mf</i>	Fabbrica	80
<i>f</i>	Metropolitana	90
<i>ff</i>	Discoteca	100
<i>fff</i>	Concerto rock	110
	Jet in partenza a 500 m	120

Tabella 1: Livello di intensità associato alle indicazioni di partitura (prima colonna) e prodotto da alcune sorgenti sonore (seconda colonna). (I valori riportati vanno presi come puramente indicativi)

## Timbro

Il timbro è quella particolare qualità del suono che permette di distinguere due suoni, rappresenta dunque, quell'attributo della sensazione uditiva che consente all'ascoltatore di identificare la fonte sonora, rendendola distinguibile da ogni altra.

Una nota suonata da una pianoforte avrà un timbro differente rispetto alla stessa nota prodotta da un violino o da un flauto. Il timbro è determinato dalle caratteristiche fisiche dello strumento, quali il mezzo utilizzato per produrre il suono (corde, pelle, ancia, ...), dipendente prevalentemente dalle *zone formanti*, o semplicemente dette *formanti*.

Per formante s'intende una frequenza di risonanza attorno alla quale un suono spettralmente ricco ha un picco di ampiezza, sono zone in frequenza dove vi è una notevole concentrazione di energia. Non devono essere confuse con le costituenti armoniche del suono.

Nella voce umana le formanti sono dovute alle risonanze del tratto vocale, ma esse possono essere individuate anche nell'emissione di strumenti acustici, come ad esempio, nei cordofoni, le formanti sono il risultato delle risonanze della tavola armonica.

Il tratto vocale e la tavola armonica sono esempi di filtri che, applicati alla vibrazioni di corde, creano formanti atti al riconoscimento del timbro di un determinato strumento.

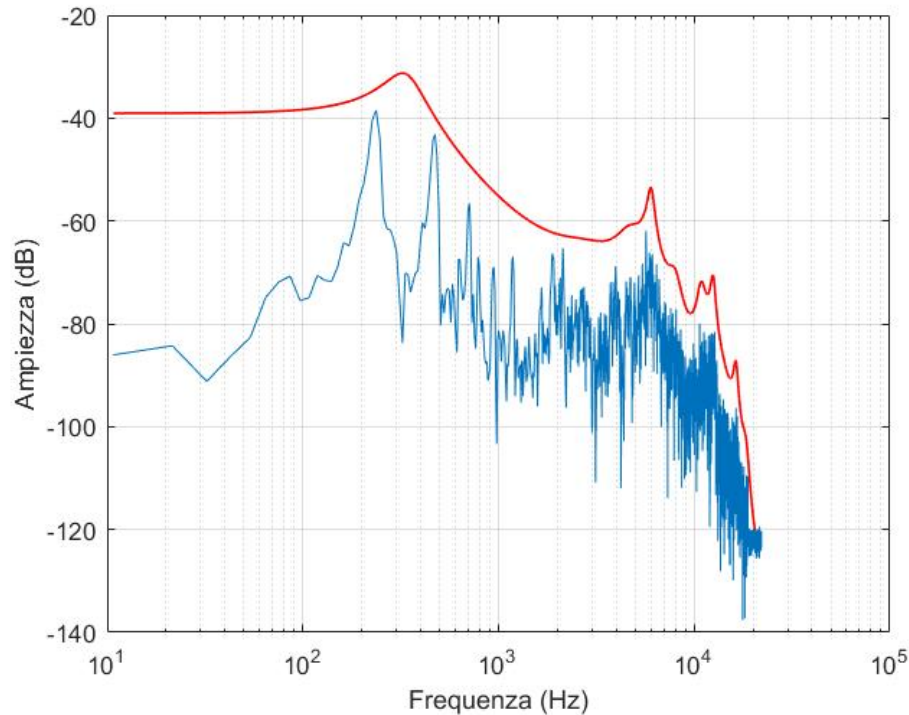


Figura 3: Esempio di formanti (linea continua rossa)

## 2.2 Analisi segnali nel dominio della frequenza

Dopo aver descritto il processo di digitalizzazione del segnale nel dominio del tempo, verranno ora fornite informazioni di base sull'analisi in frequenza dei segnali.

### 2.2.1 Segnali Periodici

Una delle classi di segnali cui si farà riferimento all'interno di questo elaborato è quella dei segnali che si ripetono periodicamente nel tempo, detti segnali periodici. Più precisamente diremo che un segnale  $f(t)$  è periodico al periodo  $T$  se si verifica che

$$f(t) = f(t + kT), \quad \forall t \in \mathbb{R}$$

Il minimo valore del periodo  $T > 0$  che soddisfa la definizione di periodicità è chiamato periodo fondamentale ed è denotato da  $T_0$ . L'inverso di questo valore rappresenta la frequenza fondamentale del segnale.

### 2.2.2 Serie di Fourier

I segnali periodici possono essere rappresentati come combinazione lineare di funzioni sinusoidali complesse mediante l'utilizzo di concetti matematici che prendono il nome di serie e trasformate di Fourier<sup>1</sup>.

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_0 t} \quad \text{con} \quad c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-in\omega_0 t} dt$$

Queste due equazioni matematiche prendono rispettivamente il nome di equazione di sintesi ed equazione di analisi della serie di Fourier. La prima consente di sintetizzare il segnale  $f(t)$  sovrapponendo le singole funzioni complesse, la seconda, invece, consente di scomporlo calcolandone i coefficienti complessi. Mediante l'identità di Eulero è possibile riscrivere queste equazioni utilizzando funzioni trigonometriche.

$$\cos(\theta) + i\sin(\theta) = e^{i\theta}$$

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{in\omega_0 t} = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(n\omega_0 t) + b_n \sin(n\omega_0 t)]$$

Dall'equazione precedente si nota che tutti i segnali periodici possono essere rappresentati come sovrapposizioni di funzioni trigonometriche con frequenza multipla di una frequenza data.

### 2.2.3 Trasformata di Fourier

La trasformata di Fourier consente un'estensione della serie di Fourier e permette la rappresentazione in frequenza di funzioni che, non essendo periodiche, non ammettono una trasformazione in serie di Fourier.

Ammettiamo un qualunque segnale come periodico di periodo  $T$  con  $T \rightarrow \infty$  un segnale  $f(t)$  non periodico, chiamiamo  $f_T(t)$  la funzione periodica di periodo  $T$  che coincide con  $f(t)$  nell'intervallo  $[-\frac{T}{2}, \frac{T}{2}]$ ; possiamo scrivere:

$$f(t) = \lim_{T \rightarrow \infty} f_T(t)$$

Essendo  $f_T(t)$  una funzione periodica, può essere sviluppata in serie di Fourier:

---

<sup>1</sup>La trasformata di Fourier è una generalizzazione ai segnali non periodici

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{inw_0 t} \quad \text{con} \quad c_n = \frac{1}{T} \int_{-\frac{T}{2}}^{\frac{T}{2}} f(t) e^{-inw_0 t} dt$$

Sapendo che  $w_0 = 2\pi v$ , chiamiamo  $v_n = nv$  la frequenza dell' $n$ -esima armonica e  $\Delta v = v_{n+1} - v = \frac{1}{T}$  la distanza tra una frequenza e la successiva; ponendo  $F(v) = c_n T$  possiamo scrivere:

$$f_T(t) = \sum_{n=-\infty}^{\infty} F(v_n) e^{i2\pi v_n t} \Delta v \quad F(v_n) = \int_{-\frac{T}{2}}^{\frac{T}{2}} f_T(t) e^{-i2\pi v_n t} dt$$

Per  $T \rightarrow \infty$  e  $\Delta v \rightarrow 0$ , la serie precedente si riduce ad un integrale:

$$f(t) = \int_{-\infty}^{\infty} F(v) e^{i2\pi v t} dv \quad F(v) = \int_{-\infty}^{\infty} f(t) e^{-i2\pi v t} dt$$

$F(v)$  viene chiamata formula di analisi o *Trasformata di Fourier* e ne rappresenta il contenuto in frequenza del segnale  $f(t)$ , ovvero il suo contenuto spettrale.

$f(t)$  viene chiamata, invece, formula di ricostruzione o *Anti-trasformata di Fourier*, e permette la ricostruzione di  $f(t)$  a partire dal suo contenuto in frequenza:  $f(t)$  viene scomposta in un numero infinito di sinusoidi continue complesse.

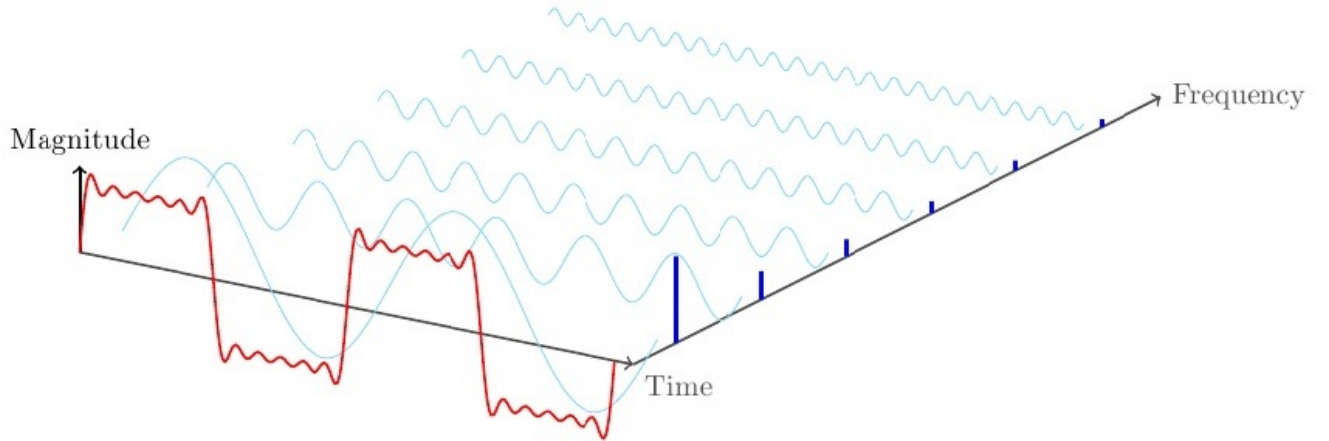


Figura 4: Dominio del tempo - frequenza

La trasformata e l'anti-trasformata di Fourier operano su segnali continui nel tempo e delle frequenze.

Per poterle operare in un calcolatore digitale è necessario adottare una discretizzazione dei segnali, introducendo una versione discreta della trasformata, la trasformata discreta di Fourier (DFT).

**DFT**

Consideriamo il segnale  $f_s(t)$  ottenuto campionando  $f(t)$  ai tempi  $n\tau$  ( $-\infty < n < \infty$ ), con  $\tau$  corrispondente il passo di campionamento, mediante una funzione impulsiva  $\delta(t)$ . L'informazione contenuta nel segnale  $f(t)$  viene approssimata con quella contenuta nel vettore discreto  $x$  formato da  $N$  campioni del segnale campionato con passo  $\tau$ .

$$x(n) = f(n\tau) \quad \text{con } n = 0, \dots, N-1$$

L'informazione in frequenza di questa funzione viene calcolata mediante la trasformata di Fourier dell'espressione precedente; essendo il segnale considerato un segnale campionato e quindi discreto, questa operazione prende il nome di *Trasformata di Fourier a tempo Discreto* (*DFT*); sapendo che  $F(\delta(t - n\tau)) = e^{-i2\pi vn\tau}$  otteniamo:

$$F_s(v) = \sum_{n=-\infty}^{\infty} x(n)e^{-2\pi vn\tau}$$

La funzione  $F_s(v)$  rappresenta lo spettro continuo del segnale  $x(n)$ , è necessario campionare anche il dominio della frequenza con intervalli equidistanti di ampiezza pari a  $F_0 = \frac{f_s}{N}$  con  $f_s = \frac{1}{\tau} = \text{frequenza di campionamento}$  dove  $F_0$  corrisponde alla risoluzione frequenziale della *DFT*.

Sapendo che :

$$X(k) = F_s(kF_0) \quad \text{con } k = 0, \dots, N-1$$

otteniamo:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi kF_0n\tau} = \sum_{n=0}^{N-1} x(n)e^{-i2\pi k\frac{1}{N\tau}n\tau} = \sum_{n=0}^{N-1} x(n)e^{-ik\frac{2\pi}{N}n}$$

Quindi:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-ik\frac{2\pi}{N}n} \quad \& \quad x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k)e^{ik\frac{2\pi}{N}n}$$

Queste due espressioni prendono il nome di trasformata e anti-trasformata discrete di Fourier.

Essendo il risultato della DFT una successione di numeri complessi, la trasformata permette di ottenere informazioni riguardo l'ampiezza e la fase delle diverse componenti sinusoidali del segnale in ingresso.

Queste due informazioni possono essere ottenute esprimendo i valori complessi in forma polare, ottenendo così l'ampiezza  $A_k$  e la fase  $\phi_k$  delle sinusoidi rispettivamente dal modulo e argomento di  $X_k$ :



$$A_k = |X_k| = \sqrt{\operatorname{Re}(X_k)^2 + \operatorname{Im}(X_k)^2} \quad \& \quad \varphi_k = \arg(X_k) = \operatorname{atan}(\operatorname{Im}(X_k), \operatorname{Re}(X_k))$$

### STFT

La trasformata di Fourier permette di ottenere informazioni delle componenti armoniche nel segnale, ma non permette di avere una valutazione temporale di quando tali frequenze siano effettivamente presenti, occorre inserire una dipendenza dal tempo nella trasformazione. Dividendo il segnale in segmenti di lunghezza fissa, si riescono ad avere informazioni circa la variazione in frequenza nel tempo effettuando la trasformata di Fourier per ogni segmento finestrato:

$$STFT_x(\tau, v) = \int_{-\infty}^{+\infty} x(t)g(t - \tau)e^{i2\pi vt} dt$$

L'espressione precedente prende il nome di trasformata di Fourier a breve termine o Short Time Fourier Transform (STFT).

Il segnale viene modificato dal comportamento della finestra presa in esame<sup>2</sup>, di conseguenza la STFT fornisce lo spettro del segnale alterato dalla presenza della finestra.

In questo senso una finestra che nel dominio del tempo provoca al segnale rapide transizioni, nel dominio della frequenza causa una forte dispersione *leakage* della potenza su tutto lo spettro come ad esempio la finestra rettangolare.

Le caratteristiche fondamentali di una finestra sono : il fattore di dispersione, la larghezza del Mainlobe(lobo principale) e l'attenuazione in dB del Sidelobes(lobi laterali).

Una finestra ideale dovrebbe avere un fattore di dispersione uguale a zero, una larghezza del Mainlobe molto stretta e un'attenuazione in dB del Sidelobe molto alta. Di seguito si illustrano due esempi di due finestre : la finestra rettangolare e la finestra di Hamming.

---

<sup>2</sup>Esistono molteplici tipologie di finestre : Hamming, Hann, Blackman, Gauss, Blackman-Harris etc...

Nella figura 5 si può notare una bassissima attenuazione del Sidelobe : intorno ai  $-13$  dB, un fattore di dispersione del 9%

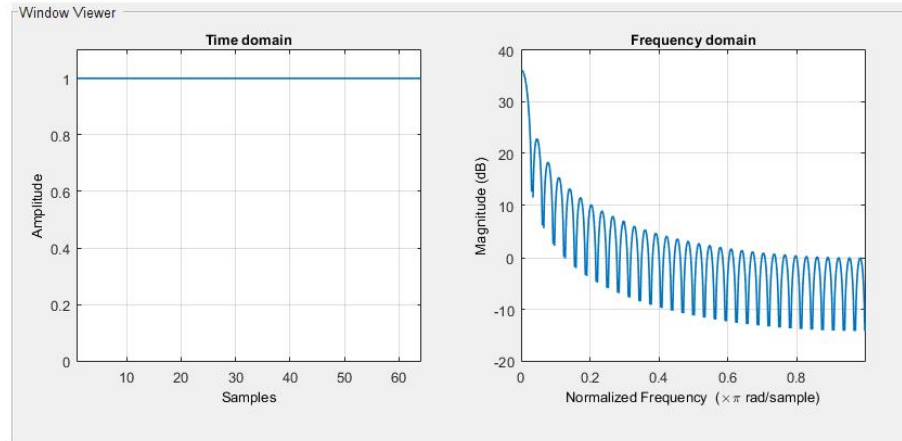


Figura 5: Finestra rettangolare

Nella figura 6 si può notare, invece che l'attenuazione dei sidelobes è molto più alta rispetto alla finestra rettangolare, questo comporterebbe una miglior lettura della STFT; e un fattore di dispersione del 0.03%<sup>3</sup>

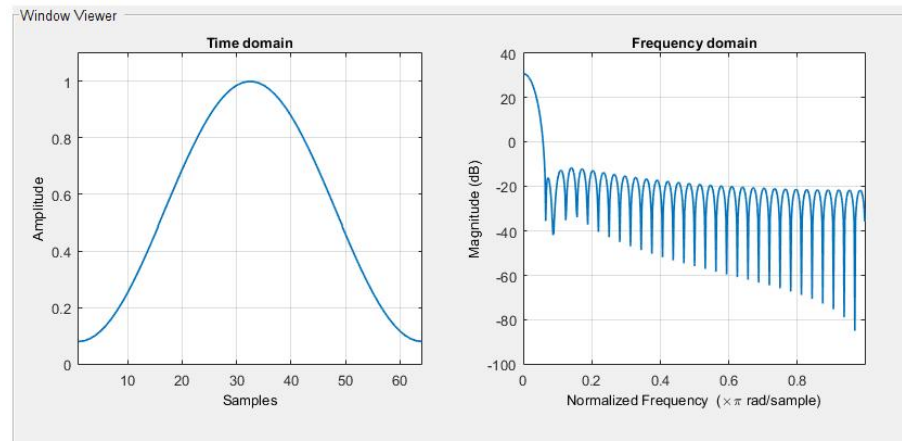


Figura 6: Finestra di Hamming

La finestra utilizzata per la maggior parte all'interno dell'elaborato è la finestra di Hamming perchè garantisce un ottimo compromesso tra larghezza del mainlobe (una

<sup>3</sup>I dati esposti sono stati elaborati tramite il tool in MATLAB "vytool"

buona precisione frequenziale) e una notevole attenuazione dei sidelobes (miglior definizione dello spettrogramma).

Di seguito si illustra un esempio di sonogramma di un fraseggio di chitarra.

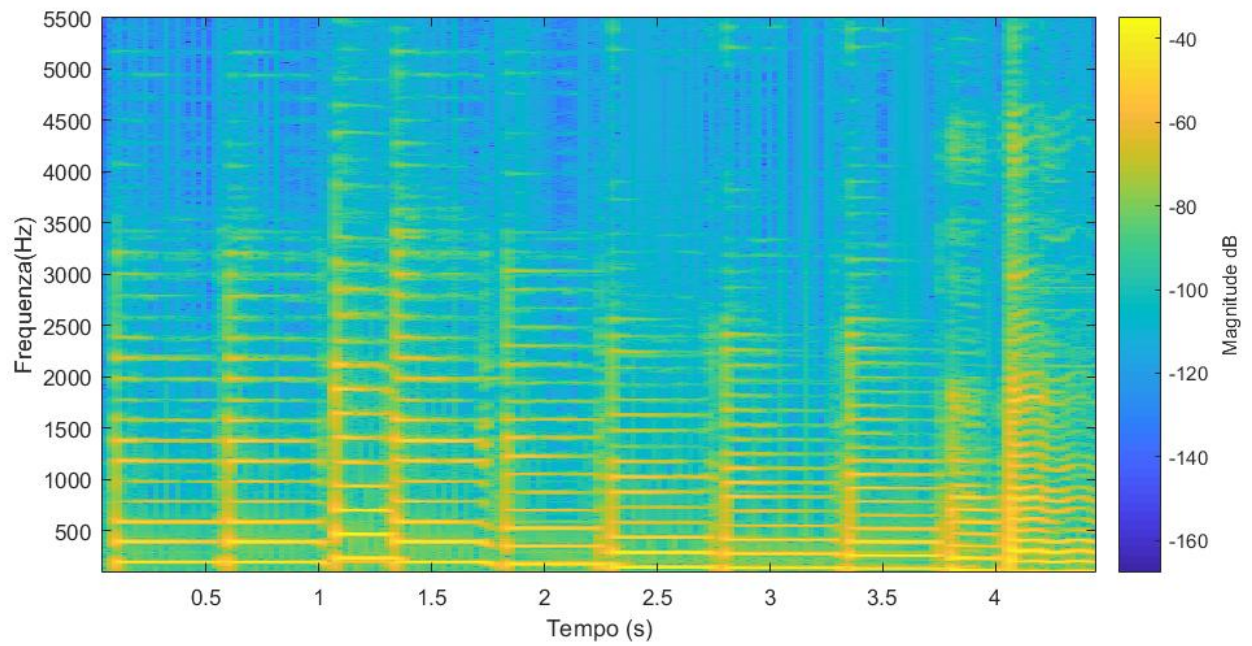


Figura 7: Sonogramma



# Capitolo 3

## Modello di Analisi

Come già accennato nell'introduzione, questo lavoro effettua analisi di coppie di segnali simili, ne estrae *features* caratteristiche e ne calcola le differenze.

In questo capitolo si illustrerà il modello di analisi utilizzato e le relative implementazioni utilizzando come segnali di test segnali ideali per convalidare il modello, per poi analizzare suoni reali come chitarre e voci.

Verranno generati toni puri di varia frequenza e chirp per convalidare il modello *Analisi armonica - Pitch, harmonic tracking* (3.2); verranno utilizzati impulsi e toni puri per convalidare il modello *Analisi attacco delle note - Onset Detection* (3.4); Per quanto riguarda la sezione *Analisi della dinamica - Potenza, intensità* (3.1 e *Analisi timbrica - Rilevamento delle formanti* (3.3) verranno utilizzati, invece, segnali complessi come chitarre e voci.

### 3.1 Analisi della dinamica

In questa sezione verranno estratte *features* che caratterizzano la dinamica di un segnale audio, ovvero verrà calcolato: il *valore efficace* del segnale (RMS), il *valore massimo di ampiezza* (valore di picco) e il valore di *crest factor* (misura del range dinamico).

Prendiamo in considerazione un piccolo riff di chitarra elettrica, dopo averlo diviso in segmenti di lunghezza di 1024 campioni (23.2 ms = lunghezza finestra di analisi) utilizzando un overlapping del 75%, calcoliamo il valore efficace mediante la formula del *valore quadratico medio*:

$$x_{rms} = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i^2}$$

dove  $n$  è la lunghezza del segmento e  $x_i$  è il valore di ampiezza del segnale. In questo modo calcoliamo ogni 23.2 ms il valore efficace del suono, ovvero il valore efficace della variazione della pressione sonora.

Allo stesso modo calcoliamo il valore massimo di ampiezza e il crest factor<sup>1</sup> rispettivamente:

$$x_{picco} = \max(|x|) \quad \& \quad x_{crest} = \frac{x_{picco}}{x_{rms}}$$

Infine, dato che i valori del segnale sono rappresentati tra  $-1$  e  $1$ , dalla formula presente nella sottosezione 2.1.1, possiamo convertirli in decibel mediante la formula :

$$dB_{fs}(A) = 20 \cdot \log_{10}(A)$$

I risultati ottenuti sono rappresentati nella Fig. 8

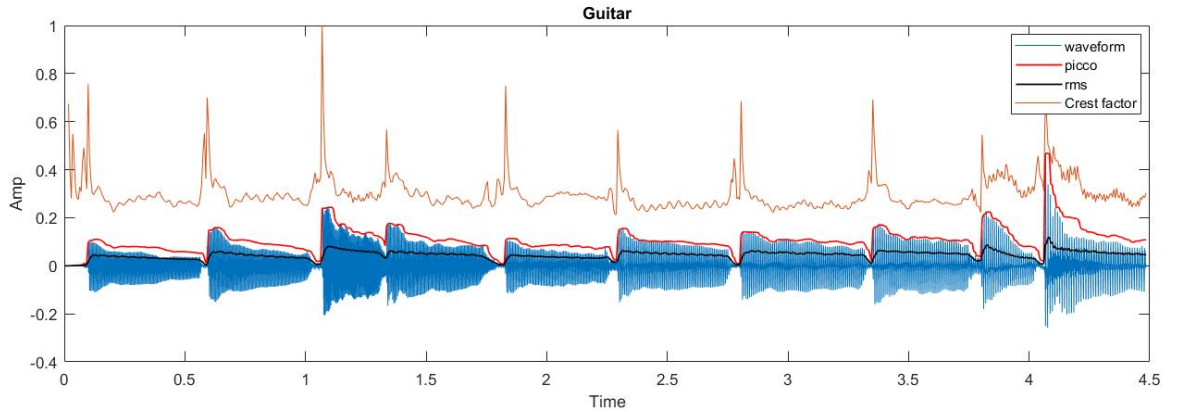


Figura 8: Esempio segnale audio - riff di chitarra

---

<sup>1</sup>Valore normalizzato tra 0 e 1

## 3.2 Analisi armonica - Harmonic tracking

Utilizzando la tecnica della doppia incisione, le tracce create sono molto simili tra di loro, ma non saranno mai identiche. Questo è dovuto dalle microvariazioni introdotte dall'esecutore durante la sua performance. Prendiamo in considerazione, ora, le variazioni in frequenza, per le variazioni nel tempo si faccia riferimento alla sezione (3.4).

Prendiamo come esempio un'incisione di chitarra, qualsiasi essa sia, le microvariazioni in frequenza sono dovute da molteplici fattori quali: umidità dell'aria, temperatura, qualità delle componenti dello strumento, pressione del dito sulla corda e la posizione del dito all'interno del tasto (*fret*).

In questa sezione si andrà ad esporre le tecniche utilizzate per il tracciamento delle armoniche illustrando i relativi risultati tramite segnali di test ideali (toni puri, chirp).

### 3.2.1 Interpolazione parabolica

Il primo metodo proposto prende in considerazione un modello matematico, il calcolo del vertice di una parabola passante per tre punti.

Nel capitolo precedente (sezione 2.2.3) si è visto l'importanza della tipologia della finestra di analisi per avere una miglior precisione sui picchi nel dominio della frequenza, per aumentare tale risoluzione possiamo aumentare la dimensione della finestra di analisi, questo comporterebbe ad avere una finestra molto grande, con lo svantaggio di avere un calcolo computazionale molto elevato. Nella figura 9 si illustra il rilevamento del picco in frequenza di una sinusoide di frequenza 880 Hz, si noti che entrambi i picchi non coincidono con il valore perfetto dell'oscillazione .

La soluzione a questo problema è effettuare un'interpolazione parabolica[2]. Una parabola è una funzione con la forma molto simile alla forma del lobo nello spettro della finestra di analisi, in scala logaritmica.

In un insieme di coordinate centrato a  $(k_\beta, 0)$ , dove  $k_\beta$  è numero del bin a cui fa riferimento il picco massimo di magnitudine  $mX(k_\beta)$  (designato come asterisco in figura), definiamo la funzione parabola tramite la forma

$$x(n) = a(n - p)^2 + b$$

tale che  $x(-1) = \alpha$ ,  $x(0) = \beta$ ,  $x(1) = \gamma$ , dove  $\alpha$ ,  $\beta$  e  $\gamma$  corrispondono ai valori dei tre *bin* più alti:

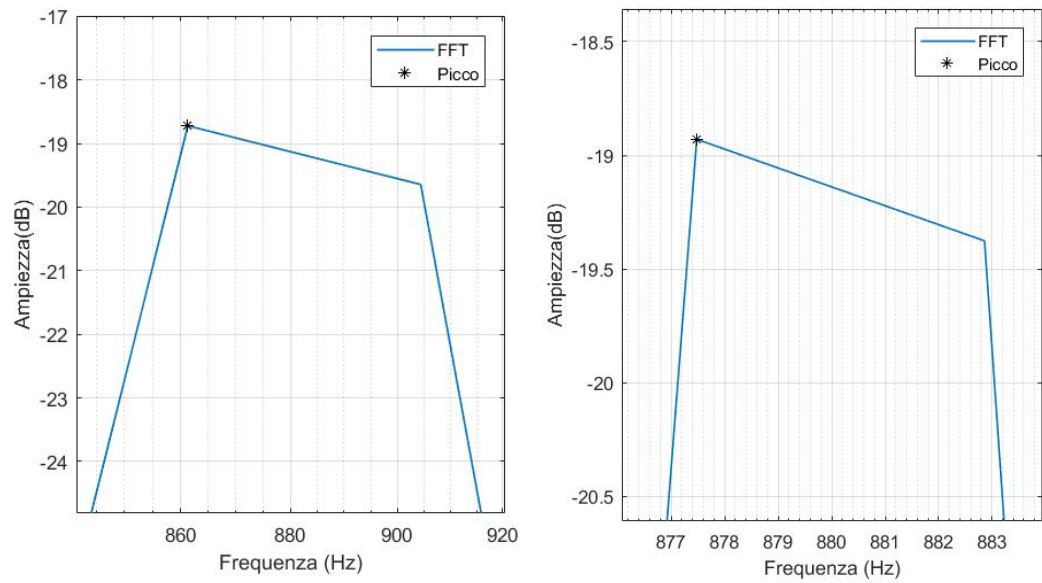


Figura 9: Picco in frequenza di una sinusoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

$$\alpha = mX[k_{\beta} - 1]$$

$$\beta = mX[k_{\beta}]$$

$$\gamma = mX[k_{\beta} + 1]$$

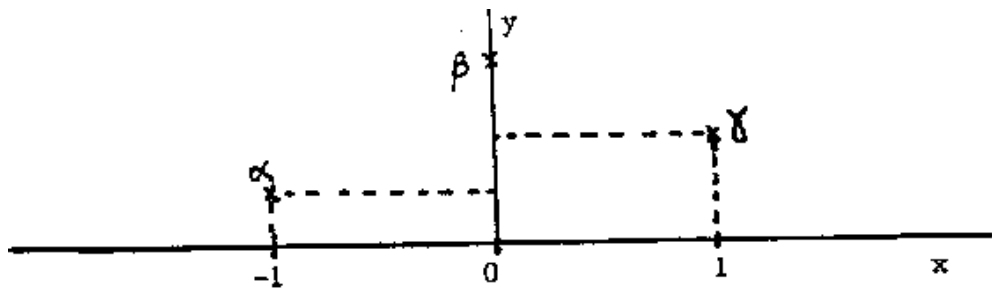


Figura 10: Coordinate dell'interpolazione parabolica

Dall'equazione generale della parabola calcoliamo la locazione  $p$  corrispondente al vertice della parabola ricostruita, ottenendo così lo scostamento dal bin del picco rilevato alla posizione  $(k_{\beta}, 0)$ :



$$p = \frac{1}{2} \frac{\alpha - \gamma}{\alpha - 2\beta + \gamma}$$

Il valore del bin corrispondente alla posizione veritiera si ottiene sommando lo scostamento appena rilevato alla posizione  $k_\beta$

$$\hat{k} = k_\beta + p$$

ottenendo il valore in Hz con la formula

$$\frac{f_s \cdot \hat{k}}{N}$$

dove  $f_s$  corrisponde alla *frequenza di campionamento*.

infine il valore di ampiezza corrispondente al vertice calcolato:

$$x(p) = \beta - \frac{1}{4}(\alpha - \gamma)p$$

Di seguito si illustra il risultato dell'interpolazione sulla sinusoide di test. La figura di sinistra rappresenta un picco al valore di 878.1297 Hz, mentre la figura di destra a 879.8692 Hz.

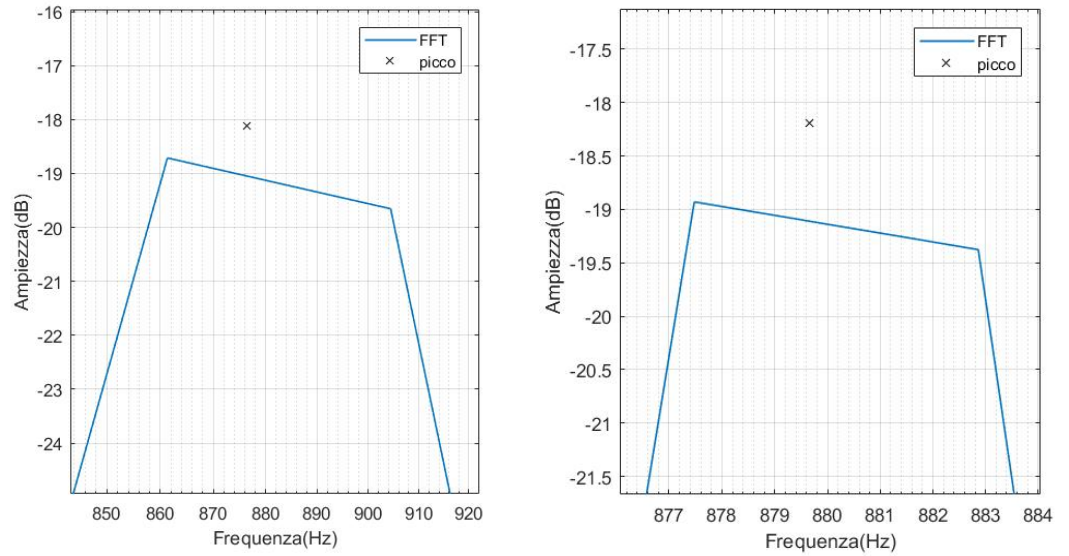


Figura 11: Interpolazione parabolica effettuata su una sinusoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

### 3.2.2 Differenza di fase

L'interpolazione parabolica permette di estrarre valori vicini al valore veritiero, ma nel nostro caso non è abbastanza preciso. Per avere una precisione maggiore si descrive un metodo che prende in considerazione la fase del segnale.

Come si è descritto in precedenza, un segnale può essere scomposto come somma di sinusoidi complesse con valori ben definiti di *frequenza*, *fase* e *magnitudine*, questi valori caratterizzano una sinusoide in un dato istante (*frame*).

La frequenza della sinusoide viene discretizzata in base alla risoluzione spettrale  $\frac{f_s}{N}$ , dove  $f_s$  corrisponde alla frequenza di campionamento e  $N$  corrisponde alla lunghezza della finestra di analisi, questo dato corrisponde al valore in frequenza tra un bin e quello successivo. Nel caso in cui la frequenza del segnale coincida esattamente con la frequenza del bin (valore multiplo di  $\frac{f_s}{N}$ ), lo spettro risulterà assente di *leakage* perchè la finestra non provocherà rapide transizioni dovute al troncamento del segnale. Nel caso in cui la frequenza del segnale non sia multipla di  $\frac{f_s}{N}$  il segnale subirà dei ripidi troncamenti alle estremità della finestra. Di seguito si illustra una sinusoide divisa in sette frame.

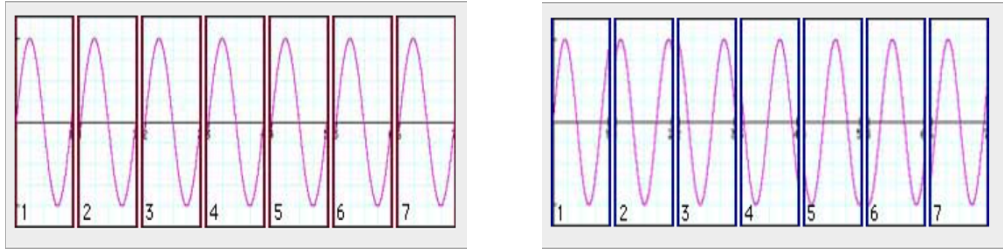


Figura 12: A sinistra una sinusoide con frequenza uguale a quella del *bin*, a destra una sinusoide con frequenza non corrispondente a quella del *bin*

Come si può notare, nella figura di sinistra, ad ogni frame il segnale inizia con lo stesso valore di fase, al contrario invece, il segnale nella figura di destra, che ha una frequenza compresa tra due bin, non inizia ogni frame con la stessa fase, chiaramente siamo in presenza di uno *scostamento di fase*. Più è alto è lo scostamento, maggiore sarà la deviazione tra il bin di riferimento.

La variazione dello scostamento di fase viene usata per calcolare il valore esatto della frequenza della nostra sinusoide.

Prendendo in considerazione il noto modello di estrazione della frequenza istantanea  $f_i[k]$  di Flanagan e Golden [3], si ottiene:

$$f_i[k] = (k + \kappa[k]) \frac{f_s}{N}$$

con:

$$\kappa[k] = \frac{N}{2\pi L} \text{princarg} \left[ \phi_l[k] - \phi_{l-1}[k] - \frac{2\pi L}{N} k \right]$$

dove  $L$  indica il numero di campioni tra una finestra e quella successiva, comunemente chiamato *stepsize* oppure *hopsiz*e,  $k$  corrisponde al bin in frequenza,  $\phi_l[k]$  indica il valore della fase al  $k$ -esimo bin del frame  $l$  e  $\kappa$  corrisponde alla deviazione in bin della frequenza istantanea dal bin di riferimento. La funzione *princarg* permette di mappare il valore di fase tra  $\pm\pi$ .

Di seguito si illustra il risultato ottenuto calcolando la frequenza istantanea sulla senoide utilizzata nell'interpolazione parabolica. La figura di sinistra rappresenta un picco al valore di 879,7985 Hz, mentre la figura di destra a 879,9991 Hz.

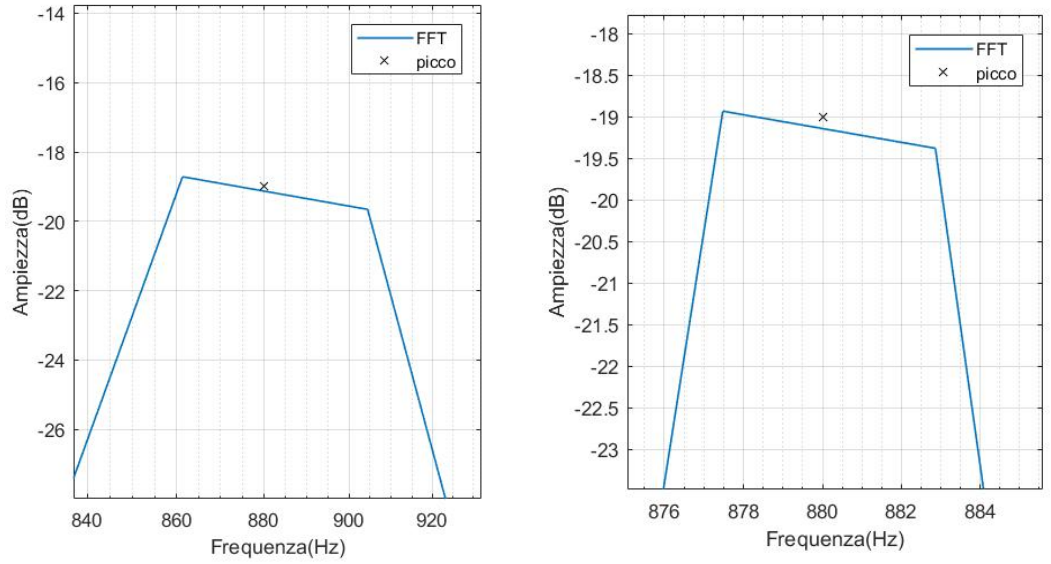


Figura 13: Interpolazione parabolica effettuata su una senoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

Infine si illustra il risultato sull'intero file audio, rappresentando entrambi i metodi in un'unica figura.

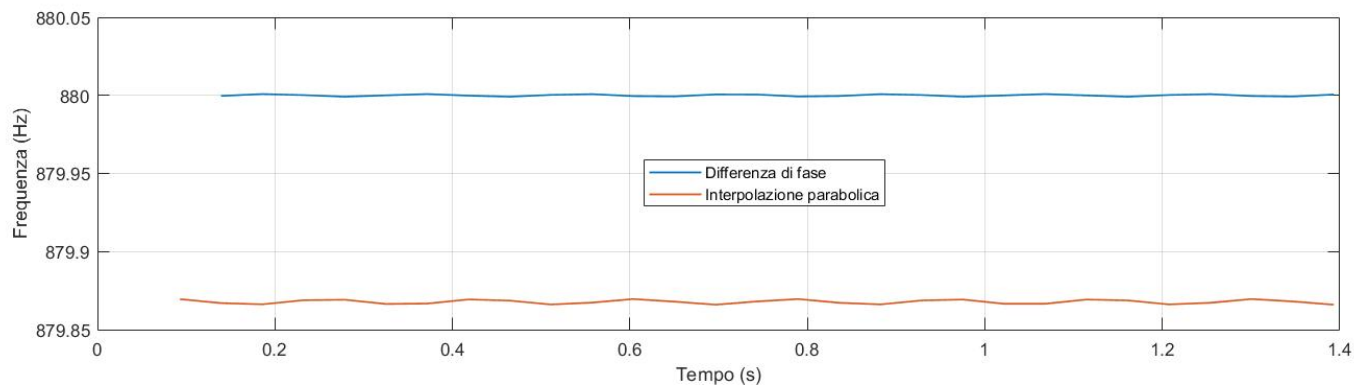


Figura 14: Confronto tra i due metodi

Calcolando la media delle singole tracce, l'implementazione che utilizza la differenza di fase presenta un valore di 880.0000 Hz. Il modello dell'interpolazione parabolica, invece, ha un valore medio di 879.8678 Hz.

### 3.2.3 Validazione dei modelli

Ai fini della validazione di modelli implementato si effettuano ulteriori test su un segnale chirp (da 200 a 500 Hz) e su un segnale complesso quale un riff di chitarra.

#### Chirp

Prendiamo in considerazione un segnale ideale che varia la sua frequenza nel tempo, ad esempio un segnale chirp di una durata di tre secondi variando la sua frequenza da 200 Hz a 500 Hz con andamento lineare. Si illustra di seguito i risultati ottenuti da entrambe le implementazioni e il tracciamento del modello più preciso.

In figura 15 viene illustrato un frammento di 0.4 secondi nel passaggio tra i 255 Hz a 295 Hz. Si noti la poca precisione del modello dell'interpolazione parabolica in confronto a quella della differenza di fase. L'andamento lineare in frequenza viene rappresentato correttamente dalla linea blu definendo stabile l'implementazione ottenuta mediante differenza di fase.

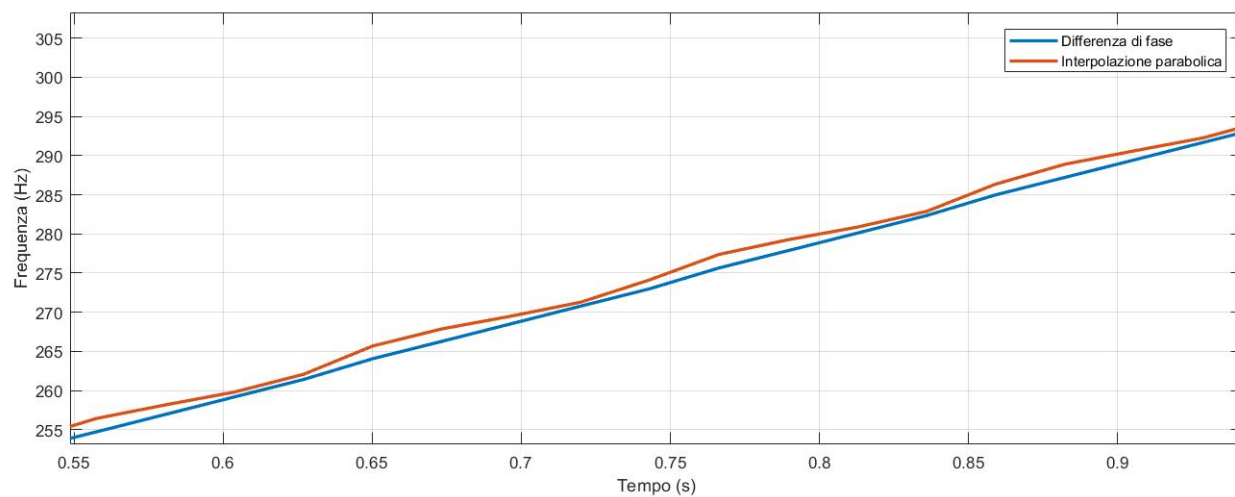


Figura 15: Chirp - Differenza di fase e Interpolazione parabolica

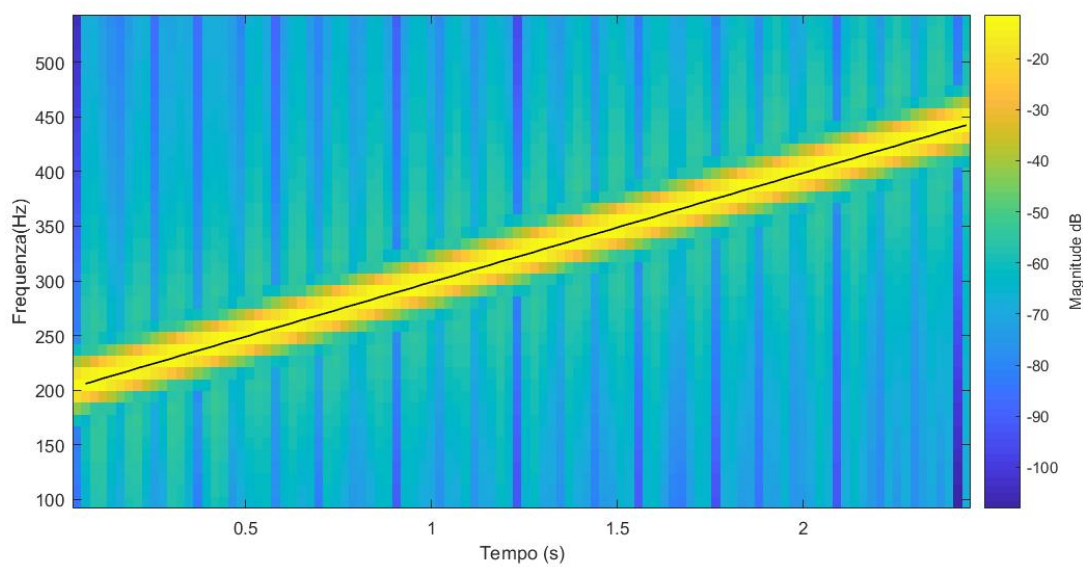


Figura 16: Chirp - Sonogramma e relativo tracciamento

### Riff di chitarra

Come ultima validazione viene effettuato un test su un segnale complesso dove il suo contenuto spettrale non è più composto da una singola componente, ma di una frequenza fondamentale e dalle sue armoniche.

Prendiamo in esame l'introduzione di chitarra del brano "Wither" del gruppo musicale statunitense Dream Theater. Di seguito si illustra il relativo spettrogramma e il tracciamento delle armoniche (Fig. 17).

Il tracciamento viene effettuato seguendo l'andamento dei singoli picchi all'interno del FFT per ogni frame di analisi, analizzando il loro percorso tenendo in considerazione lo scostamento in frequenza tra una nota e quella successiva e scartando errori dovuti da artefatti prodotti dalla finestratura come il rilevamento dei sidelobe fissando una soglia di ampiezza in relazione al valore di SNR prodotto dalla finestra per evitare di rilevare componenti di ampiezza troppo bassa denominati come *garbage peak*.

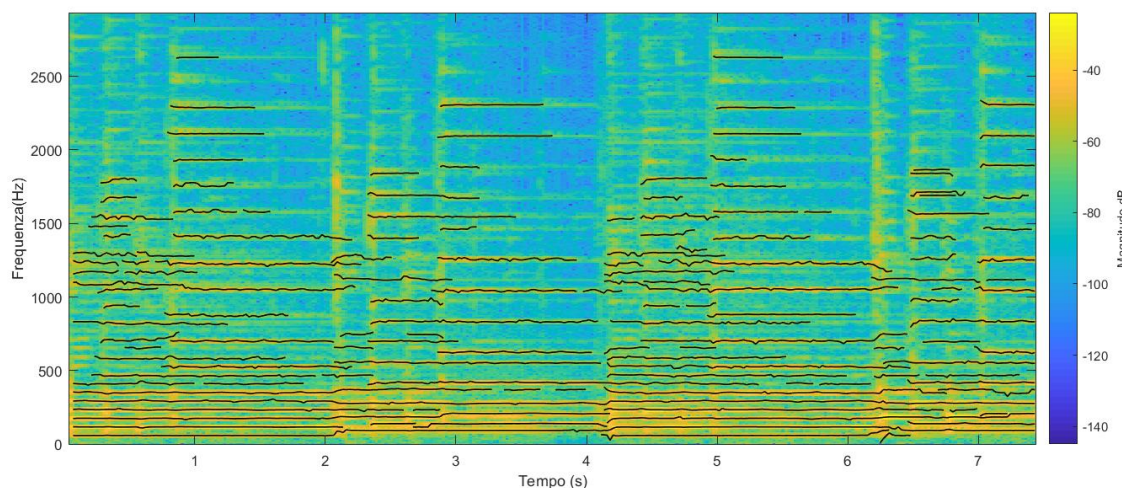


Figura 17: Spettrogramma con il tracciamento delle singole componenti

## 3.3 Analisi timbrica - Rilevamento delle formanti

Come si è accennato nel capitolo precedente (Sez. 2.1.1) il timbro è quella caratteristica del suono che permette all'ascoltatore di distinguere due suoni.

In questa sezione si andranno ad estrarre *feature* che caratterizzano il timbro del segnale, le *formanti*. Questa caratteristica oltre ad identificare il suono permette inoltre di valutare le modalità di esecuzione della performance di un musicista, ad esempio

la posizione della mano destra o di un plettro durante l'esecuzione di una chitarra elettrica. Definendo così un'analisi di equalizzazione generale, utile per definire un profilo

### **3.4 Analisi attacco delle note - Onset Detection**

## Capitolo 4

### Test effettuati e relativi risultati

4.1 Strumento : Chitarre

4.2 Strumento : Voci



## Capitolo 5

### Conclusioni e sviluppi futuri

# Bibliografia

- [1] J. R. Pierce, *La Scienza del Suono*, Zanichelli, 1988.
- [2] Julius O. Smith III, Xavier Sierra, *PARSHL: An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation*, Stanford University, California, 1987.
- [3] J. L. Flanagan & M. Golden, "Phase Vocoder", *Bell System Technical Journal*, vol.45 pp. 1493-1509, 1966.
- [4] Karin Dressler, *Sinusoidal Extraction using an Efficient Implementation of a Multi-Resolution FFT*, Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany, 2006.
- [5] De La Cuadra, Patricio, Aaron Master & Craig Sapp. *Efficient pitch detection techniques for interactive music*. Proceedings of the 2001 International Computer Music Conference, 2001.
- [6] L. Basso, "Capitalismo monopolistico e strategia operaia", *Problemi del socialismo*, vol. 8, n. 5, pp. 585-612, 1962.
- [7] L. Avirovic, J. Dodds (a cura di), *Atti del Convegno internazionale "Umberto Eco, Claudio Magris. Autori e traduttori a confronto"* ( Trieste, 27-28 novembre 1989), Udine, Campanotto, 1993.
- [8] E.L. Gans, "The Discovery of Illusion: Flaubert's Early Works, 1835-1837", unpublished Ph.D. Dissertation, Johns Hopkins University, 1967.
- [9] R. Harrison, *Bibliography of planned languages (excluding Esperanto)*. <http://www.vor.nu/langlab/bibliog.html>, 1992, agg. 1997.