

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

In questo elaborato viene descritto il progetto effettuato presso IK Multimedia (<http://www.ikmultimedia.com/>), software house italiana con sede a Modena (Mo). Il progetto svolto prende in considerazione una tecnica utilizzata nelle produzioni musicali, ovvero la *doppia incisione*. La tecnica della doppia incisione è una tecnica molto usata da svariati produttori discografici e fondamentale per garantire una notevole resa stereofonica da parte dell'ascoltatore. Durante il processo di registrazione in una produzione musicale si è soliti incidere le chitarre e altri strumenti due, quattro, otto volte la medesima parte musicale per poi ripartire a proprio piacimento le singole tracce nell'immagine stereo in modo da avvolgere l'ascoltatore. In questo elaborato si prenderanno in considerazione solamente una coppia ripartita una destra e una a sinistra nella suddetta immagine.

Lo scopo del progetto è di estrarre le differenze di *Dinamica*, *Frequenza*, *Timbro* e *Tempo* tra le due tracce in esame utilizzando *MATLAB*.

La tesi è organizzata come segue:

- nel Capitolo 1 verrà effettuata un'introduzione sul progetto svolto presso IK Multimedia andando a descrivere in maniera generale le procedure utilizzare per raggiungere gli obiettivi concordati.
- nel Capitolo 2 verrà fornita una breve panoramica sui fondamenti teorici utili alla comprensione degli argomenti esposti.
- nel Capitolo 3 verrà presentata quella che è lo stato dell'arte per l'estrazione delle features finalizzata all'estrazione della dinamica, l'estrazione del contenuto frequenziale analizzando e confrontando due modelli (interpolazione parabolica e differenza di fase), verranno descritti inoltre l'estrazione del timbro, in particolar modo finalizzato ad individuare le formanti di un suono e infine verrà presentato un modello per la rilevazione degli *onset*, ovvero l'inizio di un suono.
- nel Capitolo 4 vengono presentati i test e i risultati ottenuti.

# Capitolo 1

## Introduzione

## Capitolo 2

### Fondamenti Teorici

# 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 e chirp per convalidare il modello *Analisi armonica - Harmonic tracking* (3.2); verranno utilizzati impulsi per validare la precisione del modello *Analisi attacco delle note - Onset Detection* (3.4); Per quanto riguarda la sezione *Analisi della dinamica* (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*:

### CAPITOLO 3. MODELLO DI ANALISI

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

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, possiamo convertirli in decibel mediante la formula:

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

I risultati ottenuti sono rappresentati nella Fig. 1

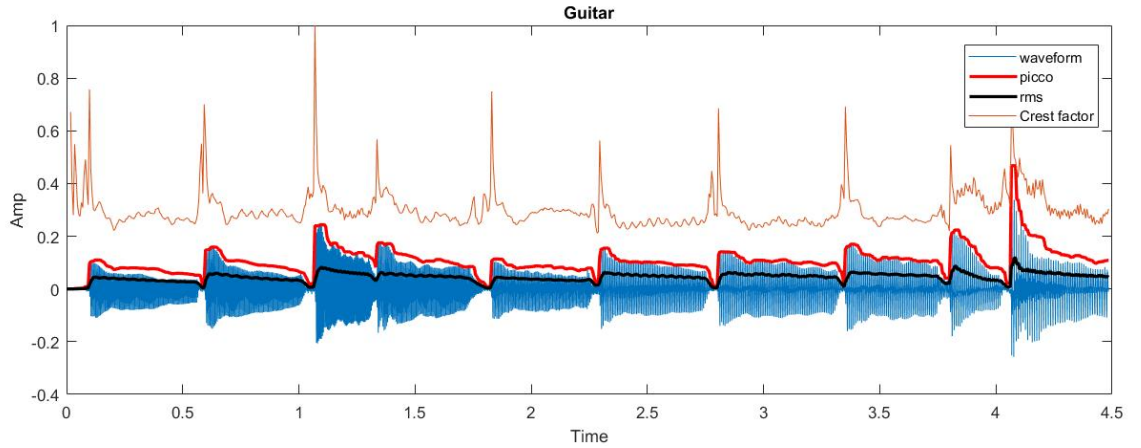


Figura 1: Esempio segnale audio - riff di chitarra

---

<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) 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 2 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:



### CAPITOLO 3. MODELLO DI ANALISI

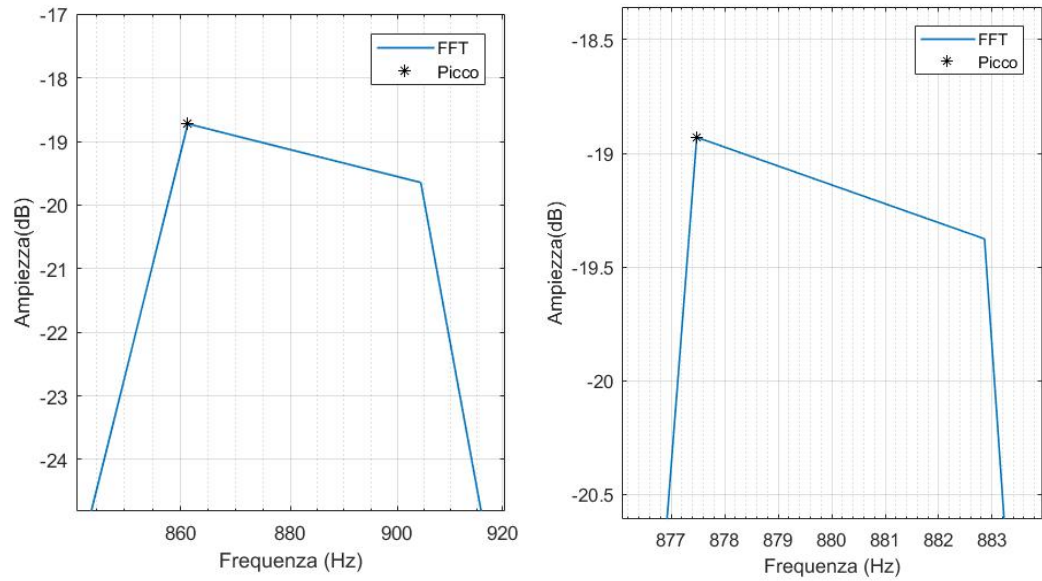


Figura 2: Picco in frequenza di una sinusoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

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

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

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

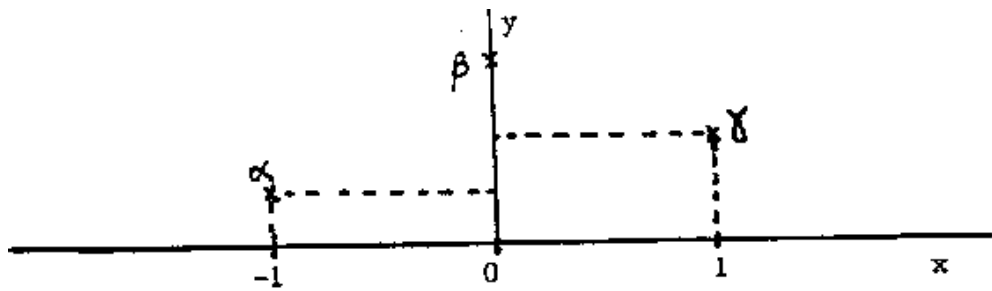


Figura 3: Coordinate dell'interpolazione parabolica

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  $(k_{\beta}, 0)$ :

### CAPITOLO 3. MODELLO DI ANALISI

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

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.

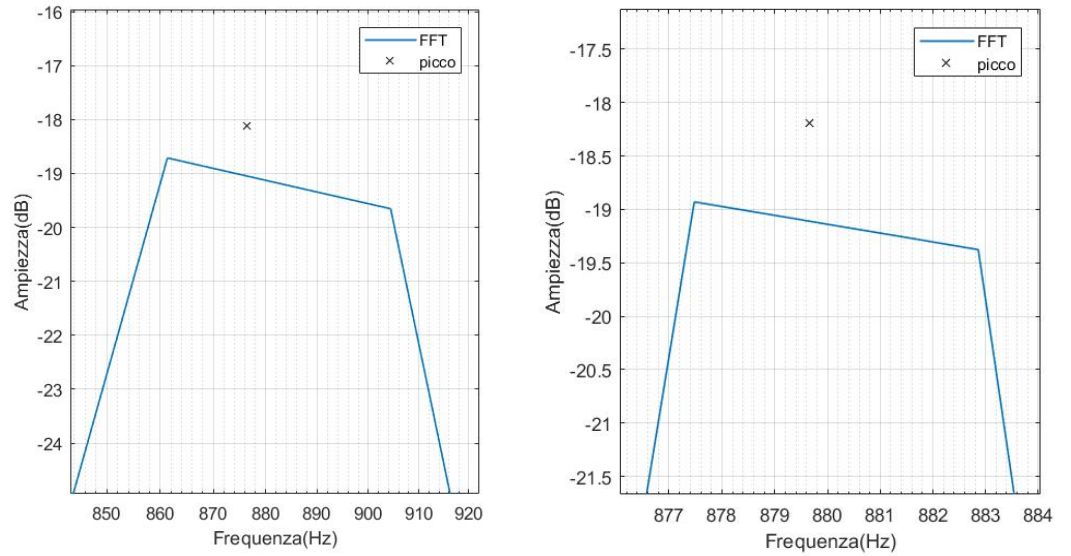


Figura 4: Interpolazione parabolica effettuata su una sinusoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

### 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.

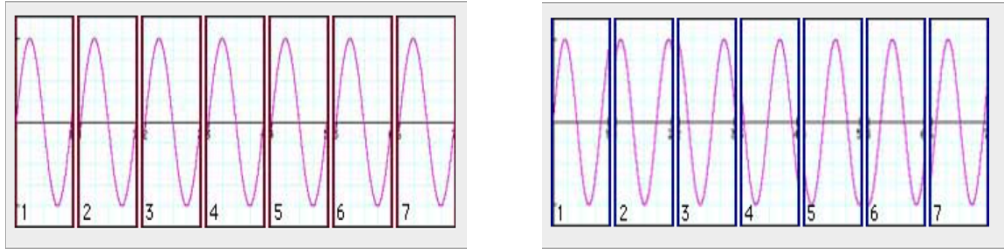


Figura 5: A sinistra una sinusoide con frequenza uguale a quella del *bin*, a destra una sinusoide con frequenza non corrispondente a quella del *bin*

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 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:

### CAPITOLO 3. MODELLO DI ANALISI

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

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 sinusoide 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.

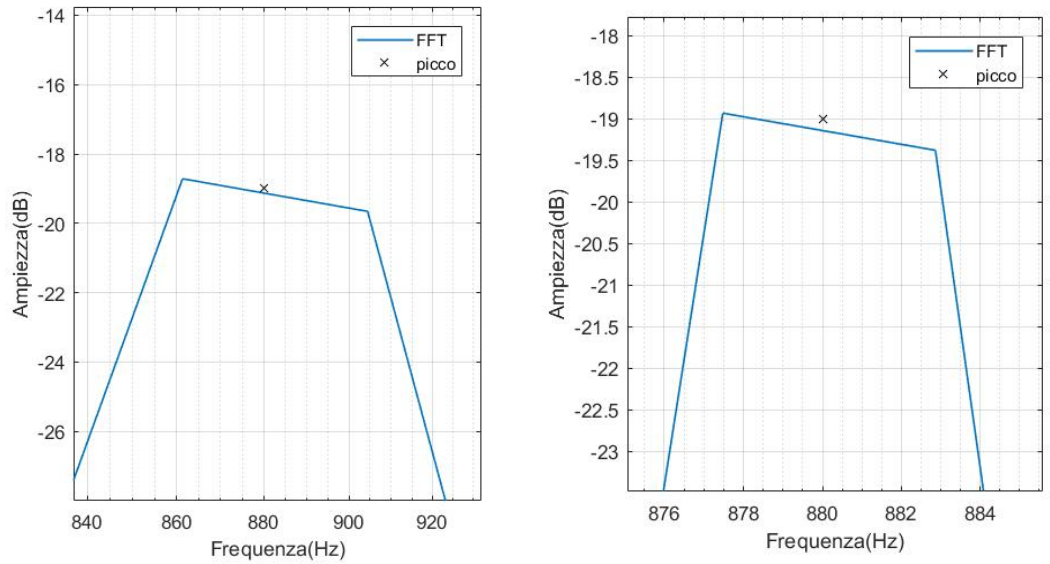


Figura 6: Differenza di fase effettuata su una sinusoide di 880 Hz calcolando la FFT con finestre di 1024 campioni (a sinistra) e 8182 campioni (a destra).

### CAPITOLO 3. MODELLO DI ANALISI

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

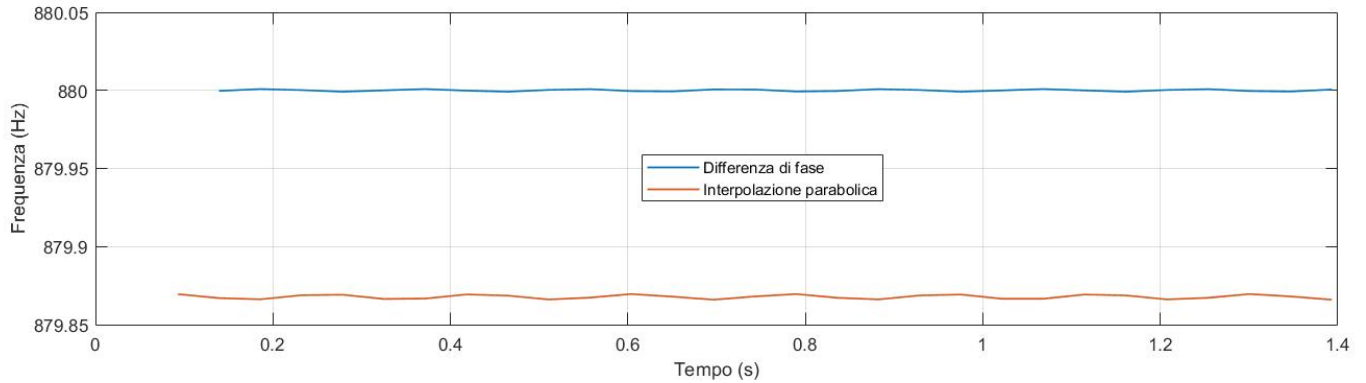


Figura 7: Confronto tra i due metodi

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 dei modelli implementati 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 8 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.

### CAPITOLO 3. MODELLO DI ANALISI

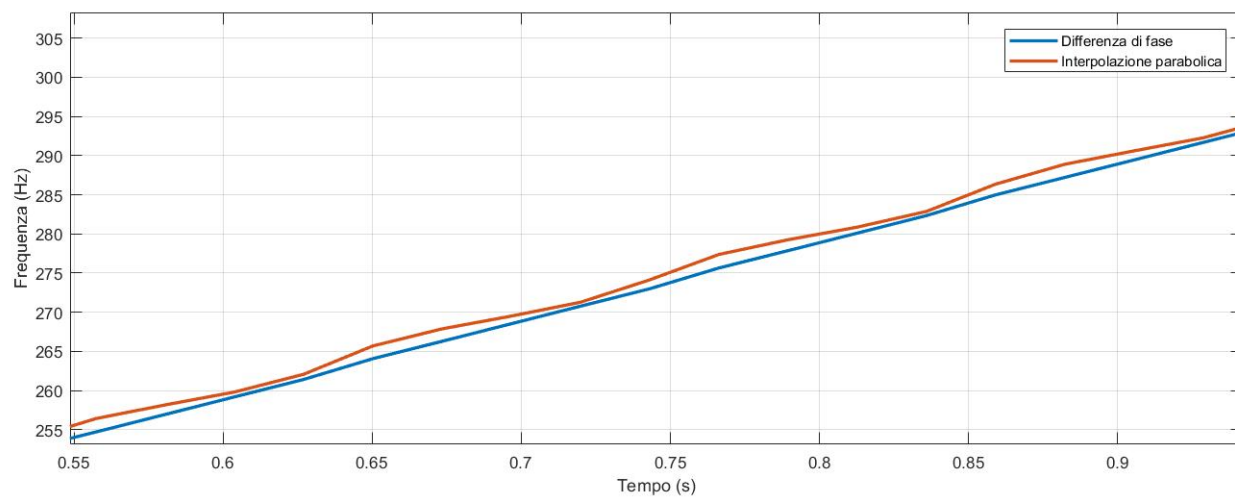


Figura 8: Chirp - Differenza di fase e Interpolazione parabolica

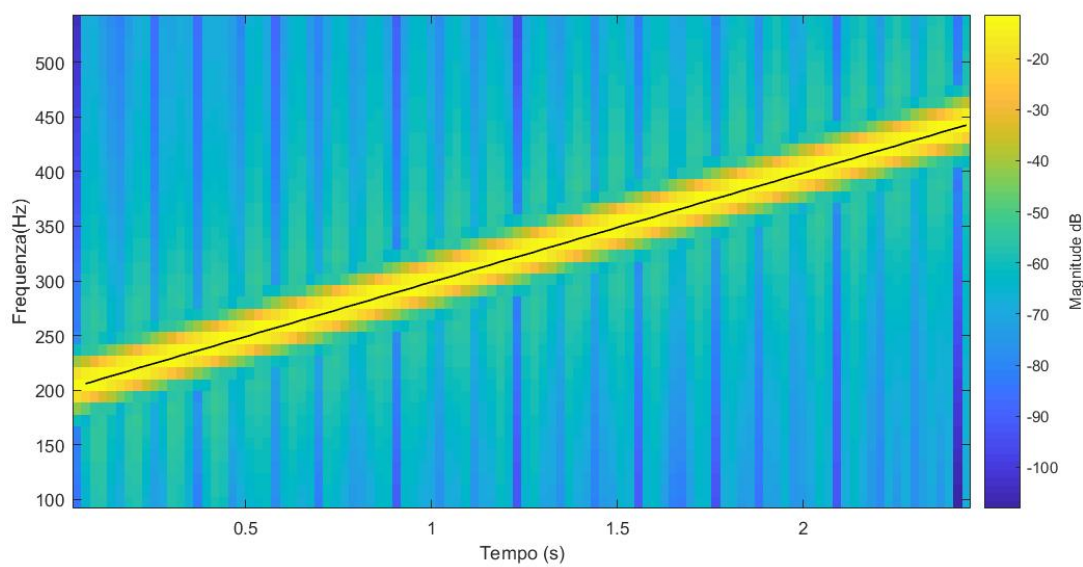


Figura 9: Chirp - Sonogramma e relativo tracciamento



#### 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. 10).

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*.

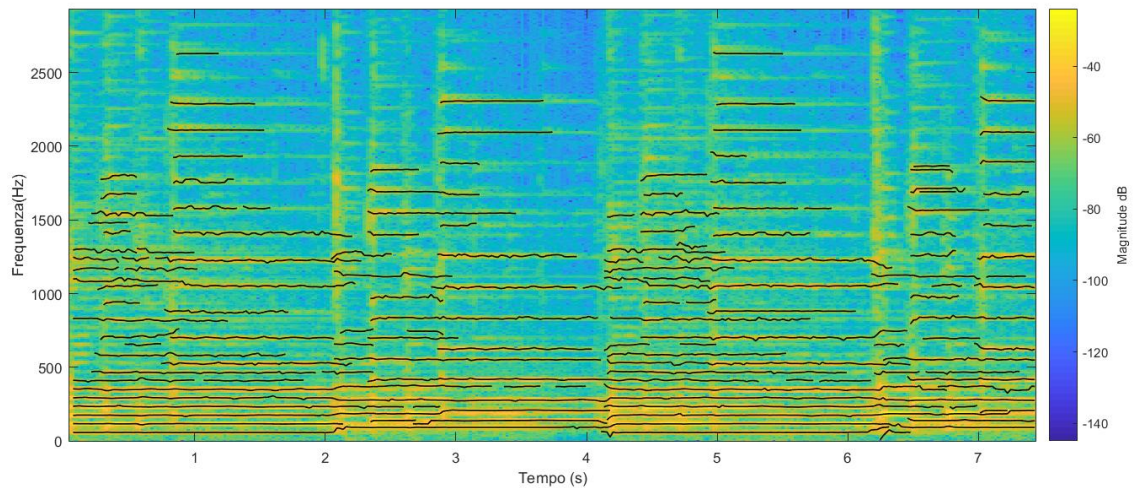


Figura 10: Spettrogramma con il tracciamento delle singole componenti

### 3.3 Analisi timbrica - Rilevamento delle formanti

Come si è accennato nel capitolo precedente (Sez. 2) 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 l'involuppo spettrale dei due segnali.

Per ottenere le formanti di un segnale, è necessario estrarre dal segnale complesso gli elementi frequenziali che caratterizzano il timbro. Per ottenere queste informazioni si utilizza un modello autoregressivo (AR) lineare *a soli poli* chiamato Linear Predictive Coding (LPC). Mediante LPC possiamo ottenere una serie di coefficienti  $a_k$  (poli di un filtro) e la risposta in frequenza del filtro risultante, corrispondente ad una approssimazione dello spettro del segnale. Tale approssimazione è definita come:

$$\hat{x}[k] = \sum_{k=1}^K a_k x(n-k)$$

dove  $K$  è l'ordine del filtro,  $a_k$  i coefficienti e  $x(n-k)$  i campioni precedenti. Questa formula esprime la combinazione lineare dei campioni precedenti, ovvero l'espressione di un filtro a risposta infinita (IIR - *Infinite Impulse Response*).

L'obiettivo principale della Linear Predictive Coding è l'estrazione dei coefficienti del filtro  $a_k$  della funzione di trasferimento definita come:

$$H(z) = \frac{G}{1 + a_1 z^{-1} + \dots + a_K z^{-K}}$$

Risolvendo il polinomio al denominatore, otteniamo le frequenze dei poli per ogni frame in cui sono stati estratti i coefficienti. Di seguito si illustrano le posizioni dei poli/formanti del segnale di un tratto vocale.

Infine si rappresenta il risultato ottenuto per un determinato frame del segnale, illustrando il contenuto spettrale, il suo involuppo e la posizione dei primi sei formanti. Si noti che l'involuppo ottenuto rappresenta fedelmente il profilo spettrale del segnale.



### CAPITOLO 3. MODELLO DI ANALISI

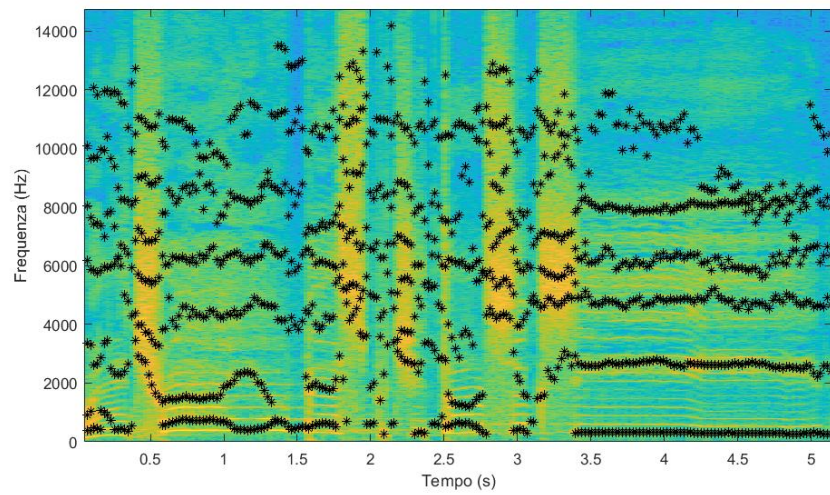


Figura 11: Poli del modello autoregressivo e il sonogramma del segnale

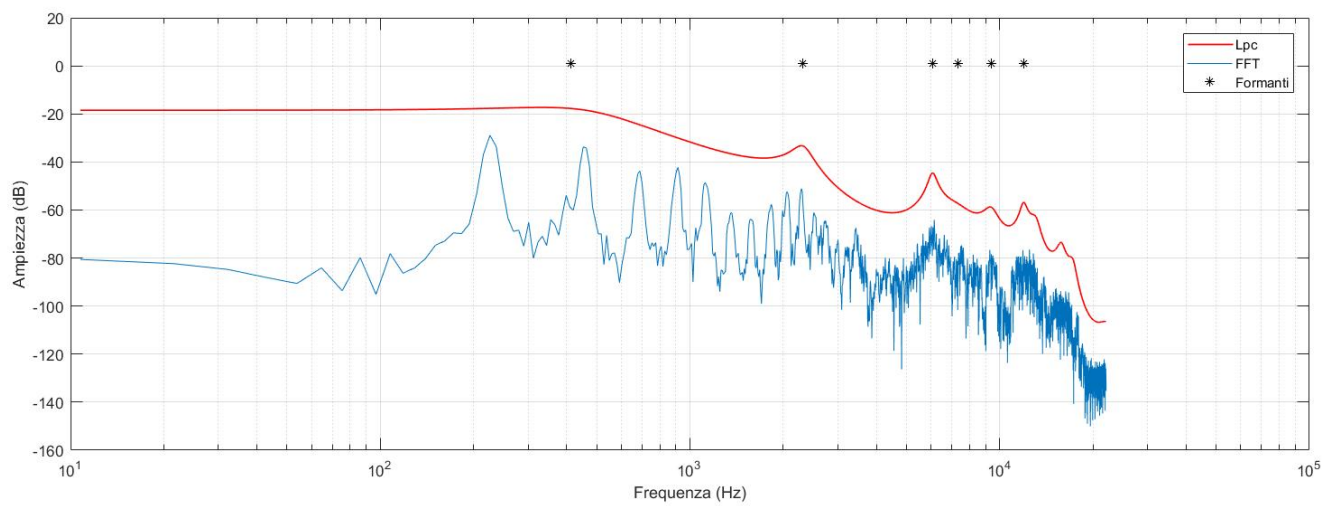


Figura 12: LPC (Inviluppo spettrale) e la posizione dei formanti in un determinato frame

### 3.4 Analisi attacco delle note - Onset Detection

Un suono, come anticipato nella sezione 2, ha un inizio e una fine. In questa sezione ci concentreremo in particolare di una caratteristica importante di un evento musicale, ovvero l'istante di una nota che caratterizza il transiente di attacco, chiamato mediante il termine inglese *onset*.

*Transiente*, *attacco* e *onset* sono termini che identificano lo stesso concetto ma non hanno lo stesso significato[9]. La Figura 13 illustra nel caso più semplice una nota isolata identificando i tre termini.

- L'*attacco* di una nota è l'intervallo di tempo nel quale il suono da intensità nulla arriva al suo massimo valore.
- I *transienti* sono brevi intervalli di tempo durante il quale il segnale si evolve rapidamente, ad esempio per segnali acustici, il transiente corrisponde all'intervallo di tempo in cui l'eccitazione viene applicata e poi smorzata/attenuata.
- L'*onset* è definito come un istante preciso nel tempo atto a identificare l'inizio di un transitorio di attacco.

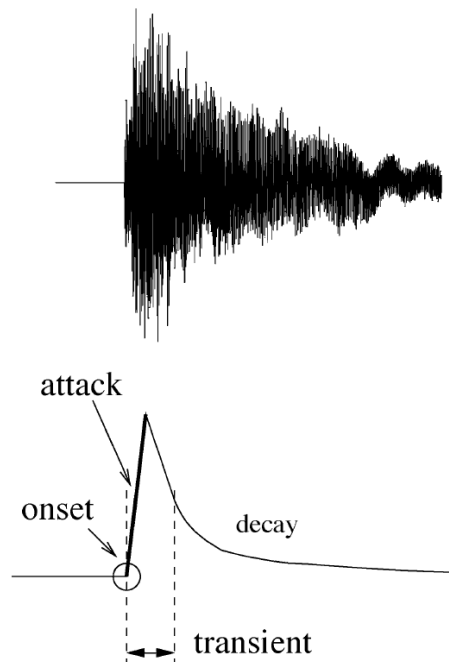


Figura 13: Attacco, transiente e onset

### CAPITOLO 3. MODELLO DI ANALISI

Maggior parte dei suoni, durante il transiente di attacco producono un'ampia dispersione di energia a banda larga. Questa situazione si incontra spesso durante la produzione di suoni come colpi di cassa e rullante di una batteria, l'eccitazione di una corda di chitarra mediante un plettro rigido oppure l'emissione vocale in particolar modo l'emissione di consonanti.

L'obiettivo è quello di identificare la collocazione temporale dell'*onset*. Viene implementato l'algoritmo di *Percussive Feature Detection*.

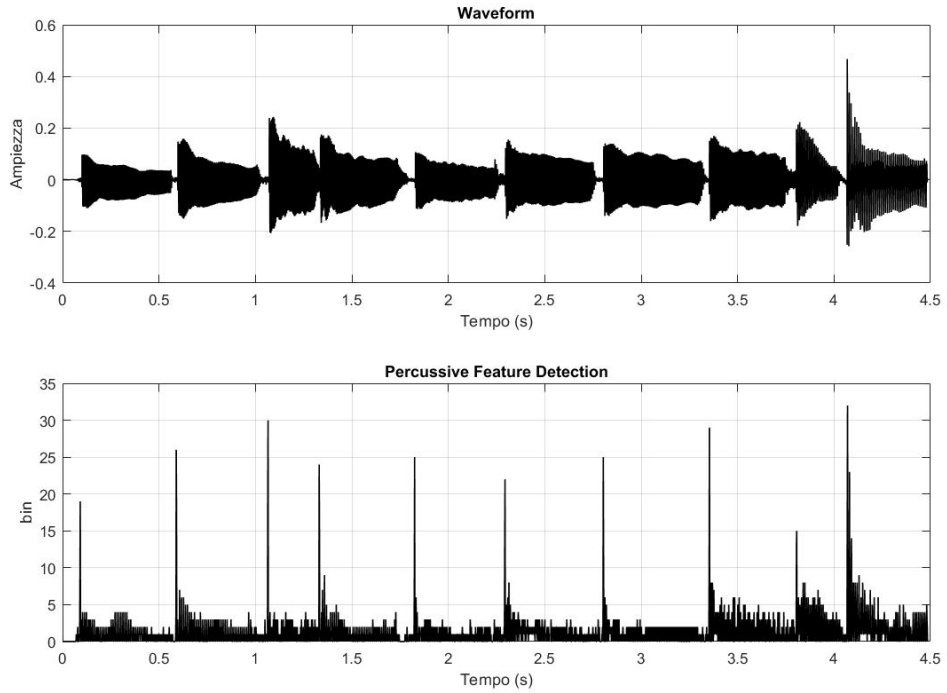


Figura 14: Percussive feature detection

Dopo aver effettuato la STFT sul nostro segnale finestrato mediante una finestra di analisi di lunghezza di 512 campioni<sup>2</sup>, con un *hopsiz*e di 64 campioni (risoluzione temporale di 1.5ms). Effettuiamo la differenza in *decibel* tra un frame temporale e quello successivo[8] definita come:

$$X'(k, m) = 20\log_{10} \frac{X(k, m-1)}{X(k, m)}$$

---

<sup>2</sup>Un valore basso permette di avere maggior risoluzione temporale ma con lo svantaggio di avere minor risoluzione spettrale

### CAPITOLO 3. MODELLO DI ANALISI

dove  $X(k, m)$  è la magnitudine al frame  $m$  alla frequenza del bin  $k$ .

Al fine di rilevare la presenza di un onset si definisce la seguente equazione :

$$Pe(m) = \sum_{k=1}^K \begin{cases} P(k, m) = 1 & \text{if } X'(k, m) > T \\ P(k, m) = 0 & \text{otherwise} \end{cases}$$

dove  $T$  è una soglia che indica l'incremento di energia espressa in dB necessaria per rilevare gli onset.  $Pe(m)$  corrisponde ad un contatore che esprime per ogni frame  $m$  il numero di bin che superano la soglia:  $P(k, m)$  contiene 1 se la condizione di soglia è soddisfatta, zero altrimenti. Si noti che l'energia reale non è significativa in  $Pe$ , si esprime solamente la quantità di dispersione a banda larga prodotta dalla percussione del suono. La figura seguente illustra l'approccio effettuato.

L'individuazione della posizione corrispondente all'inizio della nota ora è molto semplice da individuare, fissata una soglia statica per tutta la durata del segnale è possibile estrarre l'informazione temporale necessaria per ogni onset individuato.

#### 3.4.1 Validazione del modello

Ai fini della validazione del modello implementato si effettua un test mediante due impulsi generati in determinati istanti di tempo, calcolandone le differenze temporali in modo da convalidare la precisione dell'implementazione. Viene creata una finestra di un secondo dove al suo interno vengono generati due impulsi, il primo (canale *sinistro*) viene regolarmente generato a metà della finestra ( $t_1 = 0.5s$ ), mentre il secondo (canale *destro*) viene generato con un ritardo di 1024, 512, 64 campioni, rispettivamente 23.2ms, 11.6ms, 1.5ms in modo da valutare la precisione e la risoluzione temporale provocata dalla dimensione della finestra, un ritardo minore avrebbe come risultato una differenza pari a zero, dovuta alla risoluzione minima di 1.5ms. Nella figura 15 si illustra il risultato della *Percussive Feature Detection* e i risultati ottenuti effettuando la differenza tra le due rilevazioni nella tabella 1:

<b>Fig</b>	Ritardo PFD (ms)
a	23.2
b	11.6
c	1.5

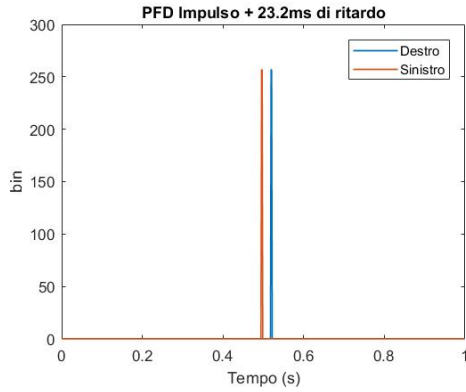
Tabella 1: Valori di ritardi in relazione della figura 15

Come si può notare, i valori risultanti del *Percussive feature detection* coincidono con i ritardi inseriti per la generazione degli impulsi. Questo modello riesce ad analizzare ritardi di 1.5ms, valori più piccoli darebbero come risultato zero perchè il PFD

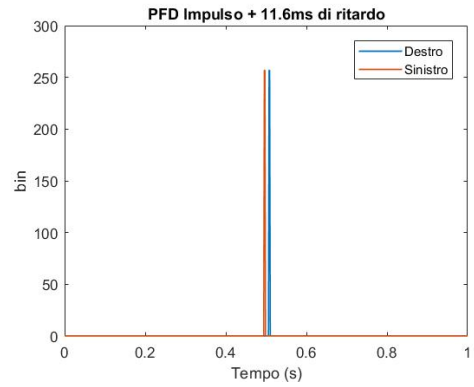
### CAPITOLO 3. MODELLO DI ANALISI

li analizzerebbe come all'interno dello stesso frame di analisi.

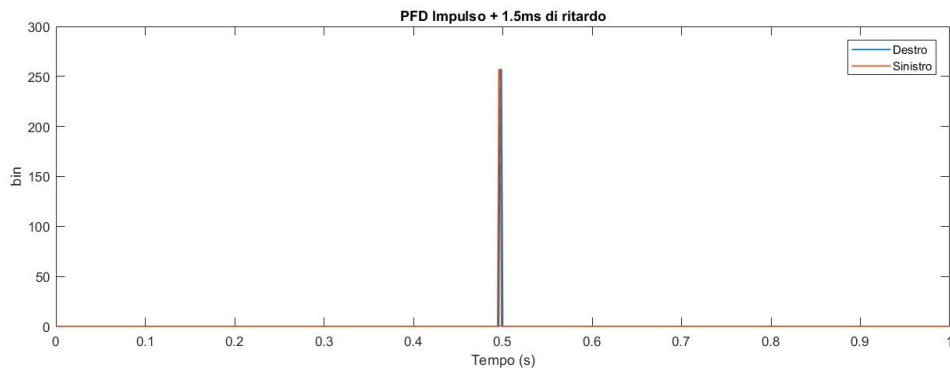
Una precisione di 1.5ms è una buona risoluzione temporale per analizzare questo tipo di ritardi.



(a) Impulsi ritardati di 23.2 ms



(b) Impulsi ritardati di 11.6 ms



(c) Impulsi ritardati di 1.5 ms

Figura 15: Coppie di impulsi ritardati

## 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] Goffredo Haus, Luca A. Ludovico, Giorgio Presti, *Automatic Annotation of Timbre Variation for Musical Instruments*, Proc. of the 13th International Symposium on CMMR, Matosinhos, Portugal, Sept. 25-28, 2017.
- [7] Amol R. Madane, Zalak Shah, Raina Shah, Sanket Thakur, *Speech Compression Using Linear Predictive Coding* Proceedings of the International Workshop on Machine Intelligence Research, 2009.
- [8] Dan Berry, Derry Fitzgerald, Eugene Coyle, Bob Lawlor, *Drum Source Separation using Percussive Feature Detection and Spectral Modulation*, ISSC, Dublin, 2005.
- [9] Juan Pablo Bello, Laurent Daudet, Samer Abdallah, Mark Brian Sandler, *a Tutorial on Onset Detection in Music Signal*, IEEE Transactions on Speech and Audio Processing, October, 2005.