

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:
Mirko Albanese
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.

Durante il processo di registrazione in una produzione musicale si è soliti incidere le chitarre e altri strumenti eseguendo 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.

Ringraziamenti

Desidero ricordare tutti coloro che mi hanno aiutato nella stesura della tesi con suggerimenti, critiche ed osservazioni. A tutti voi va la mia gratitudine, anche se spetta a me la responsabilità per ogni errore contenuto in questa tesi.

Anzitutto ringrazio il professor Goffredo Haus e il dottor Giorgio Presti rispettivamente Relatore e Co-relatore, senza la loro guida sapiente e la loro passione questa tesi non esisterebbe.

Proseguo ringraziando i ragazzi di IK Multimedia, in particolare Cedric Groyer, Alessio Ballotti e Nicolas Roussin per avermi aiutato attivamente nello svolgimento del progetto, supportandomi e sopportandomi in ogni momento. A Disi Melotti per avermi dato questa bellissima opportunità e a Lorella Cadignani per l'enorme pazienza nel processo burocratico rivelatosi infinito. Non ho avuto modo di conoscervi tutti, siete una grande squadra, continuate così!

A Giuseppe, Monia, Erika, la mia famiglia per il continuo supporto durante il mio percorso di studi e non solo.

Un grazie ai colleghi di università, in particolare a Lorenzo Romanelli, Davide Perini, Eleonora Dolif, Giulia Clerici, Roberto Alfieri e Davide Busacca, per non dimenticare i ragazzi con cui condivido la mia passione, la musica, in particolare a Giovanni, Pietro, Thomas e Matteo, per la loro carica emotiva e per la condivisione della musica che scrivo.

A Gloria Mambelli, per il continuo supporto durante lo sviluppo dell'intero progetto.

Indice

Prefazione

Ringraziamenti

1	Introduzione	1
2	Fondamenti Teorici	2
3	Modello di Analisi	3
4	Test effettuati e relativi risultati	4
5	Conclusioni e sviluppi futuri	14

Capitolo 1

Introduzione

La sensazione stereofonica che percepiamo utilizzando la tecnica della doppia incisione è dovuta dalle microvariazioni prodotte durante l'esecuzione, provocate dalla pressione del dito sulle corde, dall'enfasi e dalla concentrazione e/o umore del musicista, andando ad influire sulla singola performance. Ad esempio, un musicista che effettua una doppia incisione non applicherà la stessa pressione delle dita sulle corde, di conseguenza il contenuto frequenziale varierà di un valore impercettibile a tal punto da garantire una notevole sensazione di *chorus* naturale, gli attacchi delle note non saranno mai identici introducendo così ritardi di performance, infine non sarà nemmeno identica l'esecuzione generale dell'incisione.

Negli ultimi anni si è assistito ad un enorme crescita nello sviluppo di algoritmi e sistemi che, mediante tecniche di *Digital Signal Processing* (DSP), processassero segnali in tempo reale emulando e/o simulando comportamenti analogici di circuiti elettrici come outboard analogici (compressori, preamplificatori, equalizzatori), amplificatori per chitarra elettrica e cabinet di vario tipo, risposte all'impulso di varie stanze implementando riverberi/delay.

IK Multimedia mi propose di analizzare il comportamento percettivo della doppia incisione per un eventuale futuro prodotto basato sui risultati riprodotti dal modello qui descritto. L'obiettivo principale è quello di estrarre le differenze di un segnale stereofonico prodotto mediante la tecnica della doppia incisione.

Capitolo 2

Fondamenti Teorici

Capitolo 3

Modello di Analisi

Capitolo 4

Test effettuati e relativi risultati

In questa sezione verranno analizzate le singole coppie incise andando ad estrarre le differenze per ogni *feature*. Si descriverà il processo di estrazione dei risultati utilizzando come segnale di un riff di chitarra per poi illustrare a fine elaborato una tabella contenente le medie e le deviazioni standard delle differenze per ogni feature in determinati contesti di esecuzione (chitarra e voce).

Si prenda come riferimento un riff di chitarra registrato dal sottoscritto di durata di circa 4.5 secondi. Questo riff è formato da un totale di dieci note singole, estratte dalla scala blues nella tonalità di Sol minore. Nella figura 1 si nota che le due forme d'onda non coincidono perfettamente. Andando ad analizzare le differenze di dinamica dei valori ottenuti, si espongono i seguenti andamenti (Figura 2).

Il valore medio è illustrato nella tabella che segue:

Feature	Media (dB)	Dev std (dB)
Picco	2.3935	3.9272
RMS	2.2707	3.5766
Crest Factor	1.0211	1.2891

Tabella 1: Valori di media e deviazione standard di Picco, Rms e Crest Factor

Come si può notare nella figura 2 le variazioni sono considerevoli al momento in cui la nota viene suonata.

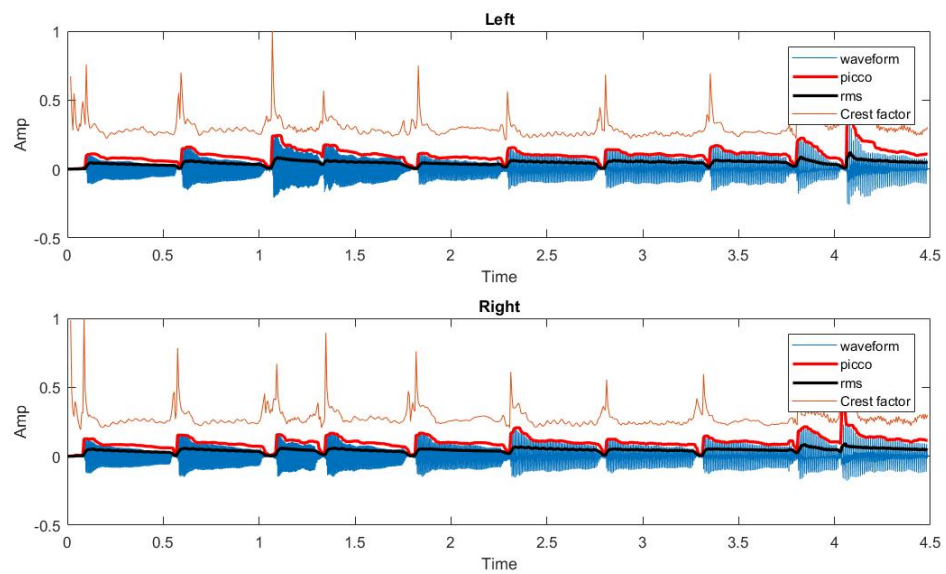


Figura 1: Forma d'onda di segnali con relativi andamenti di RMS Picco e Crest Factor

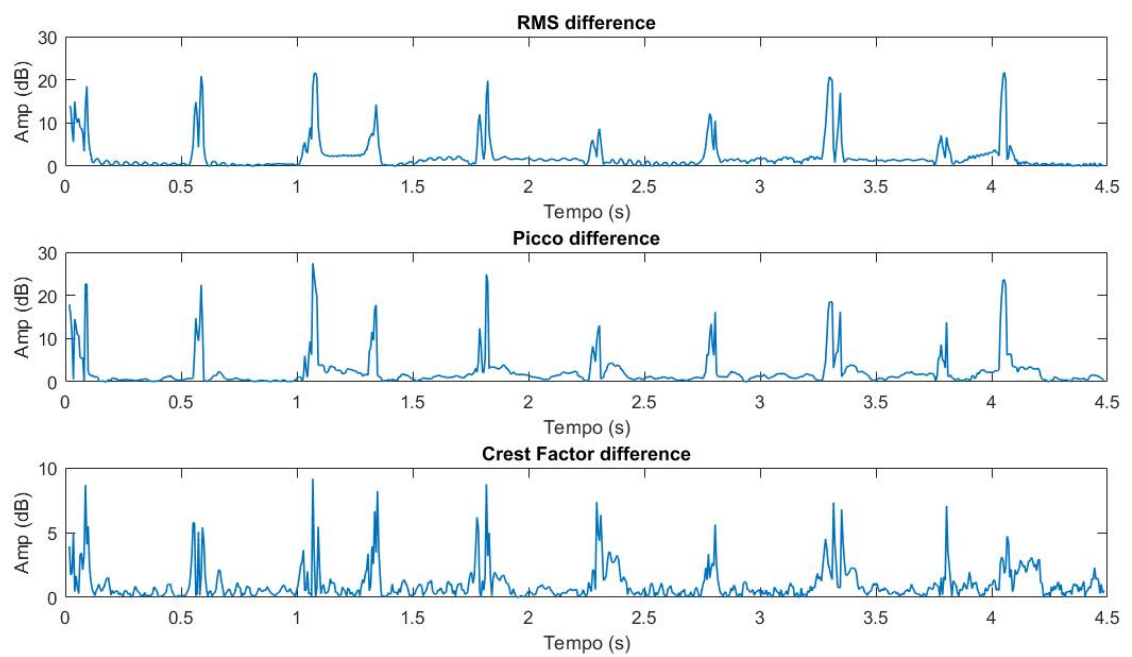


Figura 2: Differenze di RMS, Picco e Crest Factor

Per quanto riguarda il tracciamento delle armoniche viene preso in esame l'implementazione mediante differenza di fase. Di seguito si illustrano il tracciamento della fondamentale (rilevata mediante la nota tecnica "*Cepstrum*") dei due segnali e il tracciamento delle armoniche con le relative differenze.

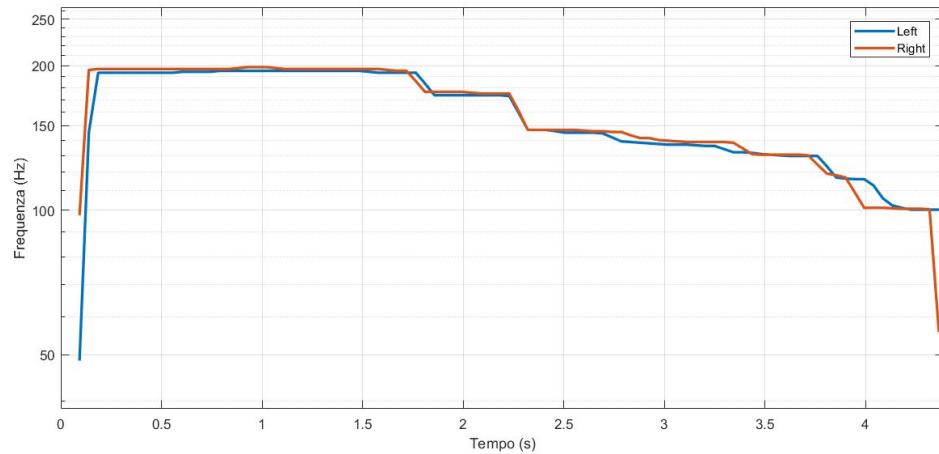


Figura 3: Rilevamento della fondamentale dei due segnali

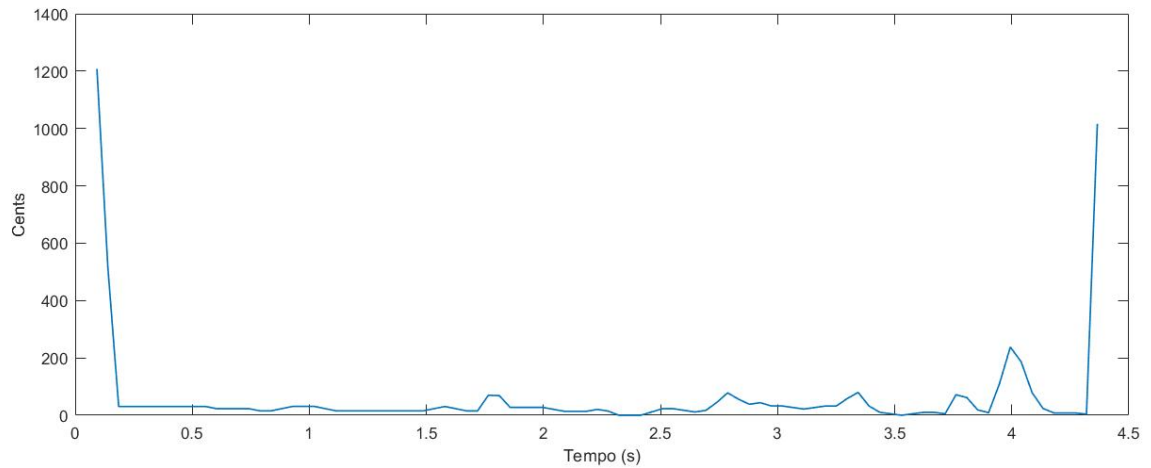


Figura 4: Variazione in Cent del rapporto tra le due fondamentali

I grafici 3 e 4 illustrano che la maggiore deviazione della fondamentale si ottiene quando avviene un cambio di nota.

Feature	Media (Cents)	Dev std (Cents)
F0	22.9987	168.8985

Tabella 2: Valori di media e deviazione standard delle variazioni della fondamentale espressi in cent

Ora andremo ad esaminare le variazioni delle armoniche tracciate dal modello di analisi. In un segnale complesso è facile avere un tracciamento disturbato da eventuali rumori prodotti da *ghost note* oppure da vibrazioni emesse dallo strumento stesso o dal musicista. È utile fissare una soglia per individuare i picchi da tracciare, data dal valore di SNR della finestra utilizzata sommata a uno scalare in dB aggiunto dall'utente in modo da cercare di evitare sidelobe.

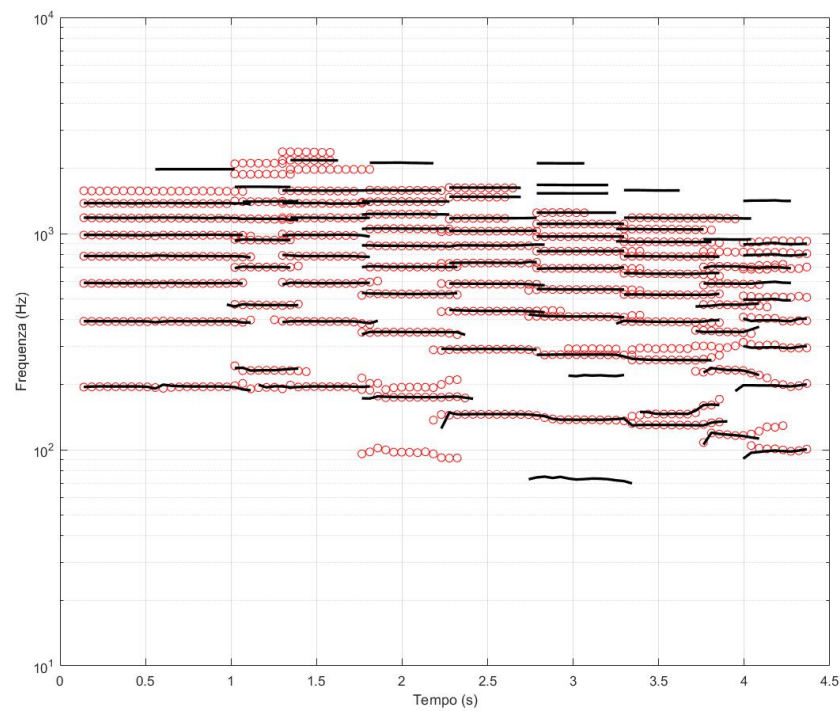


Figura 5: In rosso le armoniche del canale sinistro, in nero quelle del canale destro

In figura 5 si illustra in rosso le armoniche del segnale di sinistra, in nero quelle di destra.

Sono facilmente individuabili alcuni errori di esecuzione dove introducono artefatti dovuti molto probabilmente ad imprecisione di esecuzione oppure da una soglia di analisi troppo bassa, rilevando così eventuali sidelobe. Infatti è importante selezionare una corretta soglia di analisi in modo da estrarre le differenze per le armoniche che si corrispondono. In questo caso si noti che le prime sette armoniche si corrispondono, di conseguenza si effettuano le differenze di queste armoniche nel primo secondo di segnale e per ognuna si estrae media e deviazione standard.

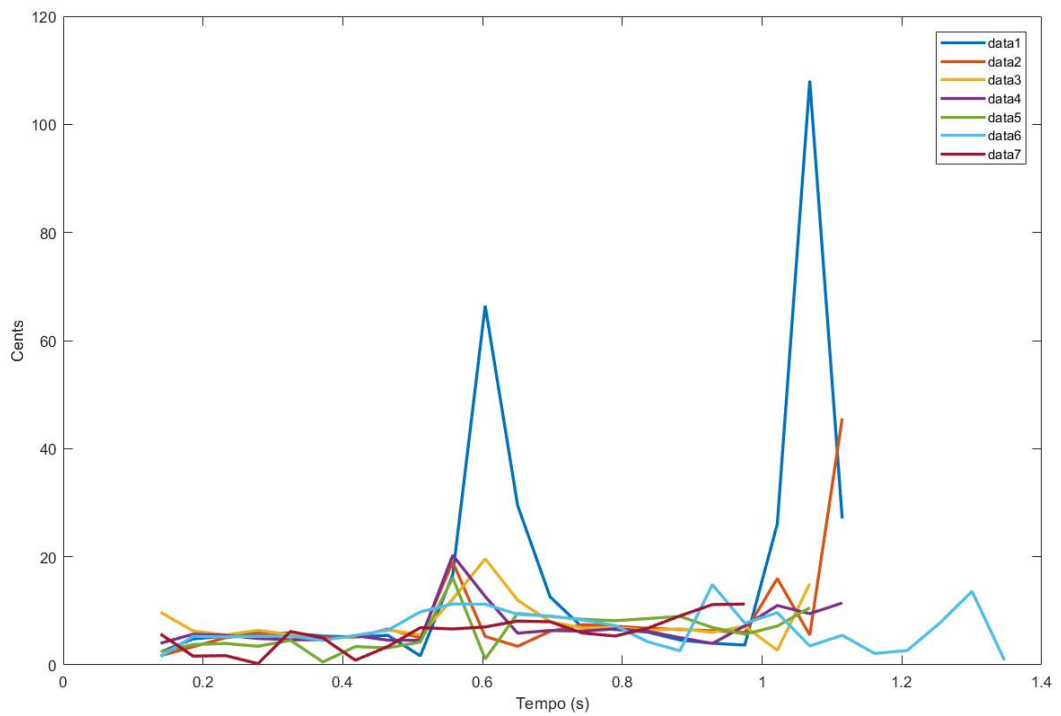


Figura 6: Variazione in cents delle prime sette armoniche del segnale

Si noti che vi sono dei picchi (*spyke*) in corrispondenza degli attacchi delle note. Ciò vuol dire che la massima deviazione la incontriamo in corrispondenza di ogni attacco. La minima deviazione, invece, la si incontra tra un attacco e quello successivo, dove la corda si assesta e continua la sua oscillazione.

Harm	Media (Cents)	Dev std (Cents)
1	16.3649	25.1818
2	8.3804	9.1425
3	1.7743	4.1552
4	7.5003	3.8366
5	5.8884	3.8492
6	6.9706	3.2110
7	5.0333	3.5714

Tabella 3: Valori di media e deviazione standard delle variazioni delle prime sette armoniche espressi in cent

Di conseguenza si calcolano la media e la deviazione standard complessiva delle variazioni delle singole armoniche in modo da ottenere un valore unico.

Media (Cents)	Dev std (Cents)
8.1769	3.7699

Tabella 4: Valori di media e deviazione standard complessivi delle variazioni delle armoniche

Per il rilevamento delle formanti viene calcolato il valore dell'ordine della Linear predictive coding in modo da evitare overfitting mediante la Minimum Description Length (MDL)[6].

Vengono inoltre limitati il numero delle formanti a sei ed estratte le differenze delle posizioni in frequenza delle formanti espressi in cents. Nella figura 7 si illustra per ogni incisione il contenuto frequenziale di un determinato frame, il suo inviluppo e la posizione delle formanti nello spettro, mentre nella tabella 6 si illustra il valore complessivo di deviazione.

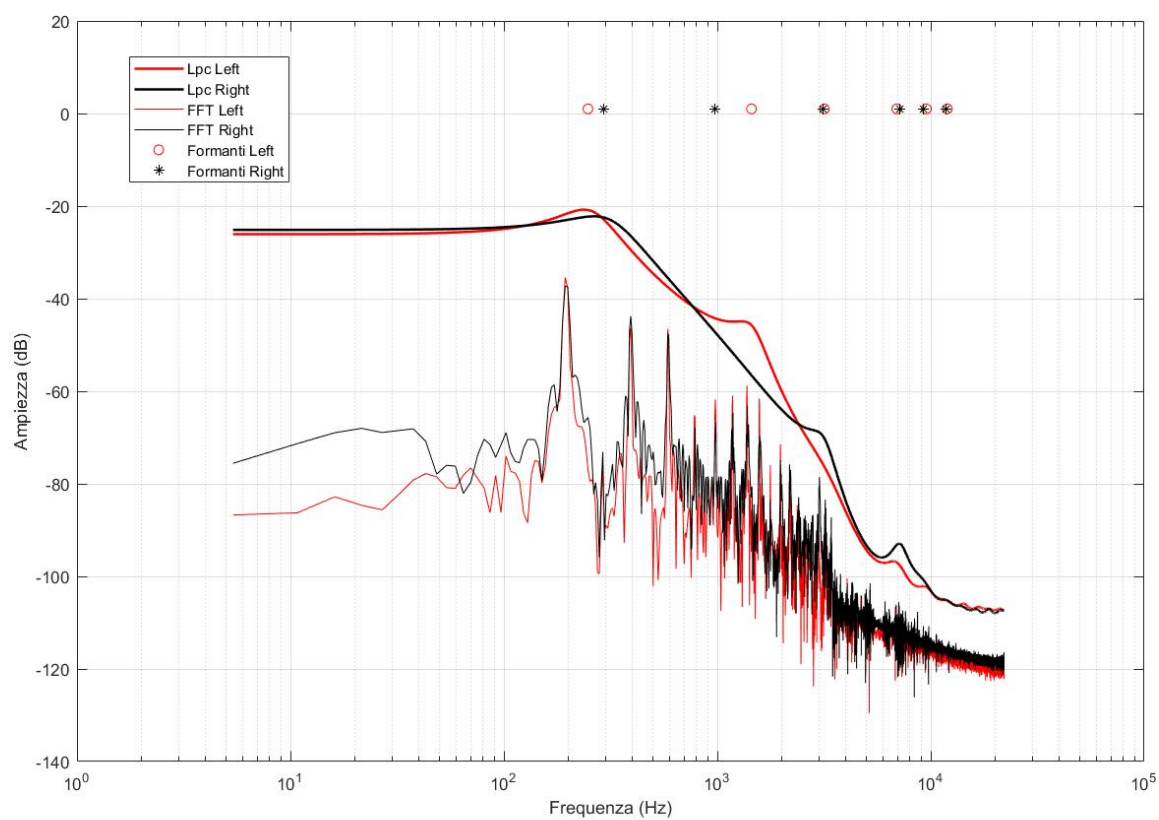


Figura 7: FFT, Formanti e LPC di entrambi i segnali

Formanti	Media (Cents)	Dev std (Cents)	Valore Medio (Hz)
1	162.8759	123.9264	429.92
2	190.7448	197.5683	1544.33
3	188.7199	192.6671	3179.64
4	146.3777	163.0557	5781.49
5	114.2583	142.1230	8062.51
6	93.3120	101.4009	10650.68

Tabella 5: Valori di media e deviazione standard delle variazioni delle prime sei formanti nel tempo e il valore medio dell'andamento di una sola formante

Media (Cents)	Dev std (Cents)
149.3814	39.5651

Tabella 6: Valori di media e deviazione standard complessivi delle variazioni delle formanti

Infine come ultima *feature* vengono estratte le differenze temporali dei due segnali. Di seguito si illustra il risultato ottenuto mediante la Percussive Feature Detection.

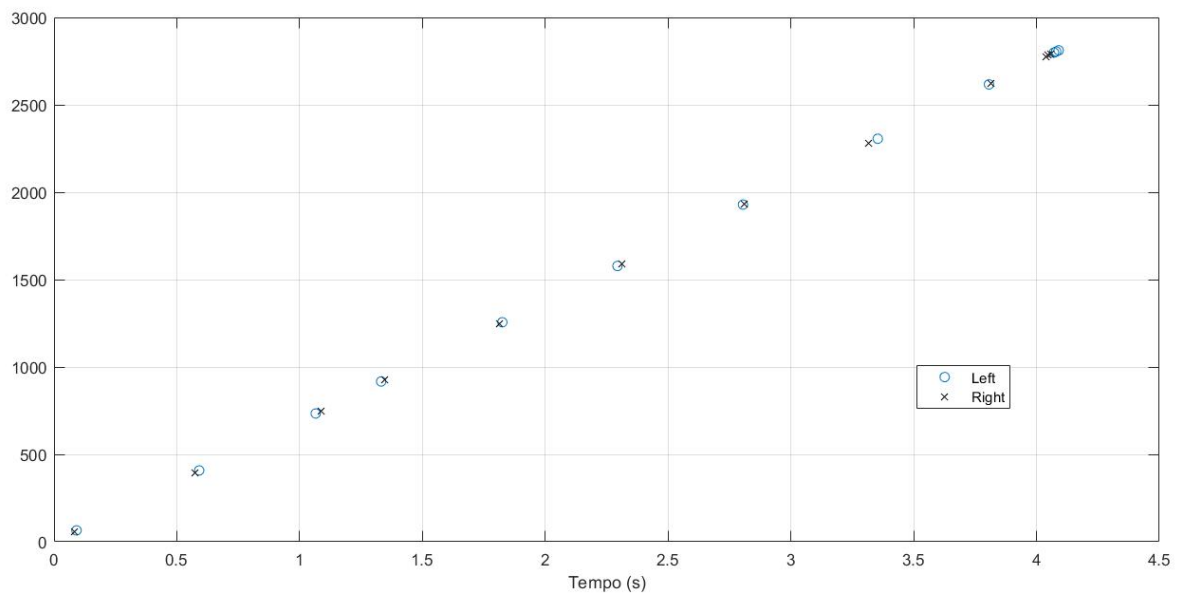


Figura 8: Posizione nel tempo degli onset

Si noti che ogni onset presenta con il suo doppiato un ritardo/anticipo dovuto all'errore umano nell'esecuzione. Ora avendo localizzato gli attacchi delle note, è possibile estrarre le differenze temporali.

Onset	Ritardi (ms)
1	8.7
2	18.9
3	21.8
4	16.0
5	10.2
6	16.0
7	5.8
8	36.3
9	7.3

Tabella 7: Valori dei singoli ritardi per ogni attacco delle note

Calcolando così la media e deviazione standard dei ritardi si estrae un valore unico.

Media (ms)	Dev std (ms)
15.6	9.5

Tabella 8: Valore di media e deviazione standard complessivo

Capitolo 5

Conclusioni e sviluppi futuri

TABELLA RIASSUNTIVA

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.