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

Mirko Albanese – 828180

Termine Tirocinio : 21 – 10 - 2017

In questo documento si descrive il lavoro svolto durante il mio tirocinio presso **IK Multimedia** (MO)

**Relatore**: Goffredo Haus;

**Correlatore Interno**: Giorgio Presti;

**Correlatore Esterno**: Disi Melotti (IK Multimedia)

**DESCRIZIONE DEL PROGETTO**

In questo progetto ho avuto il compito di effettuare test e analisi approfondite del comportamento stereofonico di una doppia incisione, o sovraincisione, di medesime parti musicali (sovraincisioni aperte in stereo). Questa tecnica è molto diffusa, effettuata in moltissime produzioni musicali per garantire una notevole apertura stereofonica. Di solito viene maggiormente utilizzata sulle chitarre acustiche ed elettriche che effettuano parti ritmiche e/o arpeggi, viene utilizzata inoltre sulle voci, elementi orchestrali di accompagnamento quali violoncello, tromboni, viola e molti altri strumenti.

L’obiettivo è di estrarre le differenze di feature fondamentali (frequenze, dinamica, attacchi delle note), in modo da consegnare ad IK Multimedia medie e deviazioni di dati statistici colti mediante tale progetto per un eventuale prodotto proposto dall’azienda.

**FASE DI TEST**

La prima fase è quella di creare dei tools in **Matlab** che mi permettono di effettuare test su coppie di segnali ideali quali impulsi e/o sinusoidi, creati sempre in matlab, dove conosco le differenze e le caratteristiche, in modo da convalidare l’ambiente di analisi testandolo con semplici segnali.

Dopo aver convalidato i tool in matlab, si effettuano prove con strumenti musicai veri e propri partendo da una chitarra acustica, poi da una chitarra elettrica distorta, fino ad arrivare a effettuare test sulle voci.

**FASE DI ANALISI**

Dopo aver acquisito le singole incisioni dai relativi strumenti e parti musicali, si effettua un analisi dettagliata delle singole coppie di incisioni, con i tool creati, in modo da estrarre le caratteristiche fondamentali di entrambi i segnali audio e osservare le variazioni di Timing, Pitch, Dinamica e qualsiasi altra informazione tra le due tracce audio in grado di fornire un’attenta panoramica delle variazioni.

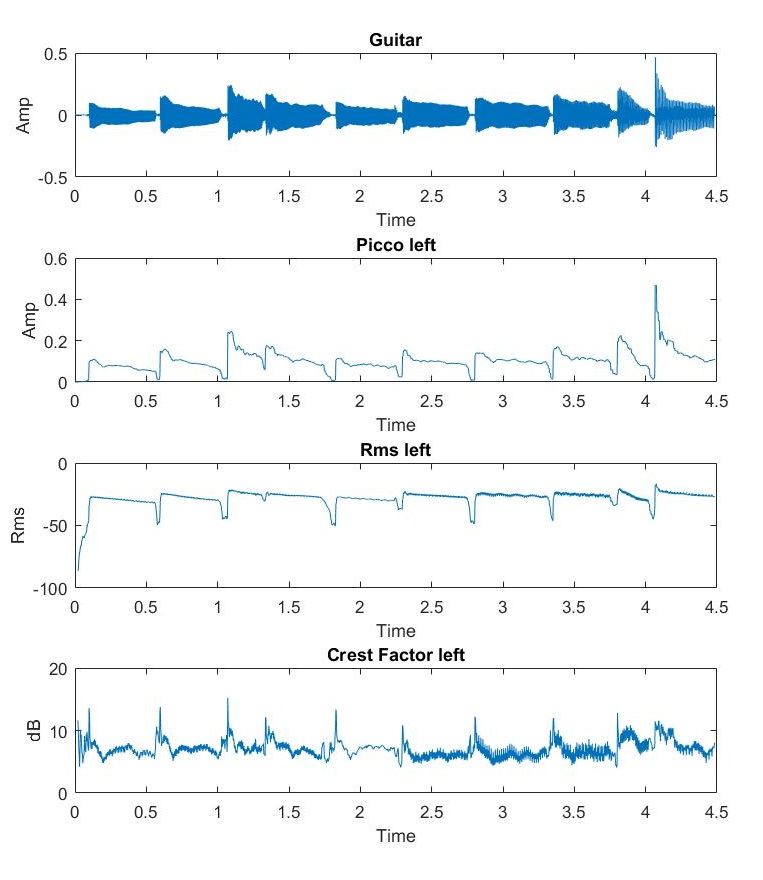
In merito sono state implementati modelli di analisi per quattro caratteristiche, quali sono:

* **Dinamica** : Energia del segnale e differenze in db estraendo media e deviazione standard;
* **Harmonic & pitch tracking :** Si definiscono le componenti frequenziali che caratterizzano il segnale in ingresso, si calcola il valore in frequenza mediante due metodi (Interpolazione parabolica e differenza di fase, mettendoli a confronto) e si tiene traccia delle singole componenti. Dopo di che si effettuano le differenze espresse in cent;
* **Analisi timbrica** : Questa caratteristica ci aiuta a definire il timbro di uno strumento e come questo strumento è stato suonato. È importante cogliere le differenze di performance tra una registrazione e la sua doppiata. Questa feature viene ottenuta mediante calcolo predittivo lineare utilizzando il modello Lpc, dove dai coefficienti in uscita riusciamo a ottenere l'inviluppo spettrale contenente le formanti del segnale. Questa tecnica porta ad un risultato dove pone al lettore un’idea generale di come lo strumento è stato suonato, ad esempio nelle chitarre elettriche la posizione del plettro o delle dita incide sul colore del suono emesso, ovvero il timbro;
* **Onset Detection** : Ultima caratteristica analizzata è il rilevamento dei cosiddetti *Onset*, ovvero l’inizio di un suono, come ad esempio l’eccitazione di una corda di un cordofono o un emissione vocale. Questa caratteristica è stata implementata mediante Percussive Feature Detection (PFD), esso è un modello che analizza il segnale nel dominio in frequenza e valuta l’andamento dell’energia nel tempo (*Spectral flux*).

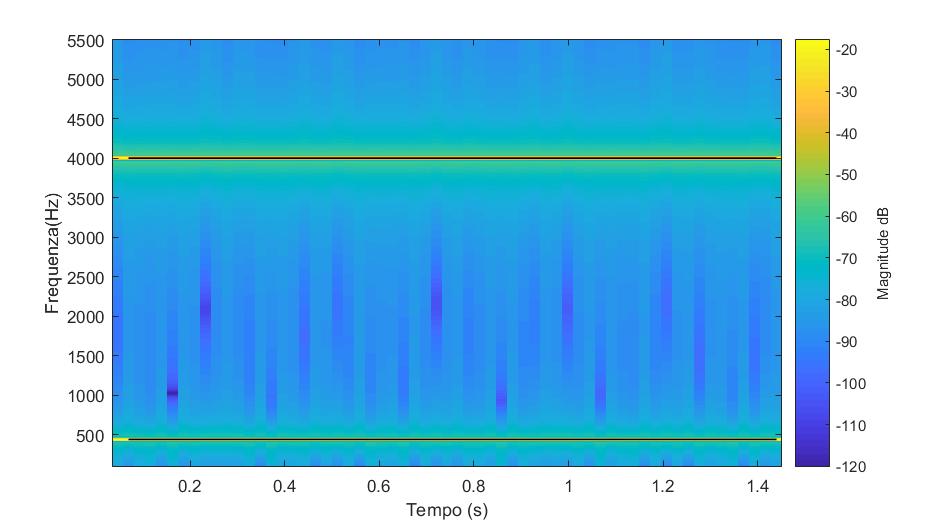
Di seguito si illustrano, per ogni caratteristica, grafici e le implementazioni effettuate:

**DINAMICA** : Per il calcolo della dinamica viene calcolato l’andamento del rms, l’andamento del picco e infine l’andamento del crest factor:

Di seguito si illustra un immagine di esempio.



* **HARMONIC TRACKING** : In questa sezione vi sono implementati due metodi di calcolo per tracciare le armoniche del suono in ingresso. Si sono implementati due metodi, **l’interpolazione parabolica** e la **differenza di fase**. Nella prima implementazione viene effettuata un’interpolazione per trovare la stima del valore corretto in frequenza, questa stima viene effettuata tra il picco rilevato e i due bin adiacenti laterali mediante la formula per ricavare il vertice di una parabola. La seconda implementazione, invece, molto più precisa va ad effettuare una differenza di fase del segnale finestrato tra un frame e quello successivo. Di seguito si illustrano immagini relativi alle due implementazioni calcolate su un tono puro a 440 Hz e 4000Hz e il suo tracciamento nello spettrogramma.



Valori del tracciamento per l’implementazione “**Interpalazione parabolica**”, dal 4° al 10° frame, tono puro a 440 Hz:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 440,4268 | 440,4280 | 442,4265 | 440,4282 | 440,4266 | 440,4277 | 440,4273 | 440,4270 | 440,4279 | 440,4265 |

Valori del tracciamento per l’implementazione “**Differenza di fase**”, dal 4° al 10° frame, tono puro a 440 Hz:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 440,0030 | 439,9978 | 440,0011 | 440,0002 | 440,9986 | 440,0025 | 439,9969 | 440,0031 | 439,9973 | 440,0018 |

Valori del tracciamento per l’implementazione “Interpalazione parabolica”, dal 4° al 10° frame, tono puro a 4000 Hz:

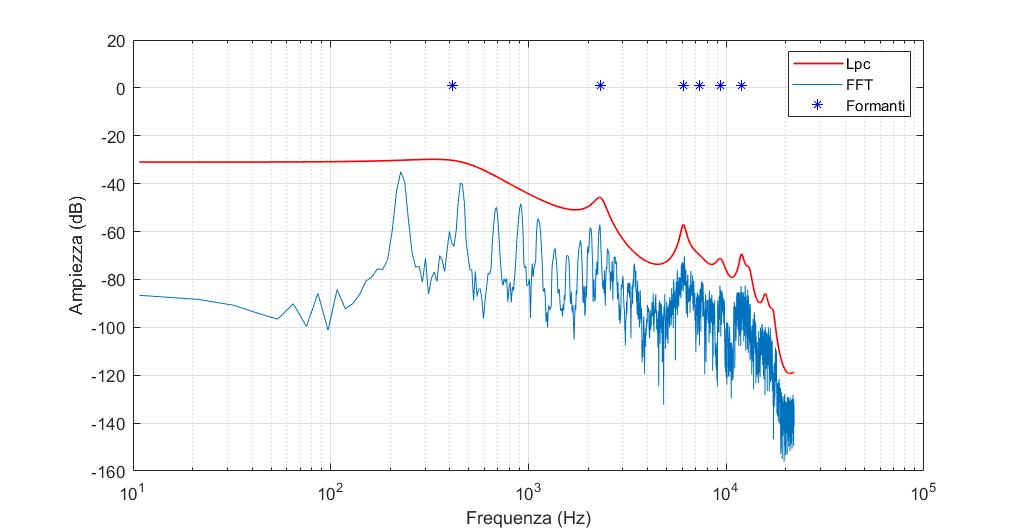
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 4000,1717 | 4000,1731 | 400,1711 | 4000,1695 | 4000,1714 | 4000,1731 | 4000,1715 | 4000,1695 | 4000,1709 | 4000,1731 |

Valori del tracciamento per l’implementazione “Differenza di fase”, dal 4° al 10° frame, tono puro a 4000 Hz:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 3999,9999 | 4000,0000 | 4000,0004 | 3999,9999 | 3999,9994 | 4000,0005 | 4000,0005 | 3999,9995 | 3999,9993 | 4000,0003 |

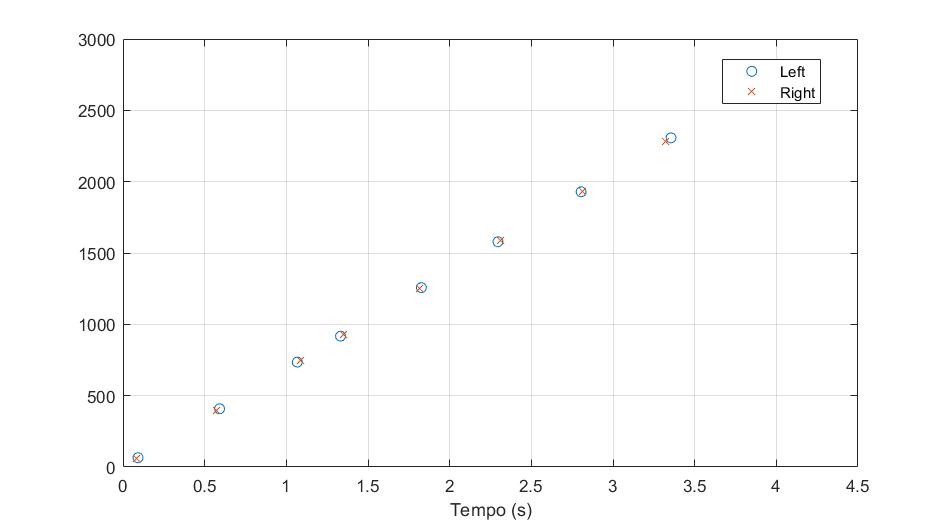
Si noti che l’implementazione mediante differenza di fase è molto più precisa.

* **ANALISI TIMBRICA [FORMANTI]** : L’analisi delle formanti (timbro) è stata ottenuta mediante calcolo predittivo lineare (LPC). Questa tecnica permette di estrarre l’inviluppo spettrale del segnale audio e il valore delle formanti mediante la risoluzione dei coefficienti polinomiali (poli).



La figura sopra riportata rappresenta, fft (linea blu), inviluppo spettrale (linea rossa) e formanti (asterischi) di un segnale vocale.

* **ONSET DETECTION** : In questa sezione vengono calcolati i ritardi degli attacchi delle note tra i due file audio. Viene utilizzata la tecnica “Percussive Feature Estraction”, ovvero viene analizzato mediante il modello *spectral flux* l’andamento dell’energia frequenziale del segnale nel tempo (frame per frame), andando a calcolare i ritardi espressi in millisecondi. Utilizzando valori di finestre di analisi e hopsize piccoli riusciamo ad avere una risoluzione temporale di 1.5ms. Di seguito si illustra il risultato dell’implementazione effettuata e i valori dei primi 8 onset.



La figura precedente illustra la posizione nel tempo degli onset del segnale audio stereo in ingresso (un riff di chitarra. I ritardi dei primi 8 onset sono illustrati di seguito. I seguenti valori sono espressi in millisecondi.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 7,3 | 18,9 | 20,4 | 16,0 | 10,2 | 16,0 | 7,3 | 36,4 |

Per ogni caratteristica vengono calcolate le differenze e viene espressa una variazione media per ogni file audio, così da stimare un andamento medio per un numero finito di test.

**BIBLIOGRAFIA**

* Julius O. Smith III, Xavier Sierra, *PARSHL: An Analysis/Synthesis Program*

*fon Non-Harmonic Sounds Based on a Sinusoidal Representation*, Stanford, University, California, 1987.

* J. L. Flanagan & M. Golden, *"Phase Vocoder"*, Bell System Technical Journal, vol.45 pp. 493-1509, 1966.
* Karin Dressler, *Sinusoidal Extraction using an E\_cient Implementation of a Multi-Resolution FFT*, Fraunhofer Institute for Digital Media Technology, Ilmenau, Germany, 2006.
* De La Cuadra, Patricio, Aaron Master & Craig Sapp. *Efficient pitch detection techniques for interactive music*. Proceedings of the 2001 International Computer Music Conference, 2001.
* Amol R. Madane, Zalak Shah, Raina Shah, Sanket Thakur, *Speech Compressione Using Linear PRedictive Coding*, Proceedings of the International Workshop on Machine Intelligence Research, 2009.
* Dan Berry, Derry Fitzgerald, Eugene Coyle, Bob Lawlor, *Drum Source SeparaTion using Percussve Feature Detection and Spectral Modulation*, ISSC, Dublin, 2005.