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

Ringraziamenti

Indice

Capitolo 1

Introduzione

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.

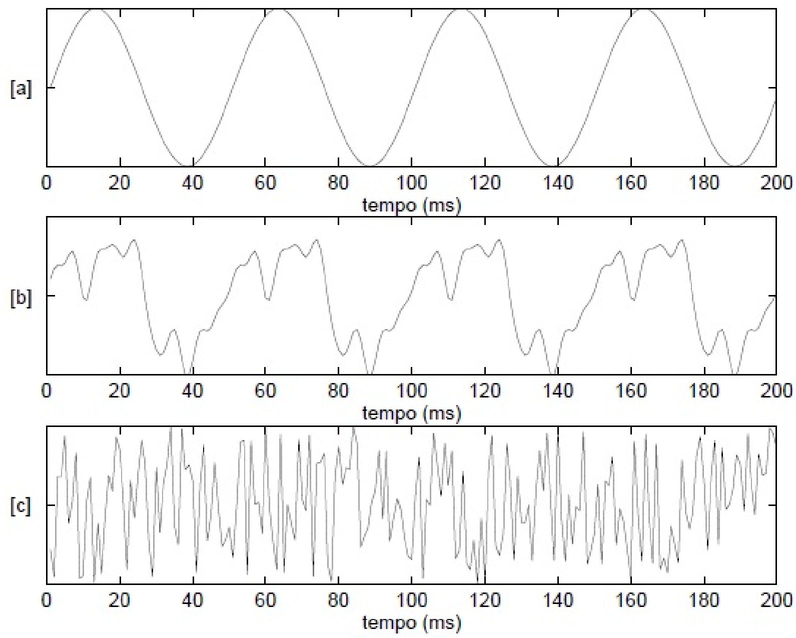

Figura 1: Andamento nel tempo di tre segnali rispettivamente: [a] sinusoidale, [b] periodico (somma di 16 sinusoidi), [c] aperiodico.

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.

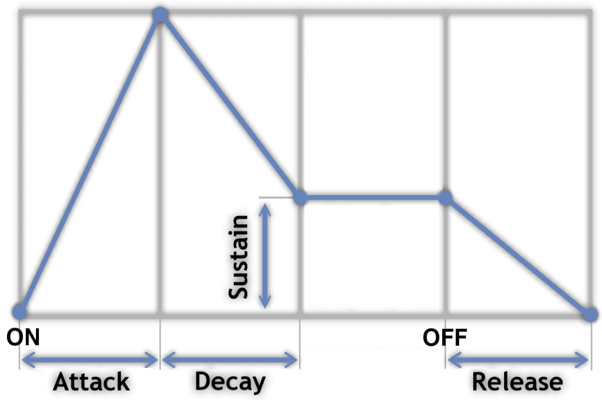

Figura 2: Inviluppo - ADSR

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, $2f$, $3f$, $4f$,.... 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 $I = 10^{-12} W/m^2$ al di sotto del quale non è più possibile percepire alcun rumore, mentre si chiama soglia del dolore il valore $I = 1 W/m^2$ 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 $p_0 = 0,00002 Pa$; 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
ppp	Libreria	40
pp	Interno auto silenziosa	50
p	Conversazione pacata	60
mp	Traffico	70
mf	Fabbrica	80
f	Metropolitana	90
ff	Discoteca	100
fff	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.

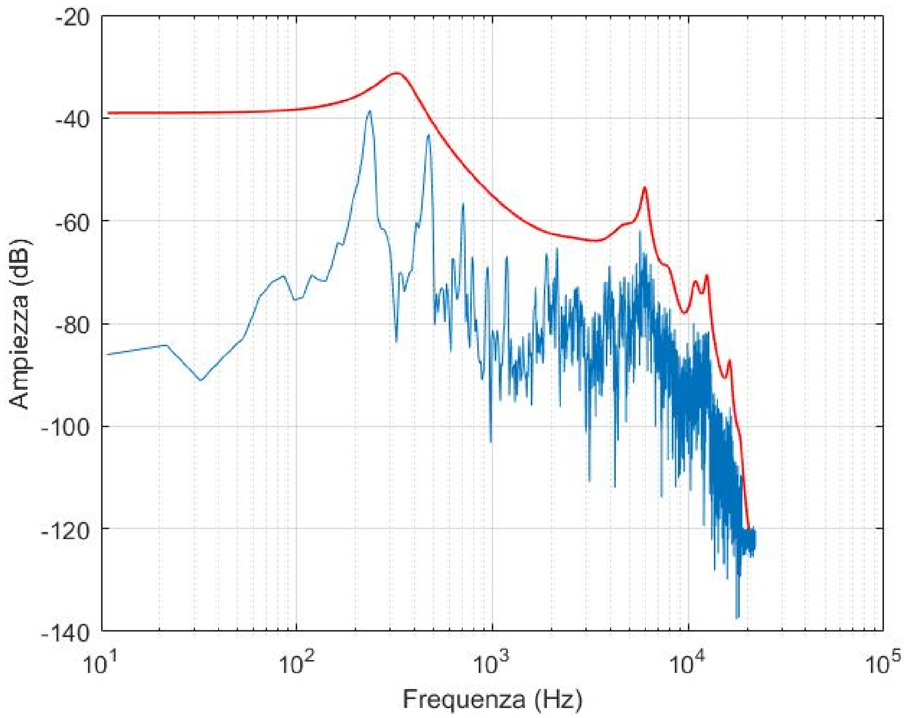

Figura 3: Esempio di formanti (linea continua rossa)

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

$$f(t) = f(t + kT), \qquad \forall t \in \Re$$

Il minimo valore del periodo $T > 0$ che soddisfa la definizione di periodicitá é chiamato periodo fondamentale ed è denotato da T0. 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[1].

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

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 decomporlo calcolandone i coefficienti complessi. Mediante l'identità di Eulero è possibile riscrivere queste equazioni utilizzando funzioni trigonometriche.

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

$$f(t) = \sum_{n=-\infty}^{\infty} c_n e^{inw_0 t} = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nw_0 t) + b_n \sin(nw_0 t)]$$

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 $T \to \infty$ 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:

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

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

[1] La trasformata di Fourier è una generalizzazione ai segnali non periodici

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

Sapendo che $w_0 = 2\pi v$, chiamiamo $v_n = nv$ la frequenza dell'ennesima 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:

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

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

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

$F(v)$ 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.

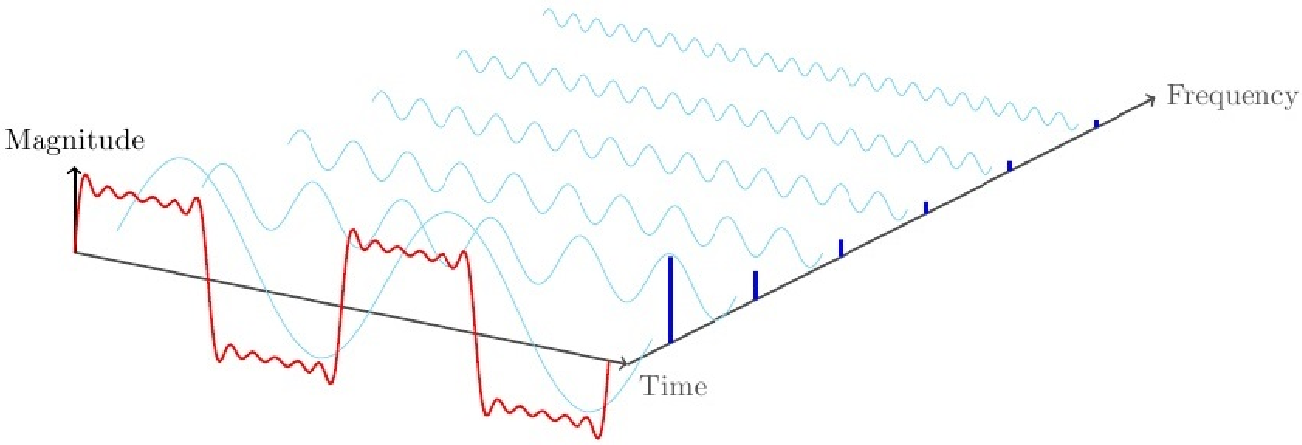

Figura 4: Dominio del tempo - frequenza

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

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

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* ($DTFT$); 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 $f_s = \frac{1}{\tau}$ = *frequenza di campionamento* dove F_0 corrisponde alla risoluzione frequenziale della DFT.

Sapendo che :

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

otteniamo:

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-i2\pi kF0n\tau} = \sum_{n=0}^{N-1} x(n)e^{-i2\pi k\frac{1}{N\tau}n\tau} = \sum_{n=0}^{N-1} x(n)e^{-ik\frac{2\pi}{N}n}$$

Quindi:

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

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 ϕ_k delle sinusoidi rispettivamente dal modulo e argomento di X_k:

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

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[2], 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'attenuazzione in dB del Sidelobes(lobi laterali).
Una finestra ideale dovrebbe avere un fattore di dispersione uguale a zero, una larghezza del Mainlobe molto bassa 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.

[2]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%

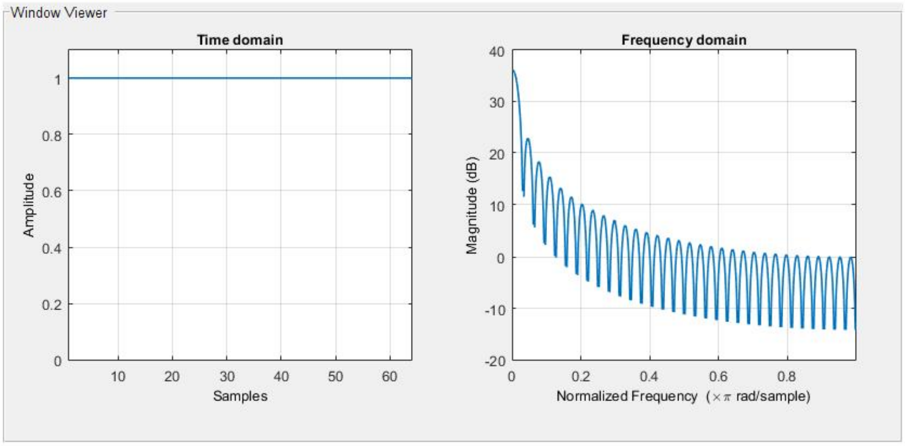

Figura 5: Finestra rettangolare

Nella figura 6 si può notare, invece che l'attenuazione dei sidelobes è molto più alta rispetto alla finestra rettangolare : intorno a -42 dB, questo comporterebbe una miglior lettura della STFT; e un fattore di dispersione del 0.03%[3]

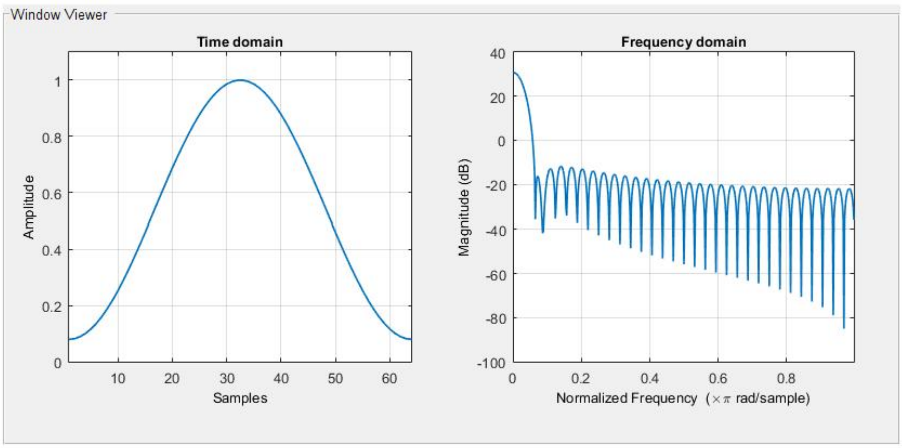

Figura 6: Finestra di Hamming

[3]I dati esposti sono stati elaborati tramite il tool in MATLAB "wvtool"

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 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.

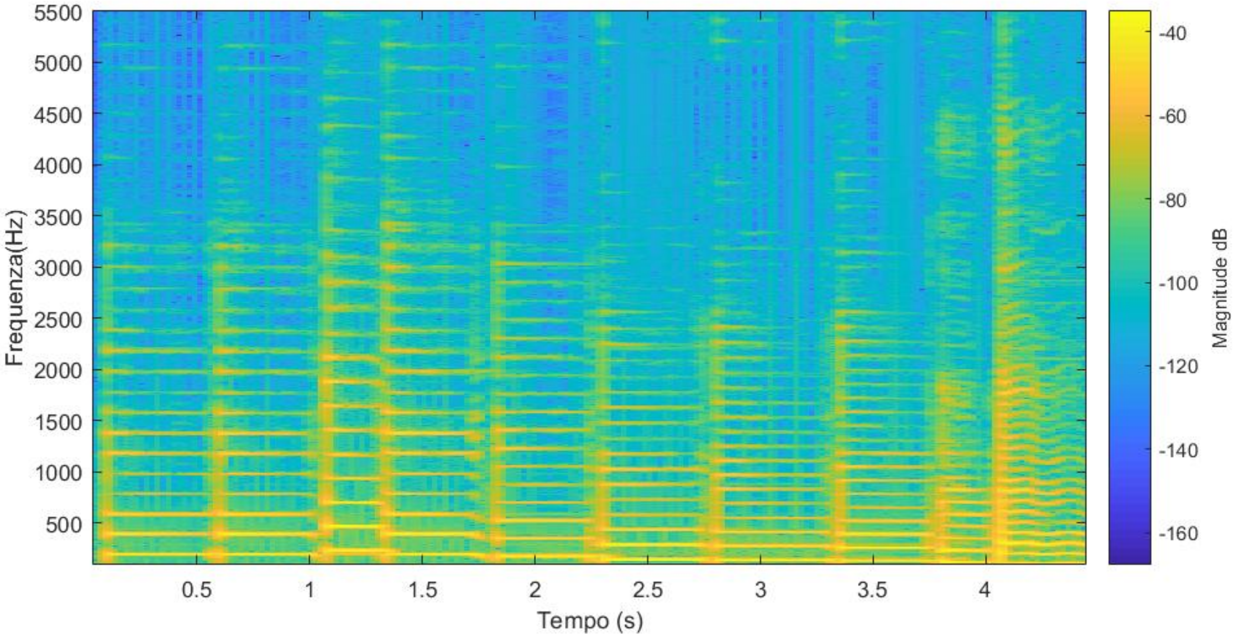

Figura 7: Sonogramma

Capitolo 3

Modello di Analisi

Capitolo 4

Test effettuati e relativi risultati

Bibliografia

[1] J. R. Pierce, *La Scienza del Suono*, Zanichelli, 1988.

[2] Julius O. Smith III, Xavier Sierra, *PARSHL: An Analysis/Synthesis Program fon 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 Compressione Using Linear PRedictive Coding* Proceedings of the International Workshop on Machine Intelligence Research, 2009.