



SAPIENZA  
UNIVERSITÀ DI ROMA

MASTER DI II° LIVELLO IN CALCOLO SCIENTIFICO

A.A. 2014/15

# Analisi di segnali acustici mediante Wavelets per il rilevamento di suoni inerenti spari in ambiente aperto.

## **Candidato**

Dr. Domenico Balestrieri

## **Tutor Interno**

Prof. Camillo Cammarota

## **Tutor Esterno**

Ing. Alessandro Rossi  
(INTECS S.p.A)

# INDICE

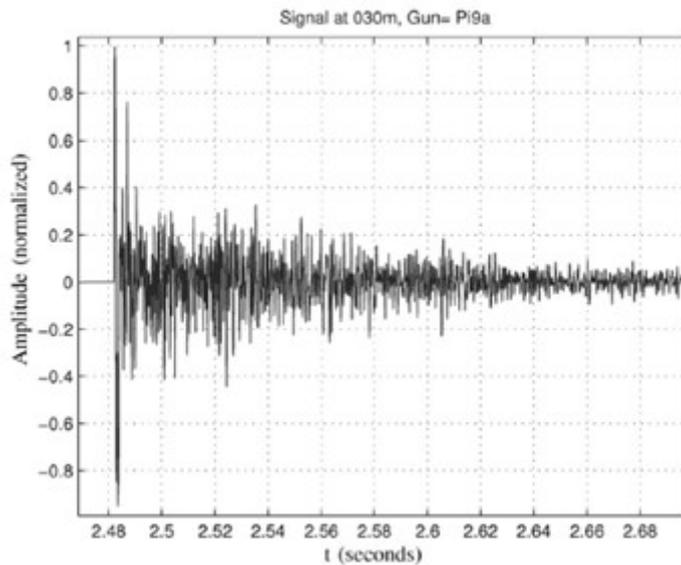
1. INTRODUZIONE.....	3
2. IL PROGETTO SMART-OPTIGRID .....	7
2.1 L'ANTENNA ACUSTICA .....	8
2.2 IL CLASSIFICATORE.....	9
3. PROBLEMATICHE NELL' ANALISI IN FREQUENZA DI SEGNALI NON STAZIONARI.	12
4. TRASFORMATA WAVELET.....	19
4.1 TRASFORMATA WAVELET CONTINUA (CWT).....	19
4.2 DISCRETE WAVELET TRANSFORM (DWT).....	24
4.3 WAVELET PER IL MULTISCALE EDGE DETECTION.....	28
4.4 MALLAT-ZHONG WAVELET (MZ-DWT) .....	31
5. DETECTION DI UN SEGNALE .....	36
6. ANALISI DI UN DETECTOR PER RILEVARE UNO SPARO IN UN SEGNALE ....	39
7. RISULTATI OTTENUTI .....	43
8. CONCLUSIONI E PROSPETTIVE FUTURE.....	50
9. BIBLIOGRAFIA .....	54

## **1. INTRODUZIONE**

La rilevazione, la classificazione e la localizzazione di spari sono di particolare interesse nei settori legati alla sicurezza pubblica, la sorveglianza, le forze dell'ordine, e l'esercito. C'è stata molta ricerca negli anni per quanto riguarda la teoria sulle armi da fuoco ( vedi [6],[7],[8],[9]), nonché molte implementazioni hardware/software di metodi computazionalmente efficienti di analisi ed elaborazione di segnali (vedi [10][11][12][13][14][15][16] solo per fare qualche esempio). Queste soluzioni per lo più utilizzano algoritmi complessi, tipicamente la short-time Fourier Transform e trasformate wavelet nonché le catene di Markov Hidden e i metodi di massima verosimiglianza etc. e sono molto efficienti nell'individuare, classificare e localizzare colpi da diverse armi da fuoco. Di contro sono piuttosto costosi in termini di potenza richiesta dovuta alle loro esigenze di calcolo rendendo il deploy sul campo di un tale sistema di classificazione piuttosto ingombrante soprattutto a causa dei requisiti di energia e, come abbiamo detto, soprattutto di potenza di calcolo. Un caso particolare di interesse è la struttura di una rete di sorveglianza contro la caccia illegale in riserve forestali tropicali. In tale ambiente, le reti di sensori a basso consumo possono fornire una soluzione praticabile, viste le ampie aree da proteggere e la quasi impossibilità di fornire sensori di maggiore potenza e con autonomia prolungata.

Per quanto riguarda la complessità che vi è dietro ai problemi di rilevazione e classificazione di un arma da fuoco, Maher (in [6],[7]) fornisce una spiegazione molto approfondita della fisica di un colpo di pistola. Il suono di uno sparo è prodotto da due fenomeni: in primo luogo dall'esplosione prodotta dai gas a rapida espansione della carica esplosiva che viene utilizzata per spingere il proiettile dalla canna della pistola. Questo shock acustico dura 3-5 ms e si propaga attraverso l'aria alla velocità del suono. In secondo luogo, se il proiettile viaggia a velocità supersonica, provoca un'onda d'urto acustica che si propaga lontano dalla traiettoria della pallottola. L'onda d'urto si espande come un cono dietro il proiettile, con il fronte d'onda che si propaga verso l'esterno alla velocità del suono.

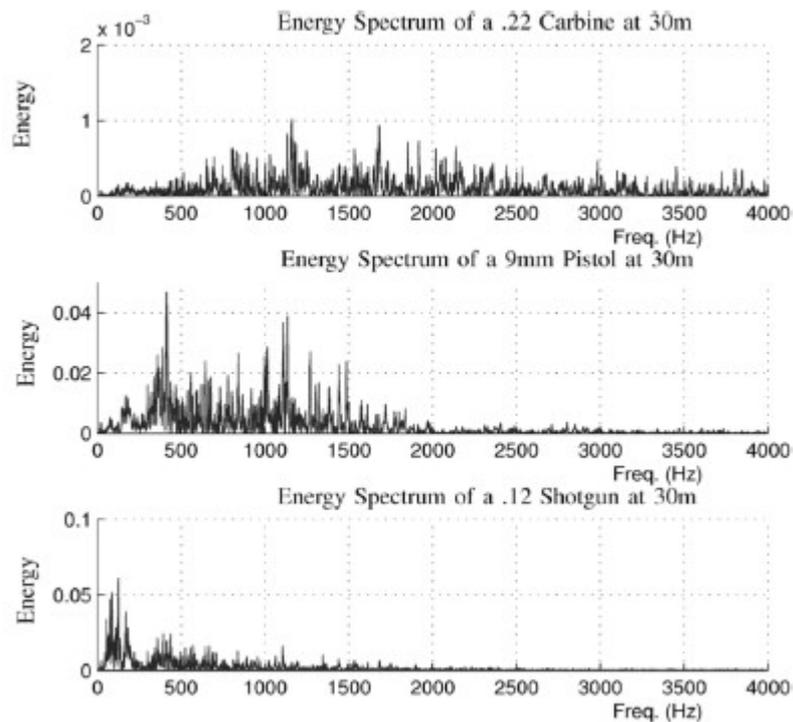
Una firma acustica tipica di uno sparo è mostrato in Figura 1.



**Figura 1: firma sonora tipica di un colpo di pistola da 9mm a 30 metri dal microfono.**

Le caratteristiche del suono di qualsiasi arma da fuoco sono determinate da fattori quali il calibro del proiettile e la canna, la lunghezza di quest'ultima, e le proprietà chimiche del propellente. Inoltre, essendo un segnale impulsivo quasi perfetto, qualsiasi misura particolare delle caratteristiche spettrali o impulsive di un particolare colpo di pistola probabilmente darà ulteriori informazioni sull'acustica dell'ambiente circostante (cioè, la risposta all'impulso nell'ambiente circostante), piuttosto che delle caratteristiche del proiettile e dell'arma ([6],[7]), che, a loro volta, dipendono da un insieme multiplo di fattori come la temperatura, la velocità del vento, densità del fogliame, umidità dell'aria, e le caratteristiche del suolo ([17]). I tentativi di rilevazione dell'onda d'urto generata dallo sparo diventano difficili nel momento in cui l'onda perde rapidamente la sua forma a causa della dispersione non lineare o scompare del tutto una volta che la velocità del proiettile scende al di sotto della velocità supersonica o colpisce un ostacolo.

D'altro canto, guardando lo spettro di potenza di tre particolari spari, si ha anche un'idea delle differenze tra armi situate alla stessa distanza (Fig. 2), che scoraggia semplicemente l'uso di un semplice metodo di filtraggio per l'attività di rilevazione.



**Figura 2: Esempi di spettri di potenza per una carabina .22, una pistola 9mm, e .12 Fucile registrato a 30 m.**

In questa tesina è stato studiato un approccio per il riconoscimento di forme d'onda relative a spari da arma da fuoco, basato sull'uso della trasformata wavelet. Questo studio è stato effettuato nell'ambito di un progetto di ricerca denominato SMART-OPTIGRID, sviluppato dalla INTECS S.p.A (società per cui l'autore e il relatore esterno lavorano) in collaborazione con l'Università "La Sapienza" (Dipartimento di Ingegneria dell'Informazione Elettronica e Telecomunicazioni - DIET) per conto del Ministero della Difesa, con lo scopo di realizzare un semplice rivelatore (detector) usato come una prima fase di elaborazione di un segnale captato da un'antenna acustica, che attiverebbe poi una classificazione successiva.

La tesina è così strutturata:

- Nel capitolo 2 verrà data una breve descrizione del progetto SMART-OPTIGRID e come è inserito il detector realizzato all'interno del sistema.
- Nel capitolo 3 verrà introdotto l'approccio classico all'analisi dei segnali non stazionari (basato sulle short-time fourier transform) e verranno evidenziate quelle problematiche per risolvere le quali sono stati fatti studi che hanno portato poi all'invenzione delle trasformate wavelet

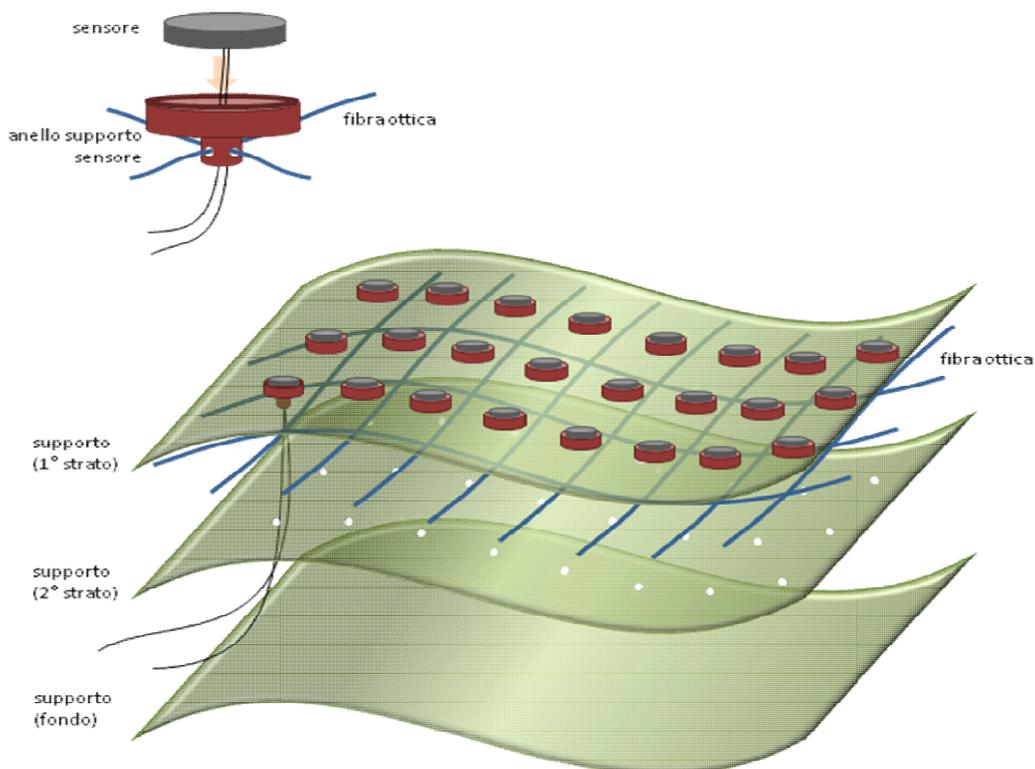
- Nel capitolo 4 verrà descritto il principio di funzionamento delle trasformate wavelet ed in particolare la trasformata wavelet di Mallat-Zhong (MZ-DWT) [2] che è alla base del detector realizzato per il riconoscimento degli spari. Nell'ambito di questa tesina, pur mantenendo un certo formalismo nella loro formulazione matematica, la descrizione sarà focalizzata soprattutto sugli aspetti intuitivi ed applicativi.
- Nel capitolo 5 verrà descritto invece il principio di funzionamento generale del detection di un segnale
- Nel capitolo 6 verrà descritto il detector realizzato nell'ambito del lavoro svolto come attività di stage per il Master. Il detector realizzato (Xie, Bao, Li, Zhang [5]) ha lo scopo di riconoscere in generale delle onde d'urto in un segnale. La tecnica utilizzata è basata sull'uso della trasformata wavelet. Tale detector, come si è precedentemente accennato, rappresenta la prima fase di elaborazione di un segnale captato dall'antenna acustica, che attiverrebbe poi la sua classificazione successiva. Lo scopo della sua introduzione nell'ambito del progetto OPTIGRID è stato quello di ridurre l'errore di classificazione dei segnali in input, che in questo modo verrebbero sottoposti, grazie all'uso del detector, ad una pre-classificazione grossolana (sparo /non-sparo) prima della fase di classificazione vera e propria.
- Nel capitolo 7 verrà descritto, a scopo esemplificativo, un caso di utilizzo del detector realizzato per riconoscere la presenza di spari dovuti ad un attività di caccia in una data area monitorata.
- Nel capitolo 8, infine, vengono tratte le conclusioni di questo lavoro e dati alcuni spunti per uno sviluppo futuro del sistema realizzato.

## 2. IL PROGETTO SMART-OPTIGRID

Il progetto OPTIGRID riguarda la realizzazione di antenna acustica 3D basata su phased array flessibile di microfoni (Flexible Acoustic Phased Array), in grado di mantenere invariato il diagramma di ascolto ad alta direttività (rilevazione della direzione in azimut ed elevazione della sorgente acustica tramite deflessione elettronica del fascio) nonostante la flessione della membrana.

Tale prestazione viene raggiunta tramite l'impiego di una griglia di fibre ottiche (Fiber Bragg Grating) in grado di consentire il calcolo in tempo reale delle posizioni relative degli elementi acustici che compongono l'antenna durante le deformazioni della membrana. Il supporto dell'antenna può quindi adattarsi a superfici di installazioni non planari come superfici curve di veicoli e superfici che variano in modo dinamico la propria forma.

Inoltre la cooperazione tra più phased array tramite collegamento in rete può consentire l'individuazione della posizione esatta (range, azimut ed elevazione) delle sorgenti acustiche tramite l'inseguimento delle stesse tramite più fasci di ascolto (triangolazione).



**Figura 3: Schema generale dell'antenna di microfoni su supporto flessibile**

## 2.1 L'ANTENNA ACUSTICA

L'utilizzo di array di microfoni è stato da tempo considerato per numerose applicazioni tra cui la separazione e l'amplificazione di sorgenti acustiche in ambienti rumorosi, l'identificazione e l'inseguimento di firme acustiche predefinite, la diagnostica ecografica ad ultrasuoni, la ricerca subacquea e l'analisi meteorologica. L'utilizzo di array di microfoni è stato anche esplorato per applicazioni di sorveglianza in ambienti outdoor.

Un array di microfoni (Figura 3) è costituito da un insieme di microfoni distribuiti nello spazio; in linea di principio è possibile ipotizzare distribuzioni di tipo lineare, planare o tridimensionale in accordo a svariate geometrie. Come ben noto, i microfoni agiscono da trasduttori convertendo il segnale di pressione in un segnale elettrico.

Utilizzando congiuntamente i segnali acquisiti dai diversi microfoni e combinandoli in maniera opportuna è possibile sintetizzare un fascio di ascolto con caratteristiche di direttività estremamente migliorate rispetto al singolo microfono; ciò determina un significativo miglioramento della qualità del segnale proveniente da una determinata direzione che si traduce in un incremento della capacità di rivelare, localizzare e riconoscere sorgenti acustiche che emettano da tale direzione.

Il principio di funzionamento di un array di microfoni si basa sulle caratteristiche di propagazione del suono. Infatti, il segnale generato da una sorgente acustica giunge ai microfoni dell'array con un ritardo che è funzione della velocità di propagazione del suono e della distanza del singolo sensore microfonico dalla sorgente. L'utilizzo congiunto dei segnali acquisiti dai diversi sensori mira ad incrementare il rapporto segnale-rumore per il suono che ha origine da una direzione desiderata. A tale scopo, l'algoritmo di beamforming (delay-and-sum beamforming,[18][19][20]) somma i segnali ricevuti dai diversi microfoni, opportunamente ritardati, in modo da combinare coerentemente i contributi provenienti da una specifica direzione; ciò equivale a formare un fascio di ascolto con elevata direttività (pencil beam) puntato in tale direzione. Modificando i ritardi da applicarsi ai segnali ricevuti dai diversi microfoni, è possibile variare la direzione di puntamento del fascio di ascolto senza alcun movimento meccanico dell'array, il che permette di esplorare ampi settori angolari in tempi rapidi.

## 2.2 IL CLASSIFICATORE

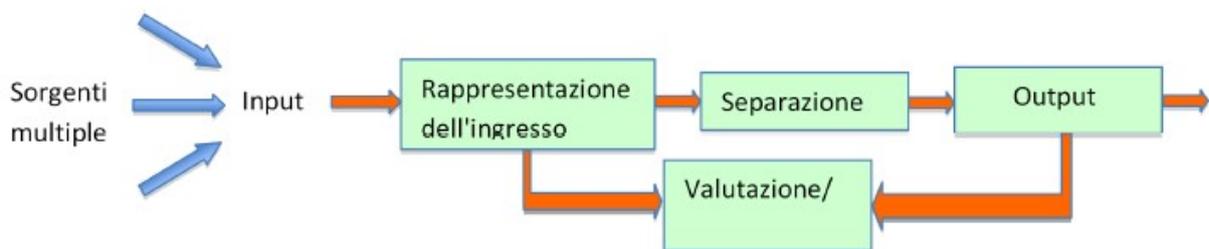
Il segnale proveniente dall' antenna acustica, attiva poi una classificazione successiva, al fine di riconoscere la "firma" sonora del segnale captato dall'antenna stessa.

Il problema del riconoscimento di una "firma" sonora richiede in generale che il segnale venga prodotto in condizioni di relativo isolamento. Per questo motivo la separazione del segnale desiderato da altri segnali ad esso contemporanei è attualmente un problema di grande interesse.

L'utilizzo di array di microfoni è giustificata anche dalla necessità di separare sorgenti multiple, qualora siano spazialmente separate.

Le proprietà del sistema sviluppato in milioni di anni di evoluzione dall'uomo e da altri animali per l'analisi della scena uditiva ha spinto i ricercatori alla ricerca di metodi per la separazione di suoni concorrenti, portando alla nascita e allo sviluppo della cosiddetta "analisi computazionale della scena sonora" (CASA).

Uno schema generale di un sistema per l'analisi di una scena sonora è rappresentato di seguito,

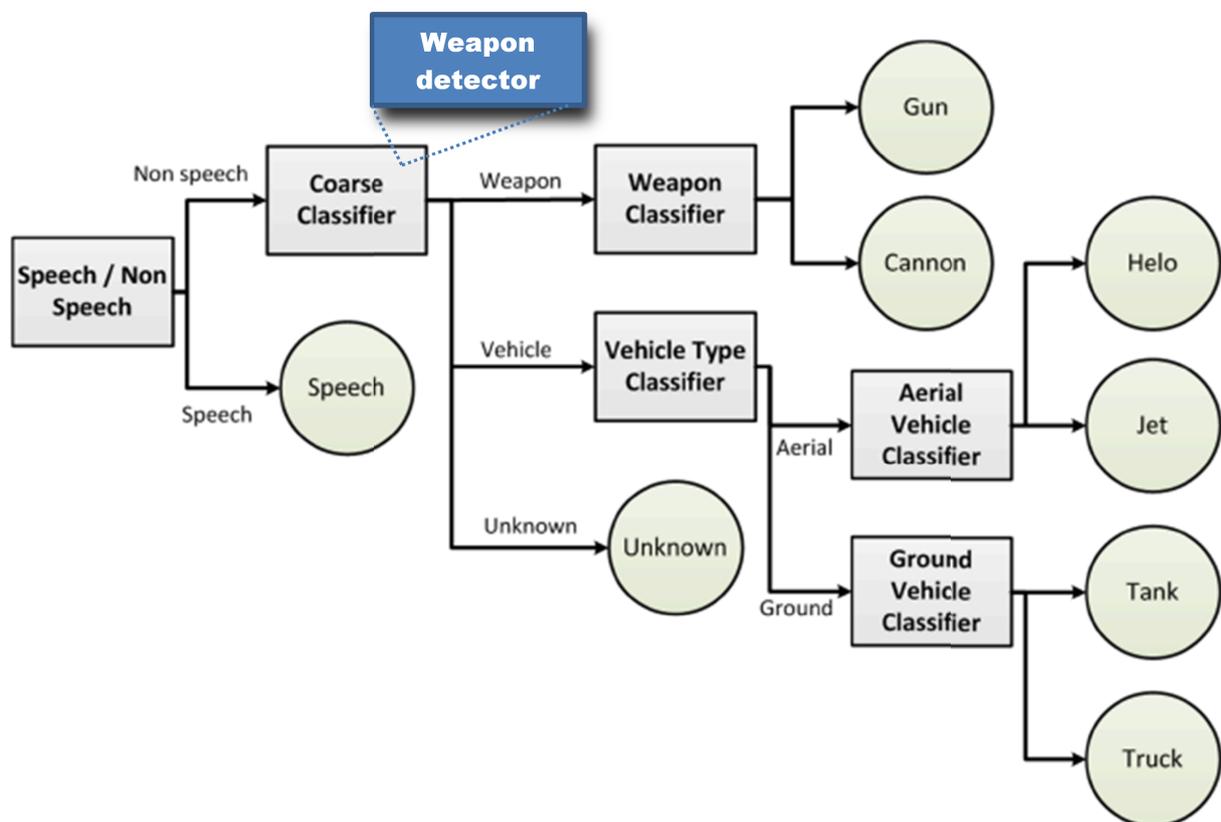


**Figura 4: schema generale di un sistema CASA**

La separazione e il riconoscimento di segnali diversi può naturalmente avvenire con diverse modalità. Ad esempio una parte del sistema potrebbe essere attivata in presenza di un segnale di tipo periodico, con una frequenza fondamentale intorno ai 200 Hz, mentre la presenza di un'altra sorgente potrebbe essere dipendente dalla direzione di provenienza del segnale. I sistemi di riconoscimento automatico si basano essenzialmente sulla rilevazione di proprietà (o features) distintive del segnale. Naturalmente, poiché in una scena possono essere presenti più segnali originati da diverse sorgenti, la detezione di features grezze del segnale può non essere sufficiente.

La giusta combinazione di features deve essere realizzata al fine di rappresentare correttamente il singolo segnale.

L'insieme di queste features sono utilizzate per l'addestramento di un classificatore e quindi la successiva analisi (e classificazione) del segnale. L'idea di base utilizzata nel progetto SMART-OPTIGRID, consiste nel non utilizzare un solo ed unico classificatore con diverse uscite, ma suddividere il problema di classificazione in semplici problemi di classificazione, possibilmente binari. E' infatti plausibile pensare che un errore nella localizzazione o distorsione nel segnale ricevuto incida poco in una classificazione binaria tra due classi molto eterogenee (ad esempio parlato/non parlato), rispetto ad una classificazione a singolo strato, dove le varie classi possono essere molto vicine in termini di diversità delle relative caratteristiche.



**Figura 5: schema generale dell'architettura di classificazione in SMART-OPTIGRID**

In Figura 5 è mostrato lo schema generale dell'architettura di classificazione usato in SMART-OPTIGRID. Il primo strato dell'architettura (a partire dalla sinistra dello schema) è dedicato al riconoscimento della presenza di una sorgente che parla. E' oggi noto che l'analisi di caratteristiche nel dominio della frequenza è sufficiente a risolvere questo

compito con una grande accuratezza, in maniera robusta rispetto alla presenza di rumore ed alla diversità della sorgente che parla (sesso, accento, disfunzioni verbali ecc.).

Il secondo strato dell'architettura classifica un segnale ricevuto all'interno di un numero ristretto di classi molto generali: veicolo, arma o sconosciuto (quest'ultimo permette anche di scartare eventuali errori nella classificazione al livello precedente).

Il sistema si conclude con tre classificatori che permettono di discriminare con più precisione il segnale ricevuto. In particolare, un primo classificatore distingue le armi automatiche dai cannoni, un secondo classificatore gli elicotteri dai jet, mentre un terzo classificatore i tank dai camion.

Il detector realizzato (in figura denominato "Weapon detector") si inserisce all'interno del secondo strato dell'architettura, ovvero nel "Coarse Classifier", con lo scopo di discernere, in modo robusto dal rumore, segnali in cui sono presenti suoni provenienti da armi da fuoco.

### **3. PROBLEMATICHE NELL' ANALISI IN FREQUENZA DI SEGNALI NON STAZIONARI**

Uno degli strumenti principali per l'analisi in frequenza dei segnali non stazionari, ovvero dei segnali il cui spettro varia nel tempo, è la Short Time Fourier Transform (STFT).

Tra la trasformata di Fourier classica (FT) e la STFT vi è una piccola differenza. Nel STFT, il segnale è diviso in segmenti abbastanza piccoli, dove questi segmenti (porzioni) del segnale possono essere assunti stazionari. A questo scopo, viene scelta una funzione finestra "w". La larghezza di questa finestra deve essere uguale al segmento del segnale dove la sua stazionarietà è valida.

La finestra viene posizionata all'inizio del segnale. Cioè, la funzione finestra si trova a  $t = 0$ . Supponiamo che la larghezza della finestra è esattamente "T" secondi.

All'istante  $t = 0$  del segnale, la funzione di finestra si sovrapporrà con i primi  $T / 2$  secondi. La funzione della finestra e il segnale sono poi moltiplicate. In questo modo, solo i primi  $T/2$  secondi del segnale viene scelto, e se ad esempio la finestra è un rettangolo, con ampiezza "1", allora il prodotto sarà proprio uguale al segnale. Quindi il risultato di tale prodotto viene considerato come un segnale a se, e di cui si calcola la FT.

Il risultato di questa trasformazione è la FT dei primi  $T / 2$  secondi del segnale. Se questa porzione del segnale è stazionario, come si presume, quindi non ci sarà alcun problema e il risultato ottenuto sarà una rappresentazione vera in frequenza dei primi  $T/2$  secondi del segnale.

Il passo successivo, sarebbe spostare questa finestra (per  $t_1$  secondi) in una nuova posizione, moltiplicando il segnale, e prendendo la FT del prodotto. Questa procedura viene seguita, fino alla fine del segnale spostando la finestra.

La seguente definizione della STFT riassume tutte le spiegazioni di cui sopra in una riga:

$$STFT_x(\tau, f) = \int_{-\infty}^{+\infty} x(t) w^*(t - \tau) e^{-j2\pi} dt$$

Dove  $x(t)$  è il segnale stesso,  $w(t)$  è la funzione finestra, e  $*$  rappresenta il complesso coniugato. Come si può vedere dall'equazione, la STFT del segnale non è altro che la FT del segnale moltiplicato per una funzione "finestra".

Ci si potrebbe chiedere a questo punto il motivo per cui, per l'analisi spettrale, sia nata la trasformata wavelet, dato che già la STFT dà la rappresentazione tempo-frequenza del segnale.

Il problema dell'uso della STFT ha le radici in ciò che è noto come il principio di indeterminazione di Heisenberg. Questo principio, originariamente applicato al moto e posizione di particelle in movimento, può essere applicato a informazioni in tempo-frequenza di un segnale. Semplicemente, questo principio afferma che non si può sapere l'esatta rappresentazione tempo-frequenza di un segnale, cioè, non si può sapere quali sono le componenti spettrali nei singoli istanti di tempo. Ciò che si può sapere sono gli intervalli di tempo in cui una certa banda di frequenze esiste (ed è quindi un problema di risoluzione).

Il problema con la STFT ha a che fare con l'ampiezza della funzione finestra che viene utilizzata.

Ricordiamo che nel FT non c'è nessun problema di risoluzione nel dominio della frequenza, cioè, sappiamo esattamente quali frequenze esistono; allo stesso modo non c'è problema in termini di risoluzione temporale nel dominio del tempo, dal momento che sappiamo il valore del segnale in ogni istante di tempo. Ciò che dà la risoluzione in frequenza ideale nel FT è il fatto che la finestra utilizzata nel FT è il suo kernel (la funzione  $e^{j\omega t}$ ), di ampiezza da meno infinito a più infinito in ogni istante.

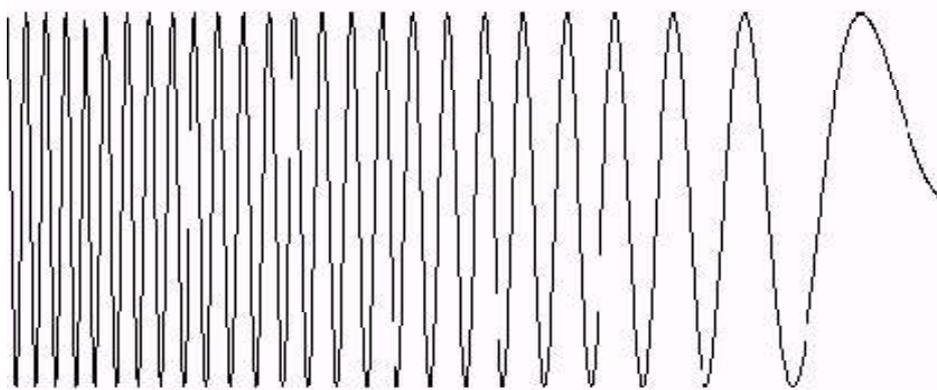
In STFT, la finestra è di ampiezza finita, così copre solo una porzione del segnale, che provoca quindi una risoluzione in frequenza più povera. Cioè meno segnale utilizziamo per la trasformata, e meno si conoscono in modo esatto le componenti di frequenza che esistono nel segnale, e quindi sappiamo solo l'esistenza di una banda di frequenze.

Se usiamo una finestra di lunghezza infinita, si ottiene la FT, che offre una risoluzione di frequenza perfetta, ma nessuna informazione di tempo. Inoltre, al fine di ottenere la stazionarietà, dobbiamo avere una finestra abbastanza breve temporalmente, in cui il segnale è considerato stazionario. Meno ampia è la finestra, e migliore è la risoluzione temporale, e meglio l'assunzione di stazionarietà, ma peggiore è la risoluzione di

frequenza. Più ampia invece è la finestra, migliore è la risoluzione in frequenza ma peggiore è la risoluzione temporale.

Si può facilmente mostrare quanto detto con un esempio.

Consideriamo il segnale in Figura 6. In questo segnale ci sono 4 componenti in frequenza a tempi diversi. L'intervallo da 0 a 250 ms è una semplice senoide di 300 Hz, e gli altri intervalli di 250 ms sono sinusoidi di 200 Hz, 100 Hz e 50 Hz rispettivamente. Tale segnale quindi non è stazionario.



**Figura 6: segnale di esempio**

Consideriamo quattro finestre di diversa lunghezza, per calcolare la STFT. La funzione finestra che usiamo è semplicemente una funzione gaussiana della forma:

$$w(t) = e^{-a \frac{t^2}{2}}$$

dove “*a*” determina la lunghezza della finestra, e “*t*” è il tempo.

La Figura 7 mostra quattro funzioni finestra di varia ampiezza, determinate dal valore di *a*.

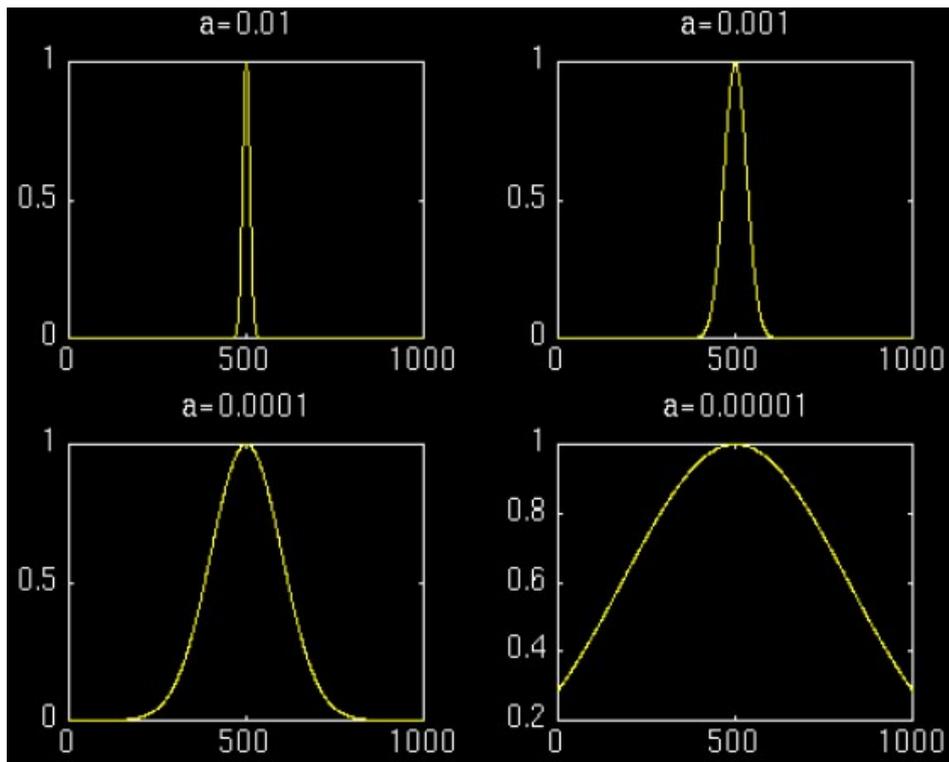


Figura 7:  $w(t)$  per valori di "a" differenti

Applichiamo la STFT con i 4 valori di  $a$  in Figura 7 al segnale dato e otteniamo le rappresentazioni frequenza-tempo del segnale, che mostrano le componenti di frequenza del segnale, e dove si trovano nel tempo (Figura 8, Figura 9, Figura 10, Figura 11).

Gli assi "x" e "y" sono tempo e frequenza, rispettivamente. Ignorando i valori presenti sugli assi, che per lo specifico esempio non è di alcun interesse, esaminiamo le curve della rappresentazione tempo-frequenza.

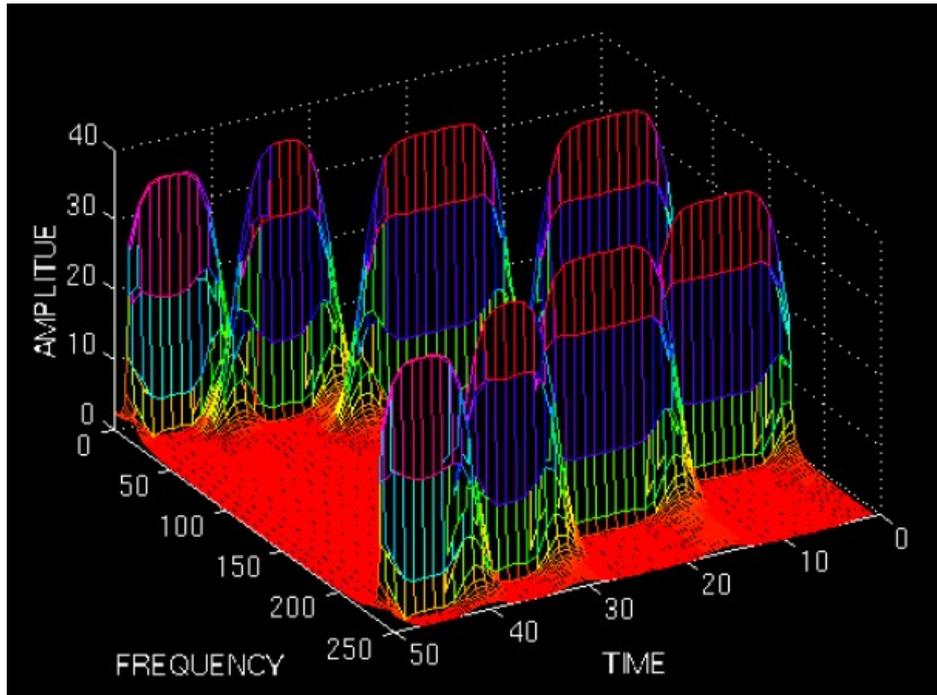


Figura 8 STFT con  $a=0.01$

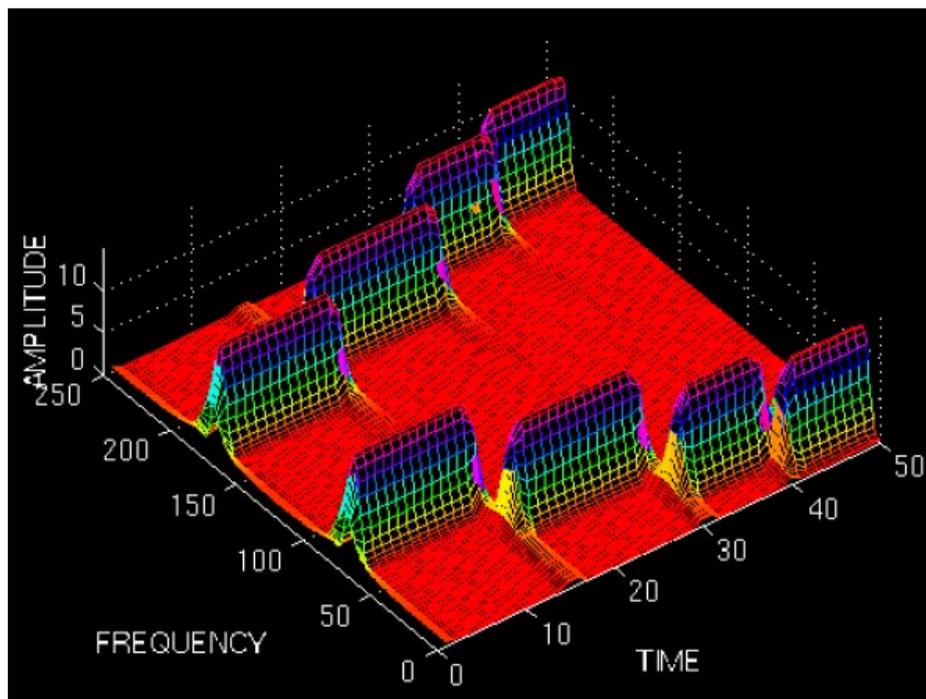


Figura 9: STFT con  $a=0.001$

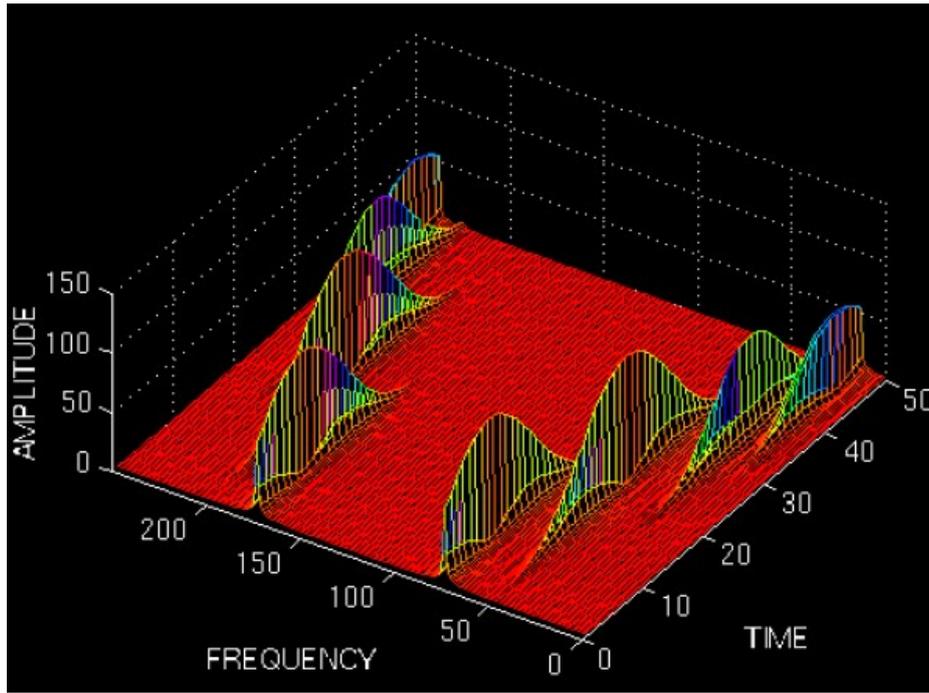


Figura 10: STFT con  $a=0.0001$

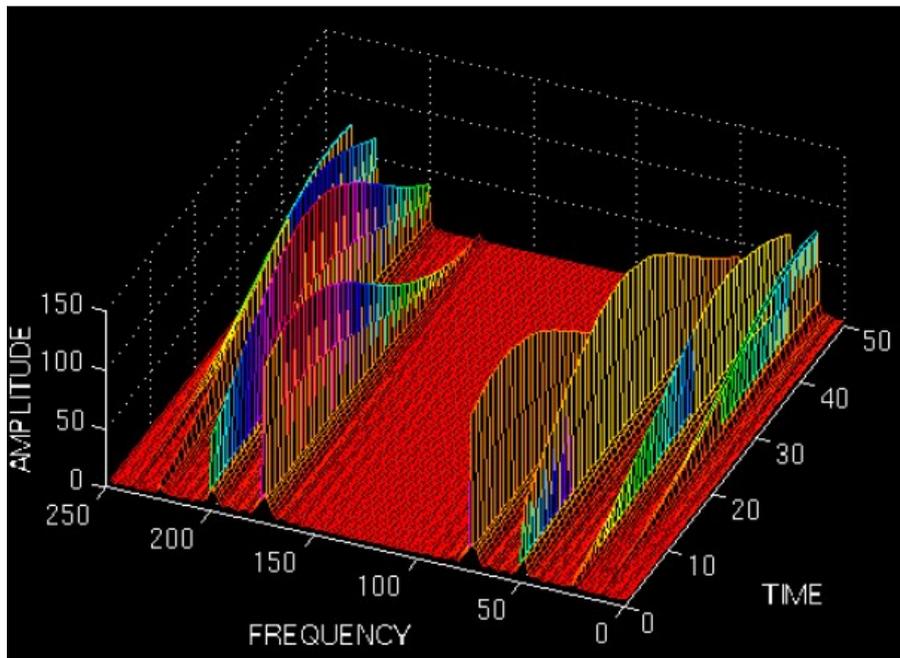


Figura 11: STFT con  $a=0.00001$

Ciò che possiamo subito notare per tutte e quattro le trasformate applicate, sono le quattro cime; infatti ci sono quattro picchi corrispondenti a quattro componenti di frequenza diverse. Si noti inoltre che, a differenza della FT, questi quattro picchi sono

situati in vari intervalli di tempo lungo l'asse del tempo. Si ricorda che il segnale originale aveva quattro componenti spettrali situati in tempi diversi.

In Figura 9 si noti che i quattro picchi sono ben separati l'uno dall'altro nel tempo. Si noti inoltre che, nel dominio della frequenza, ogni picco copre una gamma di frequenze, invece di un valore singola frequenza.

In Figura 10 i picchi non sono ben separati l'uno dall'altro nel tempo, a differenza del caso precedente, tuttavia, nel dominio della frequenza la risoluzione è migliore. Questi ultimi aspetti si enfatizzano ulteriormente aumentando la larghezza della finestra in Figura 11.

Questi esempi dovrebbero aver illustrato il problema implicito della STFT inerente la risoluzione in frequenza e temporale. Finestre "strette" danno una buona risoluzione di tempo, ma la risoluzione in frequenza povera. Viceversa, ampie finestre offrono una buona risoluzione in frequenza, ma povera in risoluzione temporale. Inoltre, ampie finestre possono violare la condizione di stazionarietà.

Una buona scelta dell'ampiezza della finestra, ovviamente, dipende dall'applicazione: se le componenti di frequenza sono ben separate l'una dall'altra nel segnale originale, allora possiamo sacrificare qualche risoluzione in frequenza e andare per buona risoluzione temporale, poiché le componenti spettrali sono già ben separate l'une dalle altre.

La trasformata wavelet, come si vedrà di seguito, risolve il problema della risoluzione.

## 4. TRASFORMATA WAVELET

### 4.1 TRASFORMATA WAVELET CONTINUA (CWT)

La trasformata continua wavelet (CWT) è stata sviluppata come un approccio alternativo per la STFT per superare il problema della risoluzione.

Ricordiamo che piccole variazioni in frequenza della Trasformata di Fourier produrrà cambiamenti ovunque nel dominio del tempo.

L'analisi wavelet viene effettuata in modo simile all'analisi STFT, nel senso che il segnale viene moltiplicato con una funzione, simile alla funzione finestra nel STFT, e la trasformata è calcolata separatamente per i diversi segmenti del segnale nel dominio del tempo.

Tuttavia, ci sono due principali differenze tra la STFT e CWT:

1. Le trasformate di Fourier dei segnali finestrati non vengono calcolate.
2. La larghezza della finestra viene modificata per il calcolo ad ogni singola componente spettrale, che è probabilmente la caratteristica più significativa della trasformata wavelet. Questo la rende "localizzata" in frequenza/scala oltre che in tempo (via traslazione della finestra come nella STFT)

Per definire la CWT, consideriamo lo spazio  $L^2(\mathbb{R})$  di Hilbert delle funzioni misurabili tali che:

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < +\infty$$

Per un segnale  $x(t)$  in  $L^2(\mathbb{R})$ , la CWT è così definita:

$$CWT_{\psi}x(\tau, s) = \int_{-\infty}^{+\infty} x(t)\psi_{\tau,s}^*(t)dt$$

Con:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-\tau}{s}\right)$$

I parametri  $\tau$  ed  $s$  sono detti rispettivamente *parametro di traslazione* e parametro di dilatazione o *scala*.

La *traslazione* è usata nello stesso senso come è stata utilizzata nella STFT: è relativa alla posizione della finestra, la finestra viene spostata attraverso il segnale. Questo termine, ovviamente, si riferisce al tempo. Tuttavia, non abbiamo un parametro di frequenza, come avevamo prima per la STFT. Invece, abbiamo un parametro di scala (che è inversamente proporzionale alla frequenza).

Il parametro di *scala* nell'analisi wavelet è simile alla scala usata nelle mappe. Come nel caso delle mappe, alte scale corrispondono ad una visione globale non dettagliata (del segnale) e scale basse corrispondono a una vista dettagliata. Analogamente, in termini di frequenza, le frequenze basse (alte scale) corrispondono ad una informazione globale di un segnale (che di solito si estende per l'intero segnale), mentre alte frequenze (basse scale) corrispondono alle informazioni dettagliate di un modello nascosto nel segnale (che di solito la durata nel tempo è relativamente breve). Fortunatamente nelle applicazioni pratiche, scale basse (alte frequenze) non durano per tutta la lunghezza del segnale, ma di solito compaiono come picchi di segnale. Alte scale (basse frequenze), invece, di solito durano per tutta la durata del segnale.

Il fattore  $\frac{1}{\sqrt{s}}$  è un fattore di normalizzazione per avere  $\|\psi_{\tau,s}\| = \|\psi\|$ , cioè di modo tale che la funzione abbia la stessa energia a ogni livello di scala.

La funzione  $\psi(t)$  è chiamata *wavelet madre*, o di analisi, e variandone i parametri di traslazione e scala si ottiene una famiglia di wavelet, proprio come nella STFT si otteneva una famiglia di finestre al variare dei parametri di traslazione in tempo e frequenza.

La funzione  $\psi(t)$  è una wavelet ammissibile se e solo se la sua trasformata di Fourier  $\hat{\psi}(\omega)$  soddisfa:

$$\int_0^{+\infty} \frac{|\hat{\psi}(\omega)|^2}{\omega} d\omega = \int_{-\infty}^0 \frac{|\hat{\psi}(\omega)|^2}{|\omega|} d\omega = C_\psi < +\infty$$

Con:

$$\hat{\psi}(\omega) = \int_{-\infty}^{+\infty} \psi(x) e^{-i\omega x} dx \quad (\text{Trasformata di Fourier})$$

Si può dimostrare che questa condizione implica che [24]:

$$\int_{-\infty}^{+\infty} \psi(t) dt = 0$$

Dunque, intuitivamente, la CWT si calcola traslando una wavelet (scalabile) continuamente sul segnale in analisi, e calcolando la correlazione tra i due.

Più dettagliatamente:

Sia  $x(t)$  il segnale da analizzare. La wavelet madre viene scelta allo scopo di essere un "prototipo" per tutte le finestre del processo. Tutte le finestre che vengono utilizzate sono la versione dilatata (o compressa) della wavelet madre.

Una volta che la wavelet madre è stata definita, il calcolo inizia con  $s = 1$  e la trasformata wavelet continua, in linea teorica, dovrebbe essere calcolata per tutti i valori di  $s$ . Tuttavia, a seconda del segnale, una trasformazione completa non è di solito necessaria. Per tutti gli scopi pratici, i segnali sono a banda limitata, e quindi, il calcolo della trasformata per un intervallo limitato di scale è di solito sufficiente. Per comodità, la procedura verrà avviata dalla scala di  $s = 1$  e continuerà per i valori crescenti di  $s$ , cioè, l'analisi partirà da alte frequenze e procedere verso le basse frequenze.

Questo primo valore di  $s$  corrisponderà al wavelet più compresso. Come si aumenta il valore di  $s$ , la Wavelet si dilaterà.

La wavelet è posto all'inizio del segnale nel punto che corrisponde al tempo  $t = 0$ . La funzione wavelet a scala "1" viene moltiplicata per il segnale e poi integrata su tutti i tempi. Il risultato dell'integrazione viene quindi moltiplicato per il numero costante  $\frac{1}{\sqrt{s}}$ . Come abbiamo detto, questa moltiplicazione ha per scopo la normalizzazione dell'energia in modo che il segnale trasformato avrà la stessa energia a tutti i livelli. Il risultato finale è il valore della trasformazione, cioè il valore della wavelet continua trasformata al tempo zero e la scala  $s = 1$ . In altre parole, è il valore che corrisponde al punto  $\tau = 0, s = 1$  nel piano scala/tempo.

La wavelet a scala  $s = 1$  viene quindi spostata verso destra della quantità  $\tau$  nella posizione  $t = \tau$ , e l'equazione di cui sopra è calcolata per ottenere il valore di trasformazione al tempo  $t = \tau, s = 1$  nel piano tempo-frequenza.

Questa procedura viene ripetuta finché la wavelet raggiunge la fine del segnale.

Poi,  $s$  è aumentato di un valore piccolo ripetendo il procedimento descritto per ogni valore di  $s$ .

La Figura 13, Figura 14, Figura 15 illustrano l'intero processo passo dopo passo.

A titolo esemplificativo, viene considerata la wavelet madre "Mexican Hat"

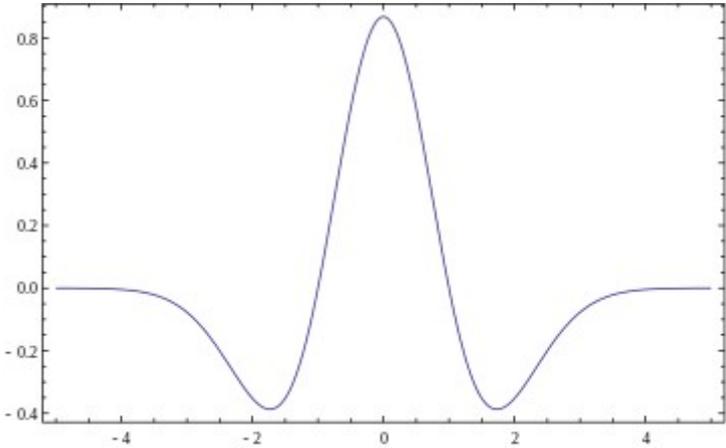


Figura 12: Mexican hat wavelet

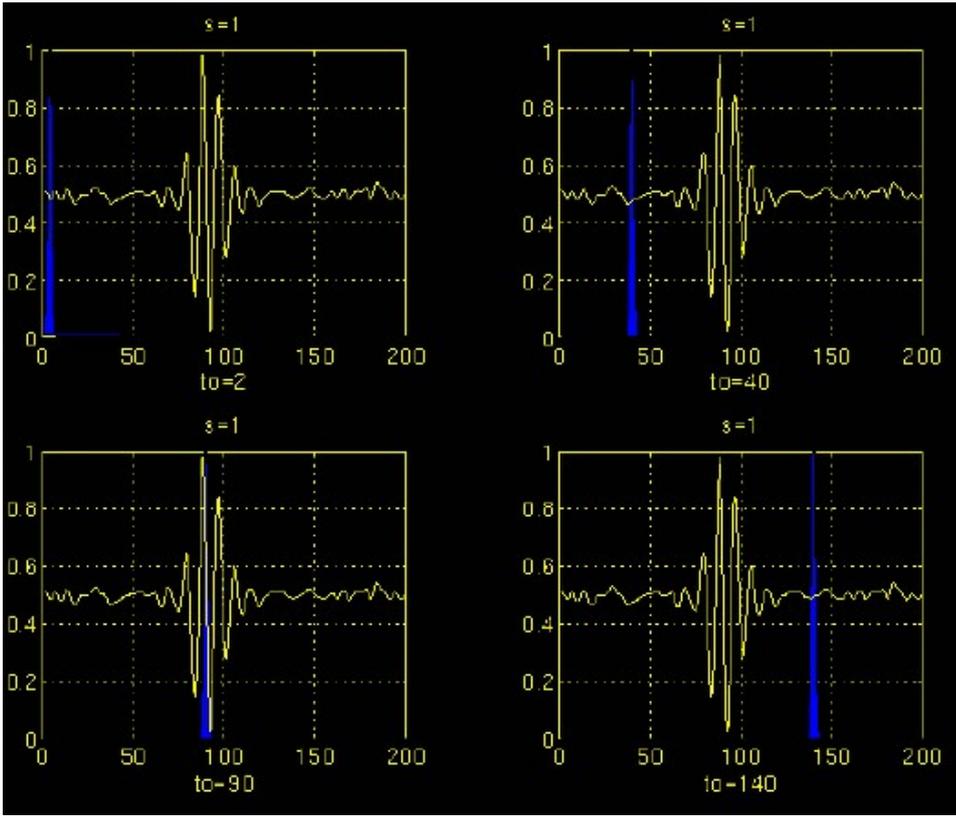


Figura 13: s=1

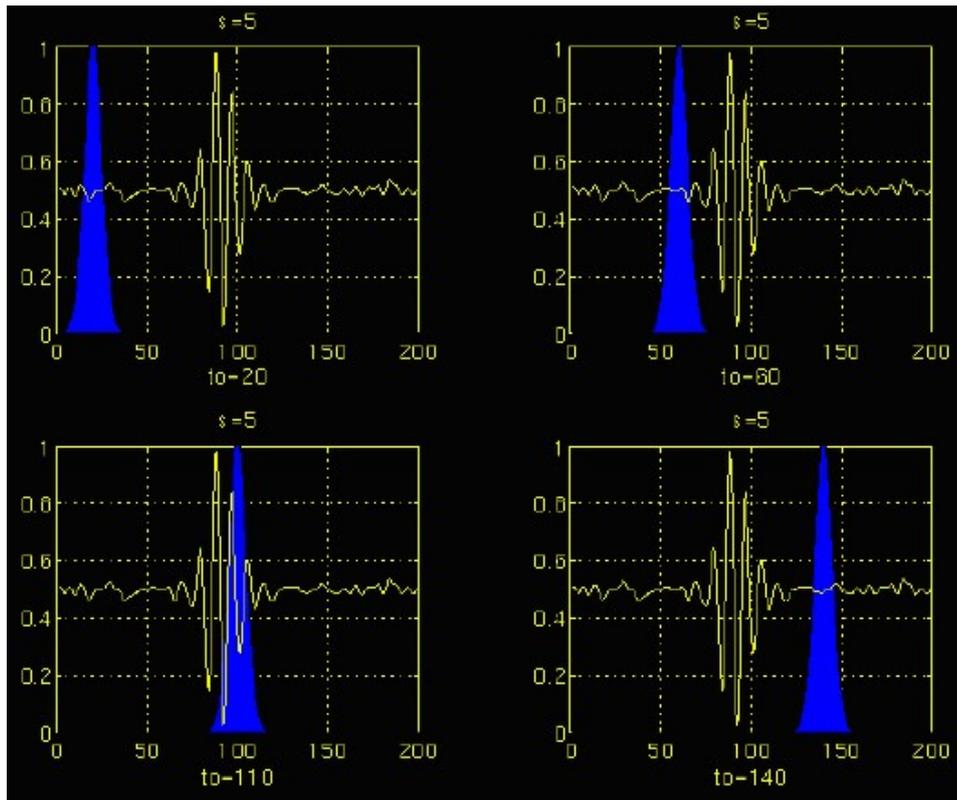


Figura 14:  $s=5$

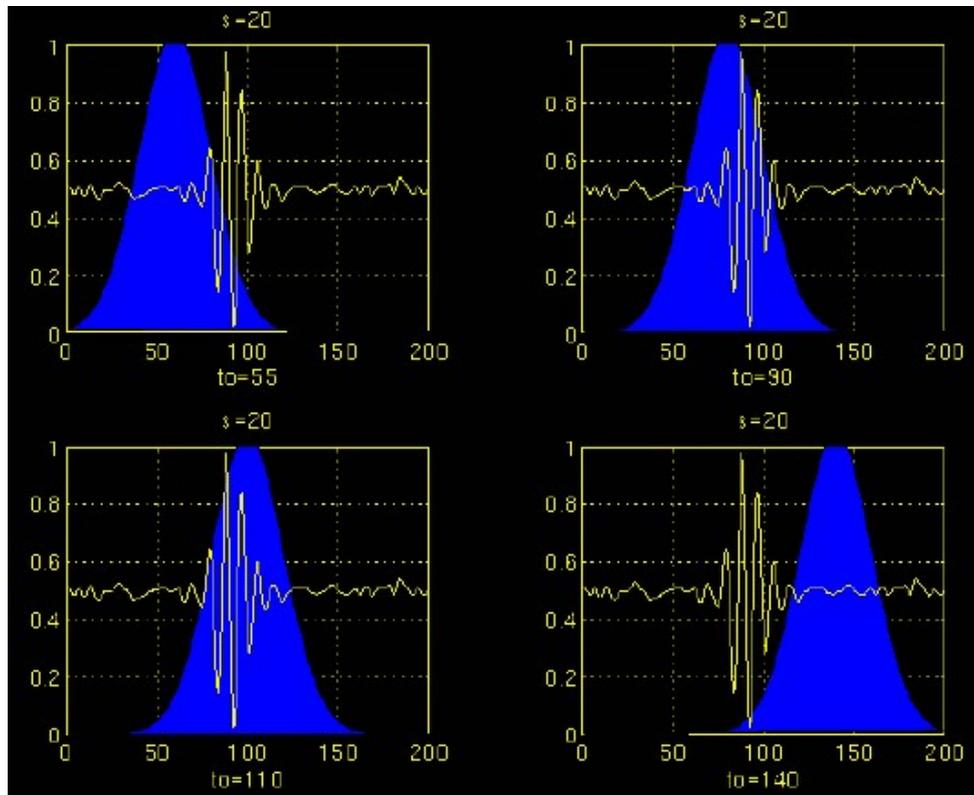


Figura 15:  $s=20$

Nella Figura 13, il segnale e la funzione wavelet sono mostrati per quattro valori diversi di  $\tau$ . Il valore di scala è 1, corrispondente al fondo scala più basso (ovvero alla più alta frequenza). Si noti quanto è stretta la finestra blu. Dovrebbe essere tanto più stretta quanto è più alta la componente in frequenza nel segnale. Sono inoltre mostrate quattro posizioni distinte della funzione wavelet. In ogni posizione la wavelet è moltiplicata per il segnale. Ovviamente, il prodotto è diverso da zero solo quando il segnale cade nella regione di definizione della wavelet, ed è zero altrove.

Questa definizione del CWT mostra che l'analisi wavelet è una misura di somiglianza tra le funzioni di base (wavelet) e il segnale stesso. Qui la somiglianza è nel senso di contenuti in frequenza simile. I coefficienti calcolati da CWT si riferiscono alla vicinanza del segnale al wavelet alla scala corrente.

Se il segnale è un componente importante della frequenza corrispondente alla scala corrente, allora il wavelet (funzione base) alla scala corrente sarà simile o vicino al segnale alla particolare posizione in cui si verifica tale componente di frequenza. Pertanto, il coefficiente CWT calcolato a questo punto del piano scala temporale sarà un numero relativamente grande. Se invece la componente spettrale corrispondente al valore corrente di  $s$  non è presente nel segnale, il valore del prodotto sarà relativamente piccolo o nullo.

Per le scale a valore più alto, d'altra parte, la trasformata wavelet continua darà valori grandi per quasi tutta la durata del segnale, poiché esistono basse frequenze in ogni momento.

La Figura 14, e Figura 15 illustrano lo stesso processo per le scale  $s = 5$  e  $s = 20$ , rispettivamente. Si noti come la larghezza della finestra (blue) cambia con frequenza decrescente e scala crescente. Come la finestra aumenta di larghezza, la trasformazione inizia raccogliendo le componenti a frequenza inferiore.

## **4.2 DISCRETE WAVELET TRANSFORM (DWT)**

La DWT è essenzialmente una rappresentazione in scala temporale di un segnale digitale che viene ottenuta utilizzando tecniche di filtraggio digitale.

Ricordiamo che la CWT è una correlazione tra una wavelet a diverse scale e il segnale, con la scala (o frequenza) utilizzate come misura di similarità. La trasformata wavelet

continua si calcola cambiando la scala della finestra di analisi, spostando la finestra nel tempo, moltiplicando per il segnale, e integrando su tutti i tempi.

Nel caso discreto, filtri di diverse frequenze di taglio sono utilizzati per analizzare il segnale a diverse scale. Il segnale viene fatto passare attraverso una serie di filtri passa alto per analizzare le alte frequenze, e viene fatto passare attraverso una serie di filtri passa-basso per analizzare le frequenze basse.

La risoluzione del segnale, che è una sorta di misura della quantità di informazioni dettagliate nel segnale, viene modificata dalle operazioni di filtraggio, e la scala viene cambiata attraverso operazioni di sovracampionamento e sottocampionamento. Il sottocampionamento di un segnale corrisponde alla riduzione del tasso di campionamento, o rimuovendo alcuni dei campioni del segnale. Quindi, sotto campionare di un fattore  $n$  riduce il numero di campioni del segnale  $n$  volte. Analogamente, sovra campionare un segnale corrisponde a incrementare il tasso di campionamento di un segnale aggiungendo nuovi campioni. Sovra campionare un segnale di un fattore  $n$  equivale a incrementare il numero dei campioni (usualmente aggiungendo degli zeri o valori interpolati dei campioni esistenti) nel segnale di un fattore  $n$ .

Il procedimento, per calcolare la DWT, inizia con passare il segnale campionato  $x$  attraverso un filtro passa basso digitale. Filtrare un segnale corrisponde all'operazione matematica di convoluzione del segnale con la risposta impulsiva del filtro (che, nel caso del passa-basso, indicheremo con  $h$ ).

L'operazione di convoluzione in tempo discreto è definito come segue:

$$y[n] = (x * h)[n] = \sum_{k=-\infty}^{\infty} x[k]h[n - k]$$

Un filtro passa-basso di metà banda rimuove tutte le frequenze che si trovano sopra la metà della più alta frequenza nel segnale. Ad esempio, se un segnale ha una massima componente di frequenza di 1000 Hz, allora il filtraggio passa-basso a metà banda elimina tutte le frequenze superiori a 500 Hz.

Dopo il passaggio del segnale attraverso un filtro passa-basso (mezza band), metà dei campioni può essere eliminato secondo la regola di Nyquist, semplicemente scartando

ogni altro campione, sottocampionando il segnale a 2, e il segnale avrà quindi metà del numero di punti. La scala del segnale è ora raddoppiata.

Si noti che il filtraggio passa basso rimuove le informazioni ad alta frequenza, ma lascia la scala invariata. Solo il processo di sottocampionamento modifica la scala. La risoluzione, d'altra parte, è legata alla quantità di informazioni nel segnale, e quindi, è influenzata dalle operazioni di filtraggio.

Il filtraggio passa-basso (metà band) elimina la metà delle frequenze, che può essere interpretato come perdere metà delle informazioni. Quindi, la risoluzione è dimezzata dopo l'operazione di filtraggio. Si noti, tuttavia, che l'operazione di sottocampionamento dopo l'applicazione del filtro non influenza la risoluzione, poiché la rimozione della metà delle componenti spettrali dal segnale rende metà del numero di campioni comunque ridondanti. Pertanto metà dei campioni possono essere tranquillamente scartati senza alcuna perdita di informazioni.

In sintesi, il filtraggio passa basso dimezza la risoluzione, ma lascia la scala invariata. Il segnale viene poi sottocampionato a 2 in quanto la metà del numero di campioni sono ridondanti (e questo raddoppia la scala).

Questa procedura può essere espressa matematicamente come:

$$y[n] = \sum_{k=-\infty}^{\infty} h[k] \cdot x[2n - k]$$

La DWT analizza successivamente il segnale in diverse bande di frequenza con risoluzioni diverse scomponendo il segnale in uno "grossolano" ed uno "dettagliato". DWT impiega due insiemi di funzioni, chiamate funzioni "smoothing" e funzioni "wavelet", che sono associati con filtri passa-basso e passa-alto, rispettivamente.

La decomposizione del segnale in diverse bande di frequenza è ottenuto semplicemente mediante operazioni di filtraggio (passa alto/passa-basso) successive del segnale nel dominio del tempo. Il segnale originale  $x[n]$  viene prima passato attraverso un filtro (metà banda) passa-alto  $g$  e passa-basso  $h$ . Dopo il filtraggio, la metà dei campioni può essere eliminato secondo la regola di Nyquist. Il segnale può quindi essere sottocampionato per 2. Ciò costituisce un livello di decomposizione e può matematicamente essere espresso come segue:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n - k] \quad y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n - k]$$

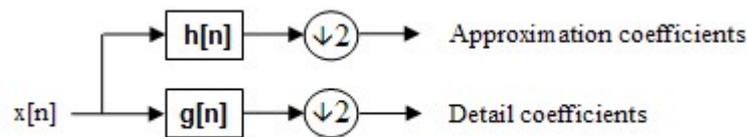


Figura 16: Schema a blocchi di analisi filtro

dove  $y_{high}[k]$  e  $y_{low}[k]$  sono l'output dei filtri passa alto e passa basso, rispettivamente, dopo il sottocampionamento per 2.

Questa scomposizione dimezza la risoluzione temporale in quanto solo la metà del numero di campioni ora caratterizza l'intero segnale. Tuttavia, questa operazione raddoppia la risoluzione in frequenza, poiché la banda di frequenza del segnale ora ricopre solo metà della banda di frequenza precedente, riducendo l'incertezza nella metà delle frequenze.

La procedura di cui sopra, può essere ripetuto per un'ulteriore decomposizione. Ad ogni livello, il filtraggio e sottocampionamento comporterà metà del numero di campioni (e quindi la metà della risoluzione temporale) e la metà della banda di frequenza ricoperta (e quindi raddoppiare la risoluzione di frequenza). La Figura 17 illustra questa procedura, dove  $x[n]$  è il segnale originale da decomposto e  $h[n]$  e  $g[n]$  sono filtri passa-basso e passa-alto rispettivamente.

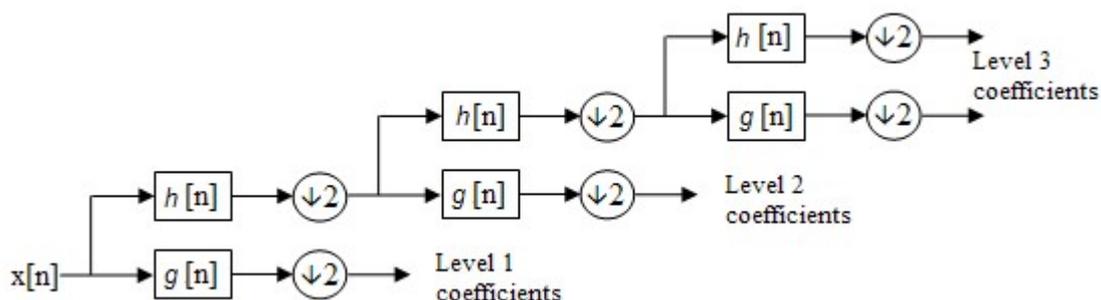


Figura 17

Ad ogni livello nel diagramma sopra il segnale viene scomposto in basse e alte frequenze. A causa del processo di decomposizione, il segnale di ingresso deve essere un multiplo di  $2^n$  dove  $n$  è il numero di livelli.

Per esempio un segnale con 32 campioni, con gamma di frequenza da 0 a  $f_n$  e 3 livelli di decomposizione, 4 scale di uscita sono prodotti:

Livello	Frequenze	campioni
3	0 a $f_n/8$	4
	$f_n/8$ a $f_n/4$	4
2	$f_n/4$ a $f_n/2$	8
1	$f_n/2$ a $f_n$	16

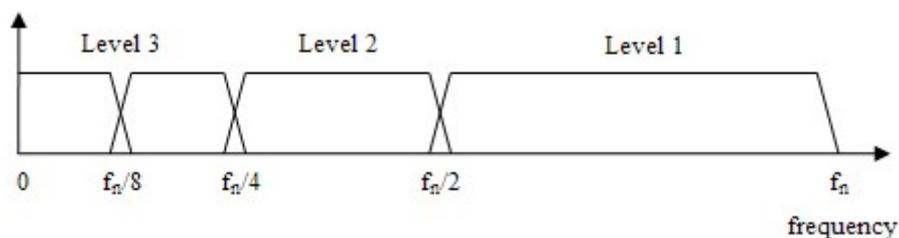


Figura 18: Rappresentazione nel dominio della frequenza del DWT

### 4.3 WAVELET PER IL MULTISCALE EDGE DETECTION

Gli Operatori gradiente sono un mezzo classico di stima degli “step changes” in segnali e immagini, e possono essere applicati per il rilevamento degli “shock edges”.

Gli operatori 2-D comunemente utilizzati per il rilevamento dei bordi (quali Roberts, Prewitt, e Sobel) si riducono ad una dimensione come un filtro FIR con risposta impulsiva  $[-1,0,1]$ . Estensioni più generali, i cosiddetti metodi derivativi filtrati, combinano lo smussamento con la stima del gradiente al fine di ridurre gli effetti del

rumore, e sono più efficaci in presenza di elevato rumore. Questi metodi sono diffusi per via di una bassa complessità di implementazione, oltre ad essere tendenzialmente localizzati, e una robustezza agli sfondi altamente variabili e ai punti multipli di cambiamento.

Un metodo derivativo filtrato, piuttosto diffuso, è la derivata della gaussiana (dG), che stima il gradiente dopo lo smoothing con una funzione gaussiana. Il livello di smoothing è determinato dalla varianza della Gaussiana.

Ottenere simultaneamente sia risultati di rilevazione che di stima è un compromesso tra il livello di smoothing e la varianza della stima della posizione dello “step-edge”, e questo compromesso è sensibile alla forma del bordo e al SNR (signal noise rate).

Il livello di smoothing ottimale non è tipicamente noto a priori, ma il problema della sua scelta, per la stima del gradiente, può essere superato in qualche misura impiegando un'analisi multiscala, cioè, combinando i risultati su più livelli di smussatura.

Infatti, la maggior parte degli “edge detectors” multiscala smussano il segnale alle varie scale e rilevano i punti di variazione del segnale dalla loro derivata prima o seconda. Gli estremi della derivata prima corrispondono ai “zero crossing” della derivata seconda e ai punti di flesso del segnale smussato.

Spieghiamo quindi in che relazione tali metodi multi scala per l'edge detection sono in relazione con la trasformata wavelet:

Indichiamo con “smoothing function” qualsiasi funzione  $\theta(x)$  il cui integrale è 1 e che converge a 0 all'infinito. Per esempio possiamo scegliere come  $\theta(x)$  la funzione Gaussiana. Supponiamo che  $\theta(x)$  è due volte differenziabile e definiamo, rispettivamente, con  $\psi^a(x)$  e  $\psi^b(x)$  la derivata prima e seconda di  $\theta(x)$ . Ovvero:

$$\psi^a(x) = \frac{d\theta(x)}{dx}$$

$$\psi^b(x) = \frac{d^2\theta(x)}{dx^2}$$

Per definizione, le funzioni  $\psi^a(x)$  e  $\psi^b(x)$  possono essere considerate *wavelets* perché il loro integrale è 0. Ovvero:

$$\int_{-\infty}^{+\infty} \psi^a(x) dx = 0$$

$$\int_{-\infty}^{+\infty} \psi^b(x) dx = 0$$

Definiamo:

$$\xi_s(x) = \frac{1}{s} \xi\left(\frac{x}{s}\right)$$

$\xi_s(x)$  rappresenta la dilatazione di un fattore di scala  $s$  di una qualsiasi funzione  $\xi(x)$  (e in particolare per le funzioni  $\psi^a(x)$  e  $\psi^b(x)$ ). Come abbiamo detto, una trasformata wavelet è calcolata attraverso la convoluzione del segnale con una wavelet dilatata. Quindi la trasformata wavelet del segnale  $f(x)$  alla scala  $s$  e posizione  $x$ , calcolata rispetto la wavelet  $\psi^a(x)$ , è definita da:

$$W_s^a f(x) = f * \psi_s^a(x) = \int_{-\infty}^{+\infty} f(u) \psi_s^a(x - u) du$$

La trasformata wavelet di  $f(x)$  rispetto  $\psi^b(x)$  è:

$$W_s^b f(x) = f * \psi_s^b(x) = \int_{-\infty}^{+\infty} f(u) \psi_s^b(x - u) du$$

Si deriva che:

$$W_s^a f(x) = f * \psi_s^a(x) = f * \left( s \frac{d\theta(x)}{dx} \right) (x) = s \frac{d}{dx} (f * \theta_s)(x)$$

$$W_s^b f(x) = f * \psi_s^b(x) = f * \left( s^2 \frac{d^2\theta(x)}{dx^2} \right) (x) = s^2 \frac{d^2}{dx^2} (f * \theta_s)(x)$$

Le trasformate wavelet  $W_s^a f(x)$  e  $W_s^b f(x)$  sono, rispettivamente, la derivata prima e seconda del segnale smussato alla scala  $s$ . Gli estremi locali di  $W_s^a f(x)$  corrispondono ai punti di "zero-crossing" di  $W_s^b f(x)$  e ai punti di flesso di  $(f * \theta_s)(x)$ . Nel caso particolare dove  $\theta(x)$  è la Gaussiana, il detection dei punti di zero-crossing è equivalente a fare l'edge detection con il metodo di Marr-Hildreth [22] mentre il detection degli estremi locali corrisponde a fare l'edge detection con il metodo Canny [23]. Quando la scala  $s$  è grande, la convoluzione con  $\theta_s(x)$  rimuove le piccole fluttuazioni del segnale; è necessario quindi solo rilevare le forti variazioni di grandi strutture.

Il detection dei punti zero-crossing o dei estremi locali è una procedura simile, ma l'approccio del detection di estremi locali ha alcuni vantaggi. Un punto di flesso di  $(f * \theta_s)(x)$  può essere o un massimo o un minimo del valore assoluto della sua derivata prima. I massimi del valore assoluto della derivata prima, sono i punti di variazione nitide di  $(f * \theta_s)(x)$  mentre il minimo corrisponde a piccole variazioni. Con un operatore di derivata seconda, è difficile distinguere questi due tipi di zero-crossing. Al contrario, con un operatore di derivata prima, si riesce a selezionare punti di variazione nette rilevando solo i massimi locali di  $|W_s^a f(x)|$ . Inoltre, zero crossing da informazioni in più sulla posizione ma non sulla differenza tra piccole fluttuazioni e discontinuità importanti.

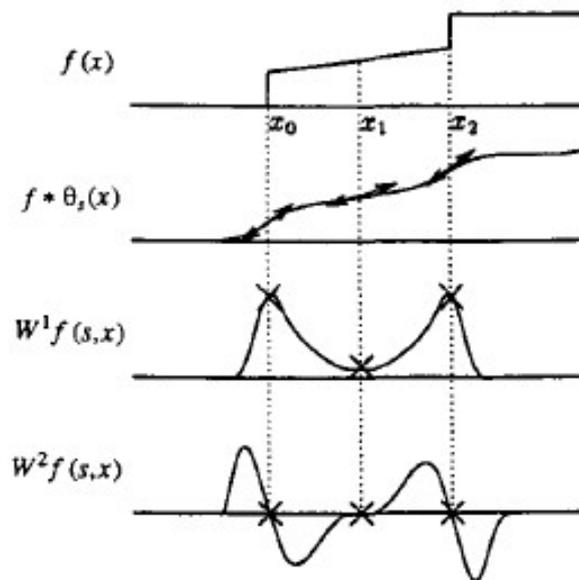


Figura 19: gli estremi di  $W^1 f(s, x)$  e i zero-crossing di  $W^2 f(s, x)$  sono i punti di flesso di  $(f * \theta_s)(x)$

#### 4.4 MALLAT-ZHONG WAVELET (MZ-DWT)

La wavelet utilizzata per la realizzazione del detector, oggetto della tesina, è quella sviluppata da Mallat e Zhong nel 1992 [2]. Si tratta di una DWT (che chiameremo MZ-DWT) basata su una funzione di smoothing  $\theta(x)$  (spline cubica) che è un'approssimazione di una gaussiana e la wavelet  $\psi(x)$ , di conseguenza, è una spline quadratica. La DWT così definita è equivalente a un edge detection di Canny ([23]).

$$\theta(x) = \begin{cases} 0 & |x| \geq 1 \\ \theta(-x) & 0 \leq x \leq 1 \\ -8x^3 - 8x^2 + 4/3 & -0.5 \leq x \leq 0 \\ 8(x+1)^3/3 & -1 \leq x \leq -0.5 \end{cases}$$

$$\psi(x) = \begin{cases} 0 & |x| \geq 1 \\ -\psi(-x) & 0 \leq x \leq 1 \\ -24x^2 - 16x & -0.5 \leq x \leq 0 \\ 8(x+1)^2 & -1 \leq x \leq -0.5 \end{cases}$$

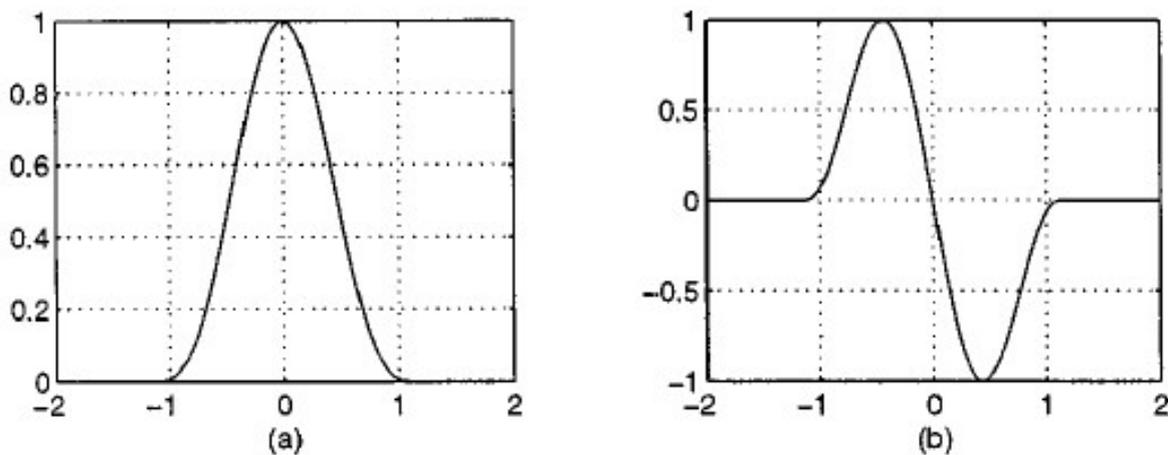


Figura 20: (a) funzione di smoothing  $\theta(x)$  (spline cubica) che è un'approssimazione della gaussiana (b) la sua derivata  $\psi(x) = \frac{d\theta(x)}{dx}$

MZ-DWT implementa l'algoritmo dG a vari livelli di smoothing. La discretizzazione è diadica in scala ( $s = 2^j, j \in \mathbb{Z}$ ) ma non è diadica nel tempo (shift), che corrisponde a una catena di filtri senza un sottocampionamento [par.4.2].

La MZ-DWT del segnale  $x(n)$ ,  $1 \leq n \leq N$  consiste di:

$$W_{2^j}x(n), \quad j = 1, 2, \dots, J-1$$

Dove  $J = \log_2 N$ , più le informazioni "grossolane" della rimanente scala, denotate da  $S_{2^j}(n)$ . Così la MZ-DWT, che consiste di  $J \times N$  punti, è "overcomplete", ovvero ha lo stesso numero di punti ad ogni scala, che contrasta con quanto detto relativamente alle trasformate wavelet discrete in 4.2, dove il numero di coefficienti decresce all'aumentare della scala.

In Figura 21, viene mostrata la risposta impulsiva della MZ-DWT per le prime cinque scale. Queste approssimano la derivata della gaussiana (dG) per vari livelli di "smussamento". Si noti come la figura (a) sia una stima non smussata.

In Figura 22 (a) è mostrato un esempio di segnale (256 campioni) in cui gli "step edges" si verificano nel primo terzo del segnale, un bordo tagliente bifacciale nel secondo terzo, e più bordi rumorosi nell'ultimo terzo. Gli estremi locali della trasformata wavelet corrispondono ai bordi del segnale originale. Solo i bordi principali vengono rilevati alle scale più basse poiché la wavelet nella convoluzione è stata ulteriormente dilatata. I bordi più importanti vengono rilevati in tutti i livelli di trasformazione mentre i bordi minori sono rilevati solo alle più alte scale. Inoltre, il tipo di bordo influenza la forma degli estremi locali.

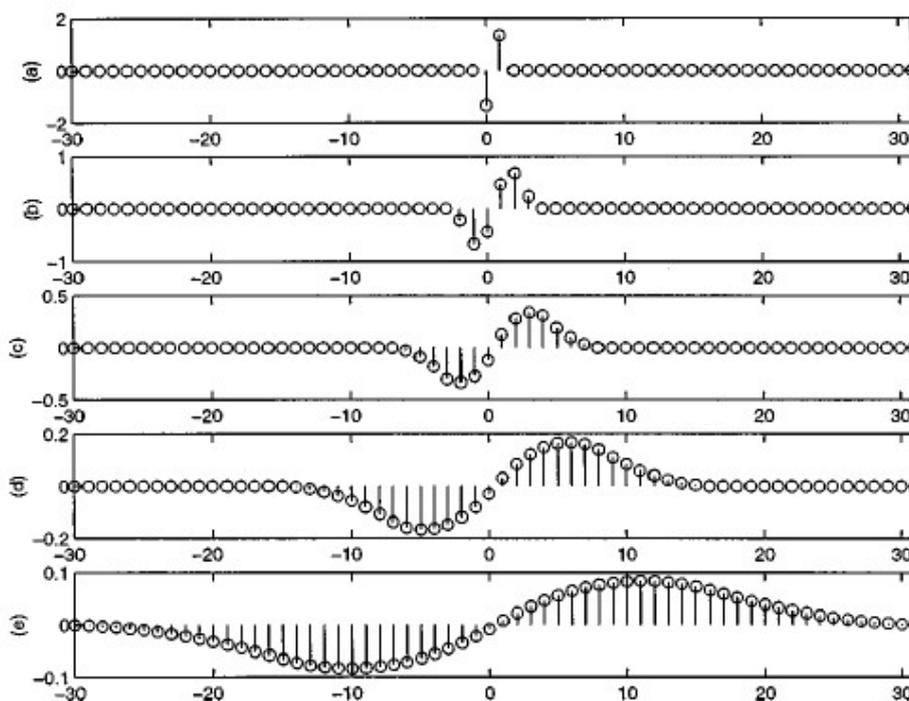


Figura 21: Risposta impulsiva (IR) della MZ-DWT per le prime le prime 5 scale

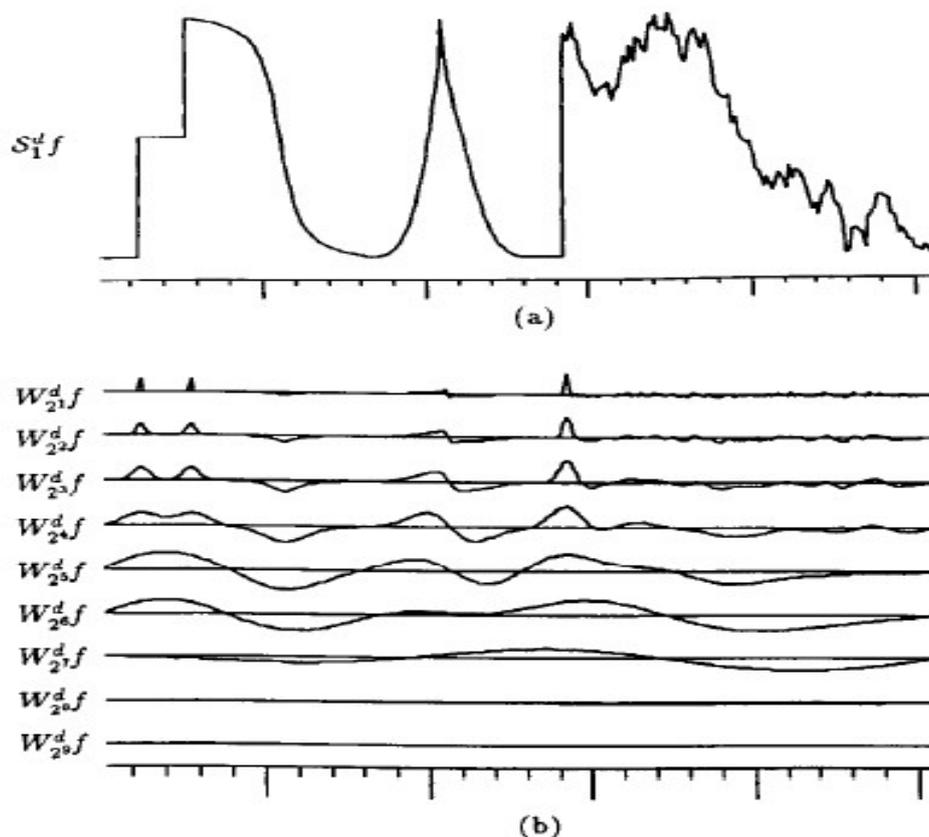


Figura 22:

L'articolo di Mallat e Zhong include un pseudo-codice per implementare la trasformata wavelet discreta diadica. Ad ogni scala diadica, il codice decompone l'input in coefficienti di dettaglio e coefficienti di approssimazione. I coefficienti di dettaglio corrispondono alle informazioni ad alta frequenza e contengono le informazioni del "bordo", mentre i coefficienti di approssimazione corrispondono alle informazioni a bassa frequenza.

```

j = 0
while (j < J)
    W_{2^{j+1}}^d f = \frac{1}{\lambda_j} \cdot S_{2^j}^d f * G_j
    S_{2^{j+1}}^d f = S_{2^j}^d f * H_j
    j = j + 1
end of while.

```

Figura 23: Pseudo-codice per la trasformata wavelet discrete MZ-DWT

$J$  è il numero totale di livelli di decomposizione che è limitata dalla lunghezza del segnale di ingresso.  $H$  e  $G$  e sono filtri discreti che caratterizzano la wavelet madre  $\psi$ .  $H$  è un filtro passa-basso mentre  $G$  è un filtro passa-alto

N	H	G
-3		
-2		
-1	0.125	
0	0.375	-2.0
1	0.375	2.0
2	0.125	
3		

**Figura 24: Finite Impulse Response (FIR) dei filtri che corrispondono alla wavelet spline quadratica**

In Figura 24, è mostrata una tabella che contiene le risposte impulsive di  $H$  e  $G$ . La risposta all'impulso di  $H$  è convoluta sul segnale  $f$  per produrre i coefficienti di approssimazione  $S_{2^j}^d f$  mentre la risposta impulsiva di  $G$  è convoluta sul segnale  $f$  per produrre i coefficienti dettaglio.

$H_j$  e  $G_j$  sono le versioni sovra campionate dei filtri originali con  $2^j - 1$  zeri inseriti tra ogni coefficiente. L'operatore di convoluzione è indicato con il simbolo  $*$ . I bordi artefatti generati dalla convoluzione discreta sono ridotti al minimo utilizzando un "riempimento" simmetrico periodico. Il valore estremo di una wavelet in prossimità di un bordo dovrebbe avere lo stesso valore a tutte le scale diadiche ma a causa della discretizzazione non è così. Le costanti  $\lambda_i$  compensano questo effetto.

j	$\lambda_j$
1	1.50
2	1.12
3	1.03
4	1.01
5	1.00

**Figura 25: coefficienti di normalizzazione per la wavelet. Per  $j > 5, \lambda_i = 1$**

Il codice è inizializzato impostando  $S_{2^0}^d f$  uguale al segnale di ingresso originale

## 5. DETECTION DI UN SEGNALE

La teoria della detezione del segnale o teoria della rilevazione è un metodo per quantificare l'abilità di distinguere, in un segnale, il segnale vero e proprio portatore di informazioni dal rumore. Se consideriamo il segnale come un insieme di dati, caratterizzato da un modello probabilistico (ignoto), il tipo più semplice di problema inferenziale è quello di decidere quale dei due modelli probabilistici (segnale portatore di informazione e rumore) descrive meglio tale insieme di dati.

Sia  $X$  una variabile casuale, e denotiamo i due modelli di probabilità come  $p(x|H_0)$  e  $p(x|H_1)$ . Qui  $H_0$  e  $H_1$  stanno per "ipotesi 0 e 1", rispettivamente. In altre parole, abbiamo due ipotesi circa come i dati  $x$  potrebbero essere stati generati.

Per semplificare la notazione, si scriverà  $p_i(x)$  per indicare  $p(x|H_i)$ .

Assumiamo quindi che la variabile  $X$  sia distribuita in accordo ad una delle due distribuzioni:

$$H_0: X \sim p_0$$

$$H_1: X \sim p_1$$

Supponiamo che  $p_0$  e  $p_1$  siano noti. In tal caso, decidere quale delle distribuzioni "fitta" meglio la variabile  $X$  è chiamata "simple binary hypothesis test" ("simple" perché le due distribuzioni sono note in modo preciso). La decisione è presa partizionando il range dei valori di  $X$  in due regioni disgiunte.

Siano  $R_0$  e  $R_1$  tali regioni e sia  $x$  il valore osservato per la variabile  $X$ . Se  $x \in R_i$ , allora stabiliamo che  $H_i$  è l'ipotesi più verosimile e cioè che i dati sono distribuiti in accordo a  $p_i$ .

La questione chiave è: come costruiamo  $R_0$  e  $R_1$ ? (da notare che  $R_0 \cup R_1$  è l'intero range di valori assunti dalla variabile  $X$  e  $R_0$  e  $R_1$  sono complementari).

Ci sono quattro possibili risultati di un test di questa forma, a seconda della decisione che prendiamo ( $H_0$  o  $H_1$ ), e la reale distribuzione dei dati (sempre  $H_0$  o  $H_1$ :). Indichiamo questi risultati con  $(0, 0)$ ,  $(0, 1)$ ,  $(1, 0)$  e  $(1, 1)$ , dove il primo argomento indica la decisione basata sulle regioni  $R_0$  e  $R_1$  e il secondo indica la reale distribuzione che ha generato i dati. Si noti che i risultati  $(0, 1)$  e  $(1, 0)$  sono errori.

Al fine di ottimizzare la scelta delle regioni di decisione, possiamo specificare un costo per l'errata decisione. Senza perdita di generalità, supponiamo che i costi non sono negativi. Sia  $c_{i,j}$  il costo associato al risultato  $(i, j)$ . I costi riflettono la relativa importanza di una corretta e scorretta decisione.

Poiché il nostro obiettivo è quello di progettare un test che fa pochi errori, è ragionevole supporre che  $c_{1,0}$  e  $c_{0,1}$  sono maggiori di  $c_{0,0}$  e  $c_{1,1}$ ; infatti spesso è ragionevole assegnare un costo nullo per decisioni corrette. Inoltre in linea generale i costi  $c_{1,0}$  e  $c_{0,1}$  possono essere differenti.

Il costo totale associato a un test (cioè con le regioni  $R_0$  e  $R_1$ ) è usualmente chiamato *Bayes Cost*, ed è definito come segue:

$$C = \sum_{i,j=0}^1 c_{i,j} P(H_j \text{ è vera}) P(\text{decisione per } H_i | H_j \text{ è vera})$$

Sia  $\pi_j := P(H_j \text{ è vera})$ .  $\pi_j$  è la *probabilità a priori di  $H_j$*  ovvero la probabilità che un'osservazione sarà generata in accordo a  $p_j$ . Il nostro obiettivo è costruire le regioni di decisioni  $R_0$  e  $R_1$  in modo da minimizzare il *Bayes Cost*.

Assumiamo che  $p_0$  e  $p_1$  sono densità continue (vale comunque nel caso discreto).

$$\begin{aligned} C &= \sum_{i,j=0}^1 c_{i,j} \pi_j P(\text{decisione per } H_i | H_j \text{ è vera}) = \\ &= \sum_{i,j=0}^1 c_{i,j} \pi_j P(X \in R_i | H_j \text{ è vera}) = \\ &= \sum_{i,j=0}^1 c_{i,j} \pi_j \int_{R_i} p_j(x) dx = \\ &= \int_{R_0} (c_{0,0} \pi_0 p_0(x) + c_{0,1} \pi_1 p_1(x)) dx + \int_{R_1} (c_{1,0} \pi_0 p_0(x) + c_{1,1} \pi_1 p_1(x)) dx \end{aligned}$$

Gli integrandi sono non-negativi, da cui segue che  $R_0$  è l'insieme degli  $x$  tali che il primo integrando è più piccolo del secondo. Cioè:

$$R_0 := \{x: (c_{0,0} \pi_0 p_0(x) + c_{0,1} \pi_1 p_1(x)) < (c_{1,0} \pi_0 p_0(x) + c_{1,1} \pi_1 p_1(x))\}$$

$$R_1 := \{x: (c_{0,0} \pi_0 p_0(x) + c_{0,1} \pi_1 p_1(x)) > (c_{1,0} \pi_0 p_0(x) + c_{1,1} \pi_1 p_1(x))\}$$

Quindi, il test (relativo ai costi assegnati) assume la seguente forma:

$$\frac{p_1(x)}{p_0(x)} \underset{H_0}{\overset{H_1}{>}} \frac{\pi_0 (c_{1,0} - c_{0,0})}{\pi_1 (c_{0,1} - c_{1,1})}$$

$\frac{p_1(x)}{p_0(x)}$  è chiamato *likelihood ratio* e il test è denominato *likelihood ratio test (LRT)*.

Il termine alla destra  $\gamma := \frac{\pi_0 (c_{1,0} - c_{0,0})}{\pi_1 (c_{0,1} - c_{1,1})}$  è una costante ( $> 0$ ) che dipende dalle probabilità a priori e dai costi (e quindi non dipende dai dati  $x$ ).  $\gamma$  rappresenta il *threshold* del test.

## 6. ANALISI DI UN DETECTOR PER RILEVARE UNO SPARO IN UN SEGNALE

Nel seguito verrà descritto un detector di shock wave edges (ovvero delle variazioni brusche del segnale che sono tipiche di uno sparo d'arma da fuoco) basato sul prodotto multi-scala, della trasformata Wavelet, proposto da Sadler [1] e migliorato da Xie et al, in [5]. Tale metodologia ha la caratteristica di avere una bassa complessità computazionale e migliore robustezza al rumore.

Questo approccio sfrutta le variazioni veloci del segnale che caratterizzano la presenza di un "bordo" (variazione brusca del segnale). Per il rilevamento di tali "bordi" è utilizzata la trasformata wavelet discreta proposta da Mallat-Zhong (MZ-DWT) [2].

Denotiamo l'operatore MZ-DWT con  $W$  e indichiamo con  $X_j[k] := Wx[k]$  il coefficiente wavelet al tempo  $k$  e alla scala  $2^j$ , dove  $x[k]$  rappresenta un'osservazione rumorosa a un tempo discreto  $k$ .

Attraverso un'analisi multi-scala si possono scegliere le scale appropriate per l'applicazione del metodo. Siano esse  $j_1 \dots j_m$ . Tale metodo è basato sul concetto di cross-correlazione delle scale per il detection di "bordi".

Si definisce il prodotto multiscala come segue:

$$\Lambda(k) = \prod_{j=j_1}^{j_m} |X_j(k)|$$

$\Lambda(k)$  è ovviamente una funzione non lineare del segnale in input.

La funzione  $\Lambda(k)$  avrà "picchi" in presenza di "bordi" mentre avrà valori relativamente piccoli altrove. L'idea di una correlazione cross-scale è stata sviluppata inizialmente da Rosenfeld per il riconoscimento dei bordi nelle immagini [3], e successivamente utilizzato nell'analisi delle immagini con l'uso di wavelet per l'eliminazione del rumore, da Xu et al. [4].

Per rilevare "i picchi" definiamo un semplice detector (multi-scale product detector):

$$L_1 = \max_{\{k\}} \prod_{j=j_1}^{j_m} |X_j(k)|$$

Le singolarità producono picchi di “cross-scala” in  $X_j[k]$  e questi sono rinforzati in  $L_1(k)$ .

I picchi di segnale si allineeranno tra le varie scale per le prime scale, ma non per tutte le scale perché aumentando il livellamento del segnale, si espanderà la risposta che separerà le singolarità nel tempo. Così la scelta di  $j_1$  troppo grande si tradurrà in picchi non allineati in  $\Lambda(k)$  (una buona scelta è  $j_1 \leq 5$ ).

Il detector (vedi [5]) può essere derivato dal likelihood ratio test con certe assunzioni riguardo le distribuzioni.

Il problema di detection può essere così formulato nel dominio della wavelet:

$$H_0: X_j[k] \sim p(X_j[k]|H_0) \quad k = 1 \dots N$$

$$H_1: X_j[k] \sim p(X_j[k]|H_1) \quad k = 1 \dots N$$

Il likelihood ratio detector è:

$$L(X) = \frac{p(X|H_1)}{p(X|H_0)} = \frac{\prod_{k=1}^N p(X(k)|H_1)}{\prod_{k=1}^N p(X(k)|H_0)}$$

Dove  $X(k) = \{X_{j_1}(k), X_{j_2}(k), \dots, X_{j_m}(k)\}$ . (n.b. assumiamo che i coefficienti siano indipendenti l'uno con l'altro).

Consideriamo il detection di un “bordo” in un segnale tracciando il massimo tra le varie scale. Per semplicità consideriamo un unico “bordo” (al tempo  $i_E$ ) a cui corrisponde un valore massimo fra le scale. Le ipotesi diventano:

$$H_0: X_j[k] \sim p(X_j[k]|H_0) \quad k = 1 \dots N$$

$$H_1: X_j[k] \sim p(X_j[k]|H_1) \quad k = i_E$$

$$X_j[k] \sim p(X_j[k]|H_0) \quad k \neq i_E$$

Dove  $i_E$  è l'istante del “bordo”.

Il detector diventa allora:

$$L(X) = p(X(k)|H_1)|_{k=i_E} \frac{\prod_{k=1, k \neq i_E}^N p(X(k)|H_0)}{\prod_{k=1}^N p(X(k)|H_0)} =$$

$$= \frac{p(X(k)|H_1)}{p(X(k)|H_0)} \Big|_{k=i_e}$$

Poiché  $i_E$  è l'istante di tempo del solo bordo sotto l'ipotesi  $H_1$ , si ha banalmente che:

$$\frac{\prod_{k=1}^N p(X(k)|H_1)}{\prod_{k=1}^N p(X(k)|H_0)} \Big|_{k=i_e} > \frac{\prod_{k=1}^N p(X(k)|H_1)}{\prod_{k=1}^N p(X(k)|H_0)} \Big|_{k \neq i_e}$$

In altre parole, nell'istante di tempo  $i_E$ , il likelihood ratio assume il massimo valore. Se ne deduce quindi che:

$$i_E = \arg \max_k \frac{\prod_{k=1}^N p(X(k)|H_1)}{\prod_{k=1}^N p(X(k)|H_0)}$$

Così il detector diventa:

$$L(X) = \frac{p(X(k)|H_1)}{p(X(k)|H_0)} \Big|_{k=i_e} = \max_k \frac{p(X(k)|H_1)}{p(X(k)|H_0)}$$

Per ottenere la forma del Multi-scale product detector, assumiamo che sotto le ipotesi  $H_1$  e  $H_0$ , il valore assoluto dei coefficienti wavelet seguono una distribuzione Log-normale. Possiamo assumere che le distribuzioni sono indipendenti tra le scale.

Per semplicità denotiamo il valore assoluto dei coefficienti  $|X_j(k)|$  con  $\dot{X}_j(k)$ . Abbiamo quindi che:

$$p(\dot{X}_j(k)|H_1) \sim LN(\mu_j^1(k), \sigma_j^2) \quad \text{e} \quad p(\dot{X}_j(k)|H_0) \sim LN(\mu_j^0(k), \sigma_j^2)$$

Dove  $\mu_j^1(k)$  e  $\mu_j^0(k)$  rappresentano il valore medio dei coefficienti alla scala  $j$  sotto le ipotesi  $H_1$  e  $H_0$  rispettivamente. Sotto le entrambe le ipotesi la varianza è  $\sigma_j^2$ .

La funzione di densità di probabilità di una distribuzione log-normale è:

$$f_Y(y; \mu, \sigma) = \frac{1}{y\sigma\sqrt{2\pi}} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \quad y > 0$$

Consideriamo i coefficienti delle scale  $2^{j_1}, \dots, 2^{j_m}$ . Abbiamo:

$$\begin{aligned} L(X) &= \max_{\{k\}} \frac{p(\dot{X}_{j_1}(k), \dots, \dot{X}_{j_m}(k)|H_1)}{p(\dot{X}_{j_1}(k), \dots, \dot{X}_{j_m}(k)|H_0)} \\ &= \max_{\{k\}} \frac{\exp(\sum_{j=j_1}^{j_m} -(\ln \dot{X}_j(k) - \mu_j^1)^2 / 2\sigma_j^2)}{\exp(\sum_{j=j_1}^{j_m} -(\ln \dot{X}_j(k) - \mu_j^0)^2 / 2\sigma_j^2)} \end{aligned}$$

$$\begin{aligned}
&= \max_{\{k\}} \prod_{j=j_1}^{j_m} (\dot{X}_j(k))^{(\mu_j^1 - \mu_j^0)/\sigma_j^2} \cdot e^{\sum_{i=j_1}^{j_m} ((\mu_j^0)^2 - (\mu_j^1)^2)/2\sigma_j^2} \\
&= \max_{\{k\}} \prod_{j=j_1}^{j_m} (\dot{X}_j(k))^{(\mu_j^1 - \mu_j^0)/\sigma_j^2} \cdot \varepsilon
\end{aligned}$$

Con:

$$\varepsilon = e^{\sum_{i=j_1}^{j_m} ((\mu_j^0)^2 - (\mu_j^1)^2)/2\sigma_j^2}$$

Se confrontiamo la funzione  $L(X)$  con il multi-scale product detector  $L_1(X)$ , per renderla in tale forma, possiamo assumere che  $(\mu_j^1 - \mu_j^0)/\sigma_j^2 = 1$  (ovvero che la varianza per scala è pari alla variazione delle medie sotto ciascuna ipotesi). Per cui la funzione test  $L(X)$  diventa:

$$L(X) = \max_{\{k\}} \prod_{j=j_1}^{j_m} (\dot{X}_j(k)) \cdot \varepsilon > \gamma$$

$L(X)$  e  $L_1(X)$  hanno la stessa forma eccetto per  $\varepsilon$  che non ha relazioni con l'osservazione  $X_j(k)$ . Con le limitazioni della varianza e della differenza delle medie, diventa un multi-scale product detector. Tale limitazione, con aggiunta l'assunzione iniziale che i coefficienti alle varie scale hanno una distribuzione log-normale, certamente non è accurata perché le distribuzioni dei coefficienti fra le scale non sono le stesse oltre al fatto che generalmente la varianza non è uguale alla differenza delle medie per scala assunte entrambi le ipotesi.

Sulla base di tale analisi, Xie et al. in [5] propongono un nuovo detector che invece tiene in considerazione la differenza delle medie e la varianza tra le varie scale:

$$L(X) = \max_{\{k\}} \prod_{j=j_1}^{j_m} (\dot{X}_j(k))^{(\mu_j^1 - \mu_j^0)/\sigma_j^2}$$

Tale detector, con l'assunzione che i coefficienti alle varie scale hanno una distribuzione log-normale, è ottimale.

## 7. RISULTATI OTTENUTI

Consideriamo una registrazione di circa 1 min acquisita in un bosco, in cui sono presenti degli spari (piuttosto distanti) di cacciatori. Si vuol individuare all'interno del segnale, gli istanti in cui sono presenti gli spari.

La registrazione acquisita è mostrata in Figura 26

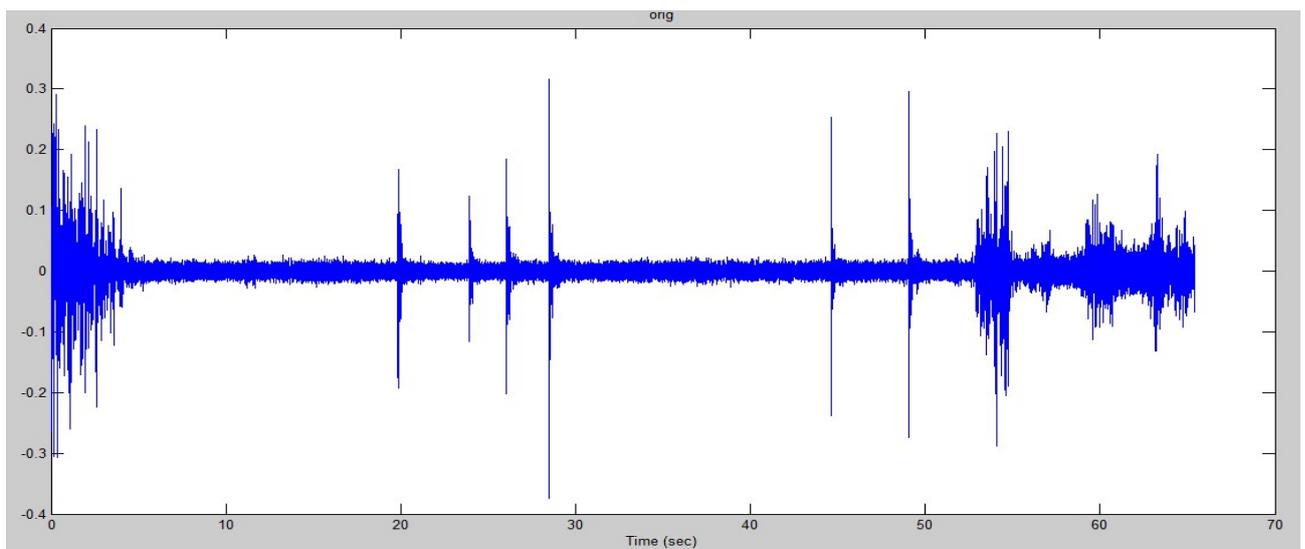


Figura 26: Segnale originale

In Figura 27 sono stati evidenziati in rosso i tratti del segnale in cui è presente uno sparo mentre in verde i tratti del segnale in cui vi è un rumore di fondo piuttosto accentuato dovuto al fruscio dei rami causato dal vento.

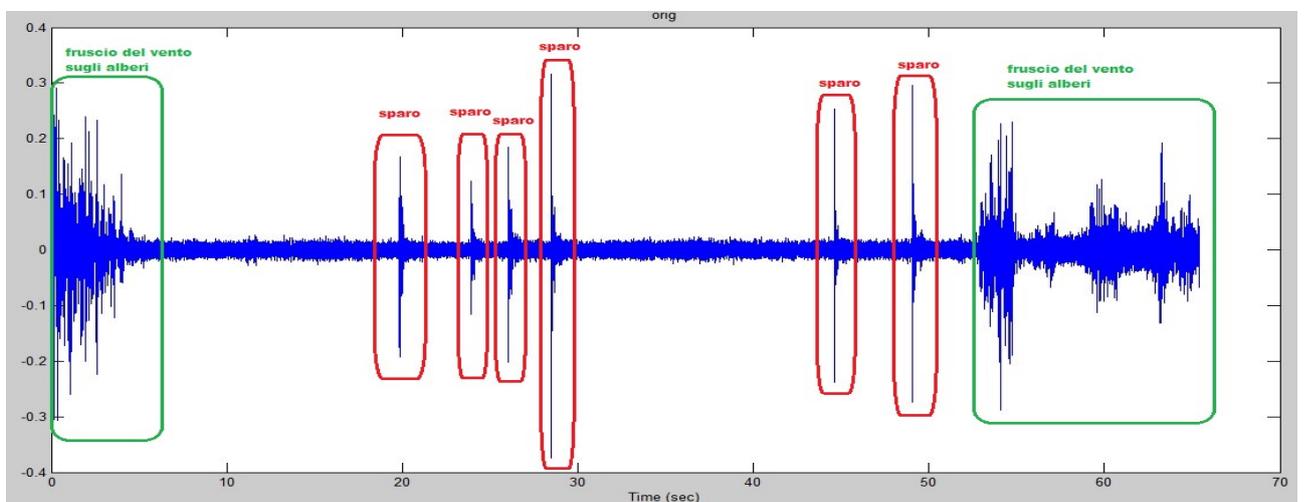


Figura 27: 6 spari e fruscio del vento evidenziati

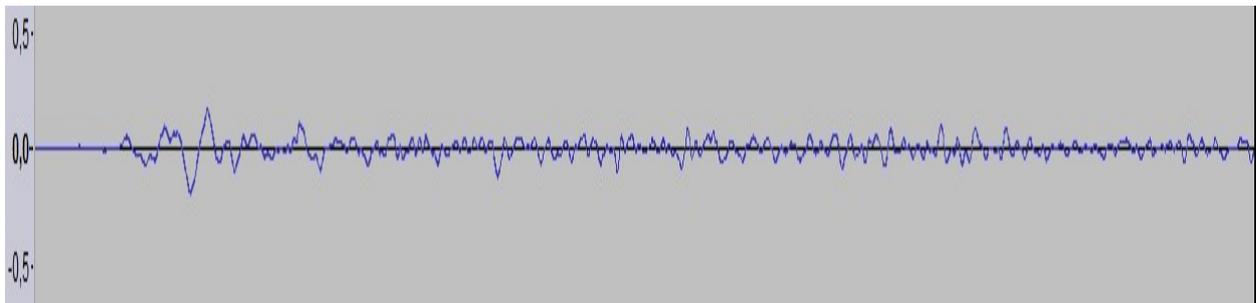
Consideriamo il seguente detector descritto in cap. 6:

$$L(X) = \max_{\{k\}} \prod_{j=j_1}^{j_m} (\hat{X}_j(k))^{(\mu_j^1 - \mu_j^0) / \sigma_j^2} > \gamma$$

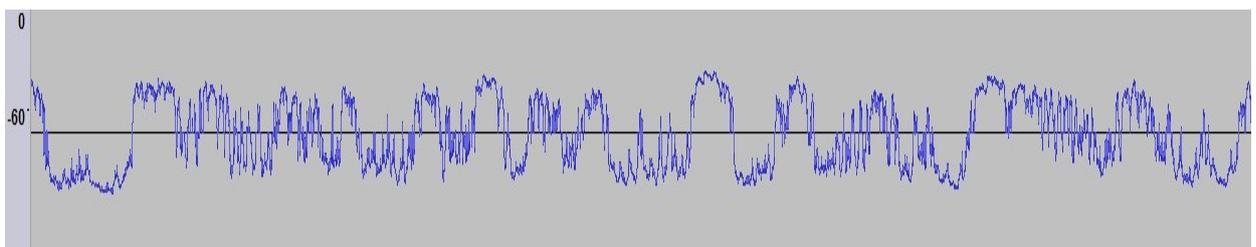
Per calcolare il suo valore, è necessario calcolare i parametri  $\mu_j^1(k)$  e  $\mu_j^0(k)$  che, ricordiamo, rappresentano il valore medio dei coefficienti alla scala  $j$  della trasformata wavelet sotto le ipotesi  $H_1$  (presenza di sparo) e  $H_0$  (assenza di sparo) rispettivamente. Sotto le entrambe le ipotesi la varianza è  $\sigma_j^2$  (possiamo prendere il valore calcolato considerando indifferentemente entrambe le ipotesi in quanto la maggior parte del segnale è costituito da rumore di fondo). Inoltre è necessario calcolare il valore del parametro  $\gamma$  necessario per eseguire il test.

A tale scopo, quindi, sono stati estratti dal segnale due segmenti: uno che contiene solo rumore di fondo e l'altro in cui vi è presente uno sparo.

In Figura 28 è mostrato uno dei tratti di segnale (in scala ingrandita) che rappresenta uno sparo mentre in Figura 29 è mostrato un tratto di segnale che rappresenta un rumore di fondo (visualizzato in db in quanto maggiormente evidenziato).

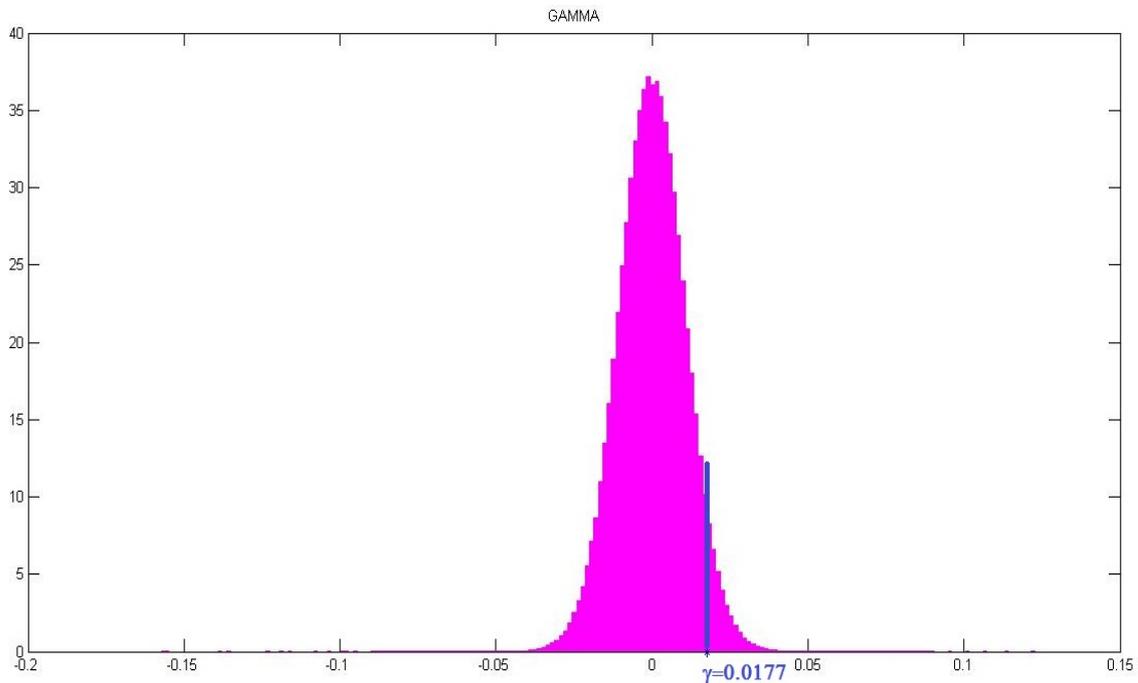


**Figura 28: sparo (distante) di un cacciatore**



**Figura 29: rumore di fondo (db)**

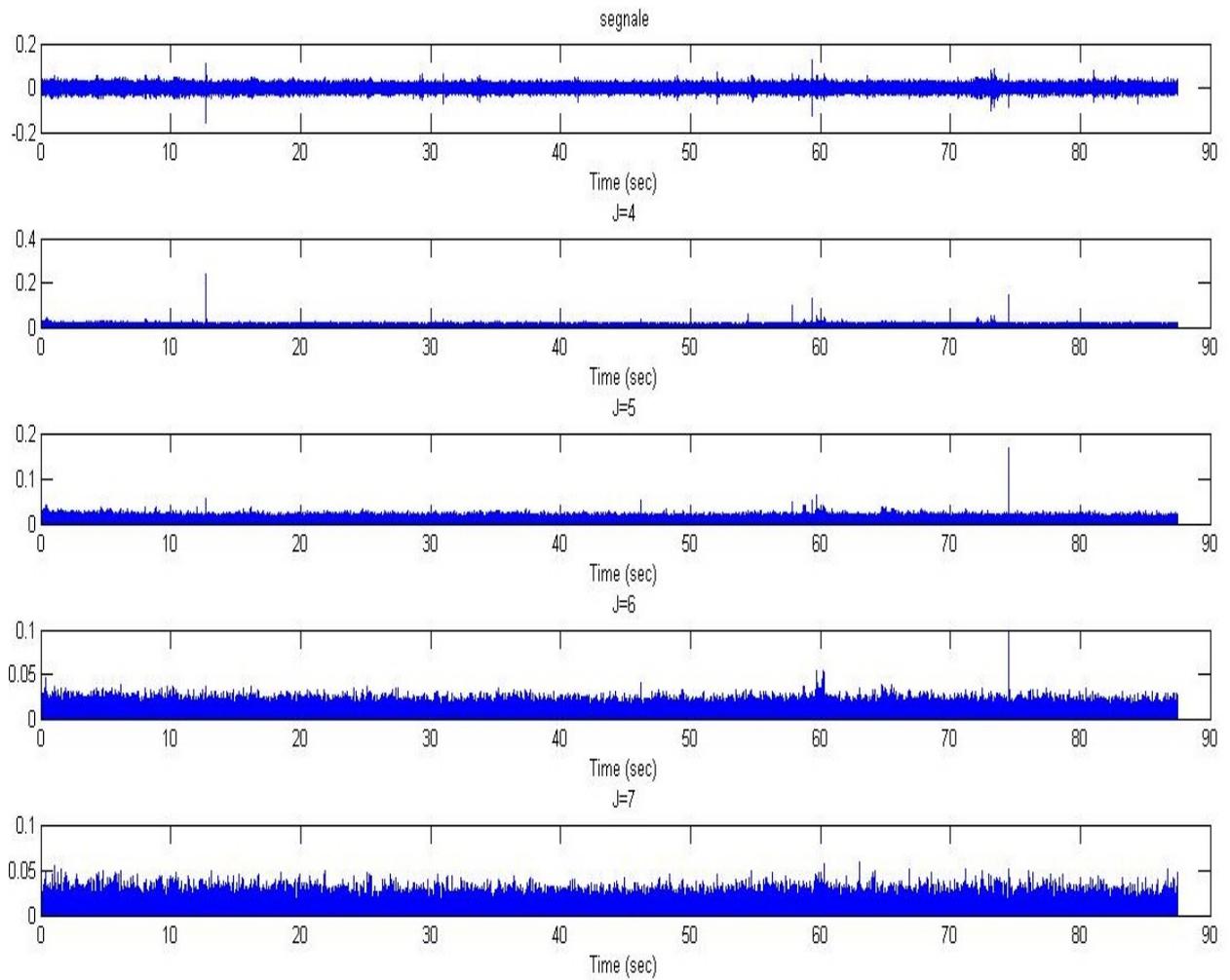
Per calcolare il parametro  $\gamma$ , si è considerata la distribuzione del segnale in assenza di sparo. Tale distribuzione, come mostrato in Figura 30, ha le caratteristiche di una distribuzione Gaussiana.



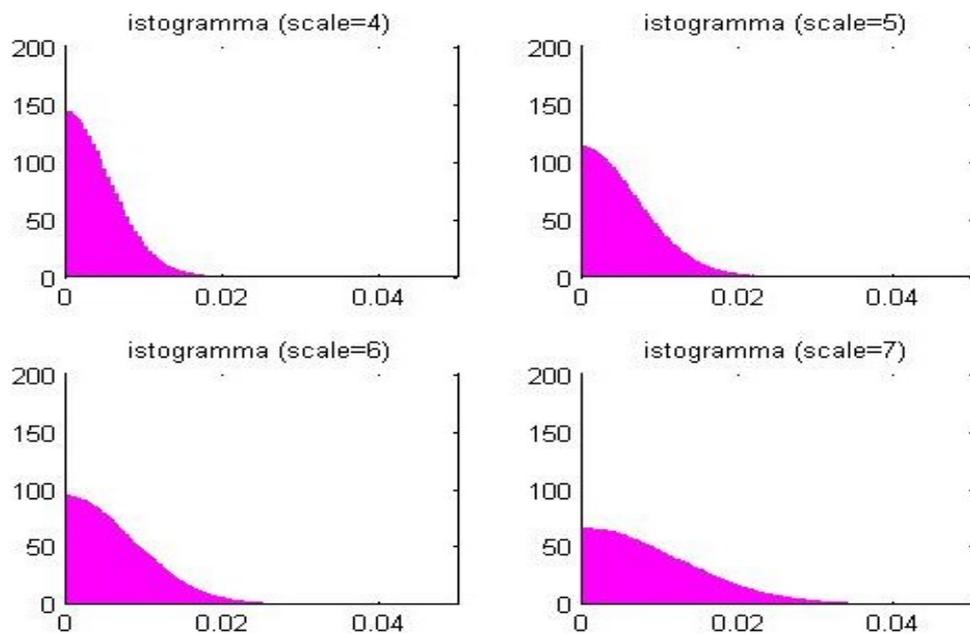
**Figura 30: distribuzione del rumore di fondo e calcolo di gamma**

In tale distribuzione, è stato individuato quel valore dell'ampiezza del segnale (asse x) tale che è superiore nel 95% dei valori di ampiezza dell'intero segnale preso come campione del rumore di fondo (Figura 30). Tale valore calcolato rappresenterà il valore  $\gamma$  della soglia utilizzata per la fase di detection.

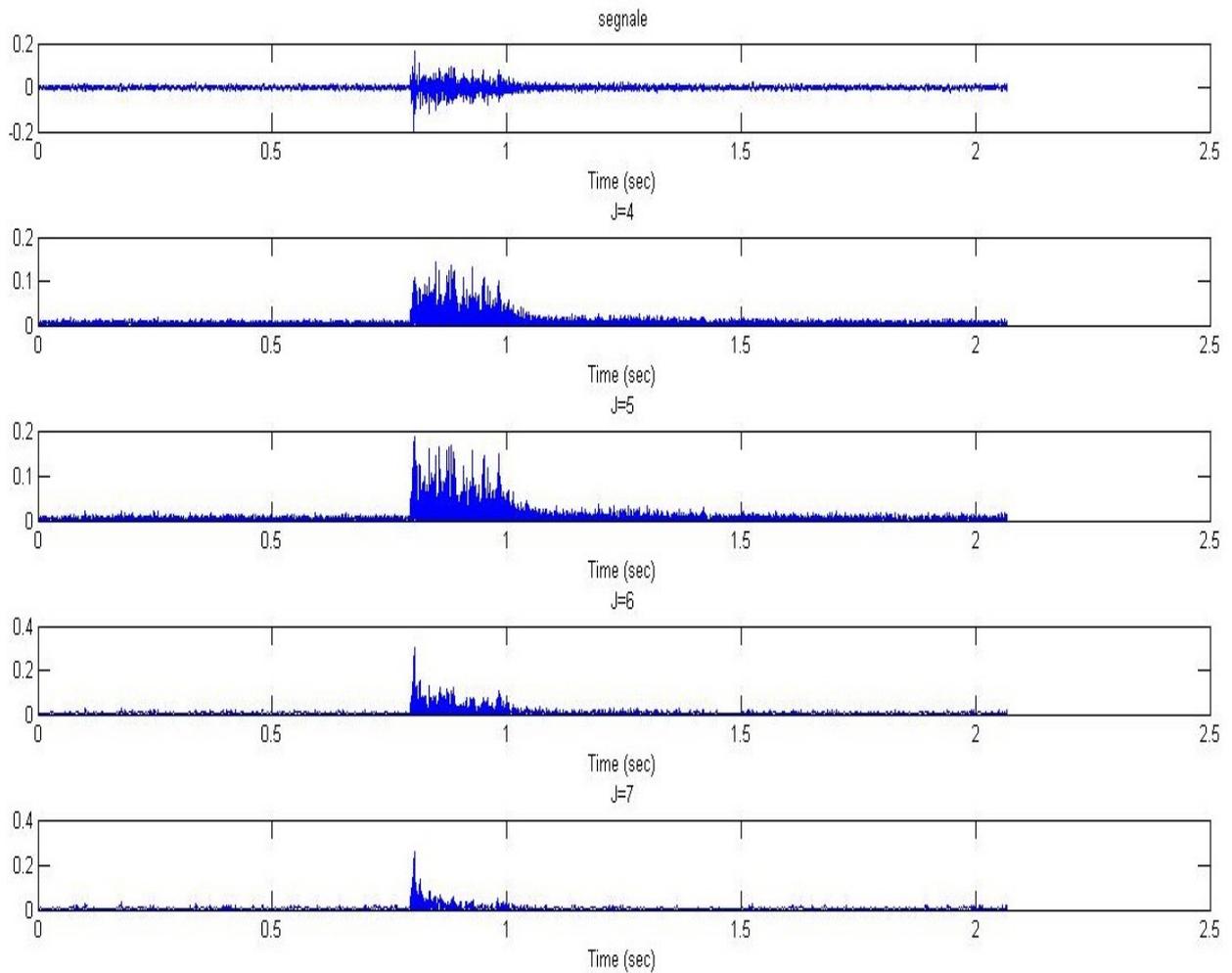
Per il calcolo dei parametri  $\mu_j^1(k)$  e  $\mu_j^0(k)$  consideriamo le trasformate wavelet secondo Mallat-Zhong, per le sole scale 4,5,6,7 ed in valore assoluto. Per la costruzione del multiscale detector, sono state scelte tali scale in quanto sono quelle in cui il segnale, di cui si vuol fare il "detect", ha i valori dei coefficienti (in valore assoluto) più alti rispetto il rumore di fondo. Per cui applicando il prodotto tra le varie scale, il rumore tende a diventare sempre più piccolo mentre i picchi relativi ai coefficienti del segnale si pronunciano sempre più (grazie anche all'allineamento in queste scale).



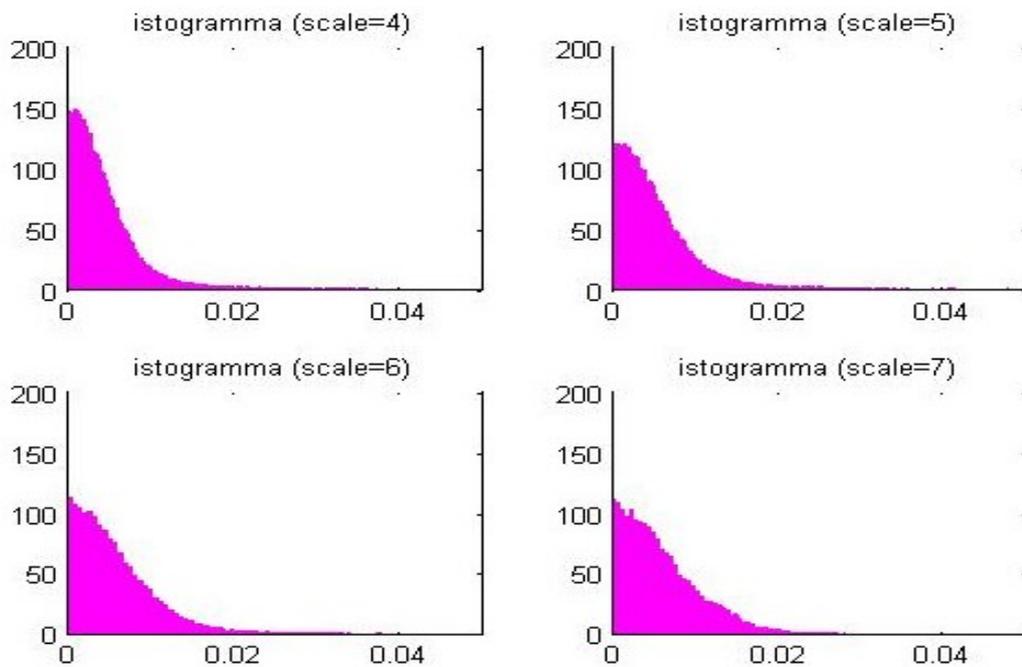
**Figura 31:  $|X_j(k)|$  del rumore di fondo per  $j=4,5,6,7$**



**Figura 32: distribuzione di  $|X_j(k)|$  del rumore di fondo per  $j=4,5,6,7$**



**Figura 33:  $|X_j(k)|$  dello sparo per  $j=4,5,6,7$**



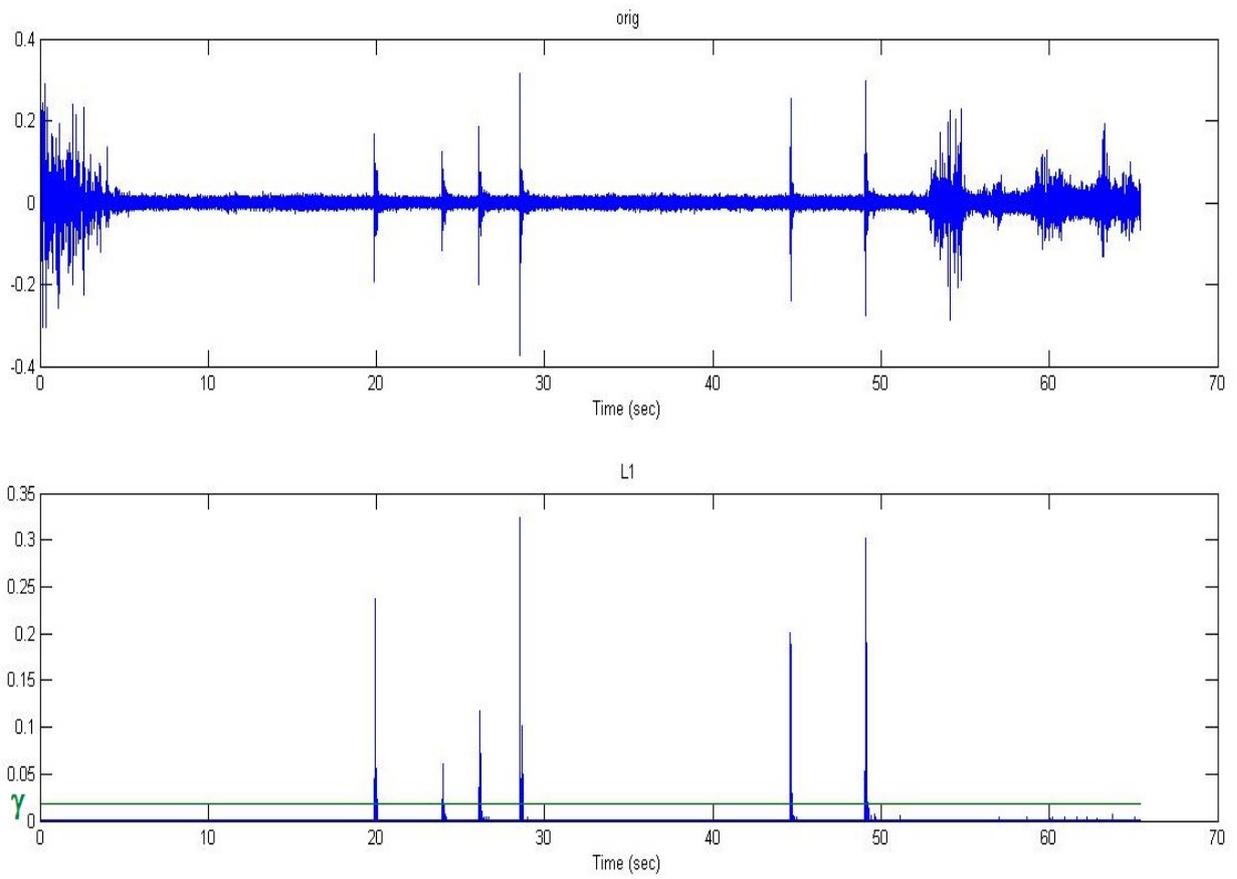
**Figura 34: distribuzione di  $|X_j(k)|$  del rumore di fondo per  $j=4,5,6,7$**

In Figura 31 e Figura 32 sono mostrati i grafici del valore assoluto dei coefficienti della trasformata wavelet (per le scale 4,5,6,7) del segnale relativo al rumore di fondo e la loro distribuzione rispettivamente. Analogamente per il segnale relativo allo sparo in Figura 33 e Figura 34.

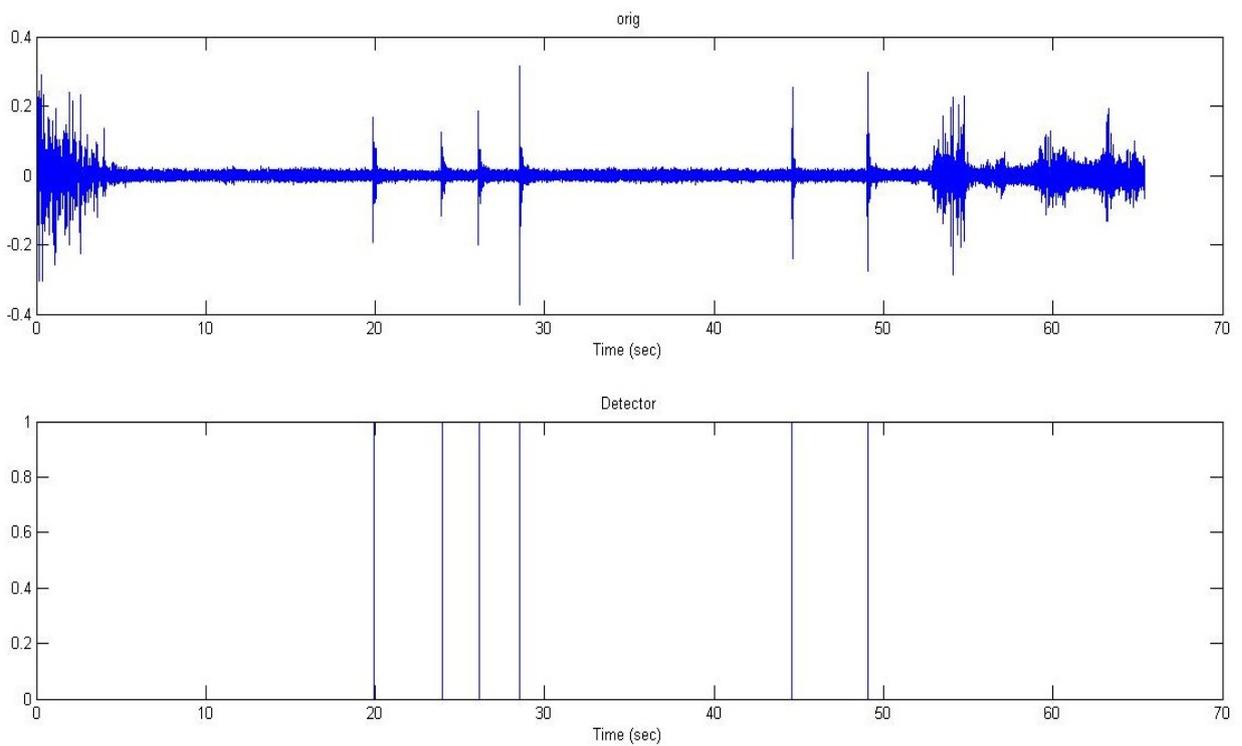
Ricordiamo che per il calcolo dei parametri  $\mu_j^1(k)$  e  $\mu_j^0(k)$ , le distribuzioni del valore assoluto  $|X_j(k)|$  vengono approssimate (“fit”) a delle distribuzioni Log-Normali (vedi par. 6), per cui sono stati ricavati, per ciascuna scala, i seguenti valori:

<b>Scala</b>	$\mu^0$	$\mu^1$	$\sigma$
4	-5.731681	-5.815268	1.284675
5	-5.518958	-5.577276	1.305151
6	-5.446998	-5.400936	1.263922
7	-5.465412	-5.048762	1.213501

A questo punto ci sono tutti gli elementi per poter applicare la formula per il detection al segnale dato, ottenendo il grafico in Figura 35. I valori assunti dalla funzione  $L(X)$  assume dei picchi nei valori negli istanti di tempo in cui lo sparo è presente mentre valori piuttosto piccoli nel resto del segnale (anche in presenza di accentuato rumore dovuto al fruscio degli alberi). Il parametro  $\gamma$  funge da soglia per i valori ottenuti per prendere solo in considerazione quelli relativi agli spari. Per questi valori viene quindi calcolato un massimo locale ottenendo il risultato in Figura 36.



**Figura 35: grafico di L(X)**



**Figura 36: output del detector: 0=assenza dello sparo 1=presenza dello sparo**

## 8. CONCLUSIONI E PROSPETTIVE FUTURE

Abbiamo visto che i coefficienti della trasformata forniscono una rappresentazione compatta che mostra la distribuzione dell'energia del segnale in tempo e frequenza. Per l'analisi dei segnali audio, la trasformata si è rivelata computazionalmente efficiente e molto robusta al rumore. Questo lo renderebbe applicabile su microcontrollori di bassa potenza e quindi a basso consumo energetico, con una conseguente praticità di utilizzo a scopo di monitoraggio in ambienti territorialmente molto estesi.

E' possibile pensare come un possibile sviluppo futuro l'utilizzo dei coefficienti wavelets, calcolati per costruire il detector, come features da usare anche per caratterizzare le classi di segnali da riconoscere, definite nel progetto SMART OPTIGRID, per poterli classificare.

Per ridurre la dimensionalità dei vettori di features estratte che ne deriverebbero dai coefficienti alle singole scale della trasformata, si possono utilizzare statistiche sul set dei coefficienti. A titolo di esempio si possono considerare [23]:

- La media del valore assoluto dei coefficienti in ogni sottobanda. Queste features forniscono informazioni sulla distribuzione in frequenza del segnale audio.
- La deviazione standard dei coefficienti in ogni sottobanda. Queste features forniscono informazioni sulla quantità di variazione della distribuzione di frequenza.
- i rapporti dei valori medi tra sottobande adiacenti. Anche queste features forniscono informazioni sulla distribuzione di frequenza.

Un ulteriore sviluppo sul progetto, come suggerito in [5], è quello di utilizzare una versione approssimata del segnale per il calcolo dei parametri statistici.

Si potrebbe estendere tale concetto anche per l'operazione di estrazione delle features utilizzate correntemente per la classificazione. SMART-OPTIGRID per la caratterizzazione delle classi definisce uno spazio vettoriale a 63 dimensioni in cui ogni vettore è un insieme di feature.

Alcune di tali features sono:

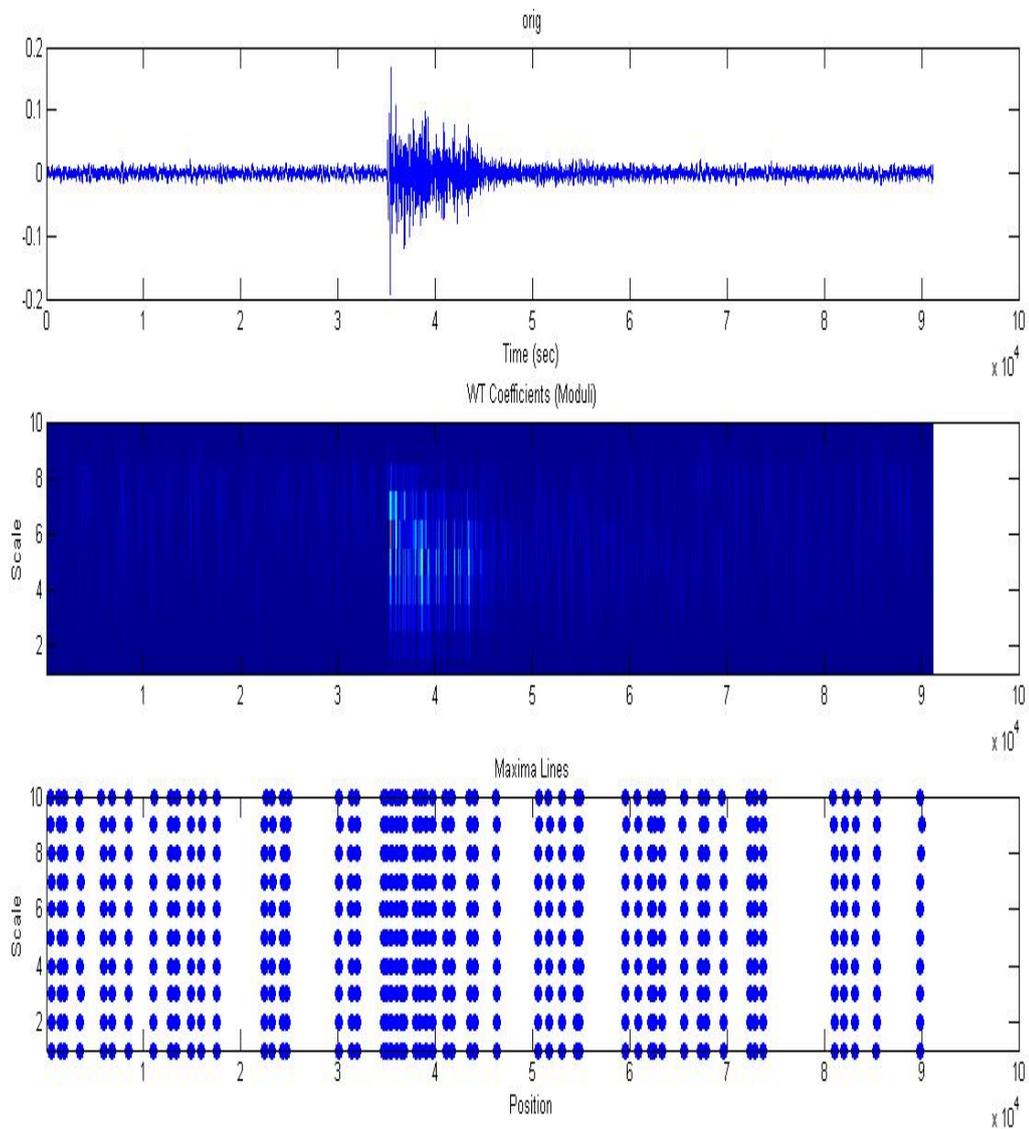
1. Zero Crossing Count
2. VZCR Variance of Zero Crossing Rate
3. HZCRR High Order Zero Crossing Rate Ratio

4. RMS Root Mean Square
5. LEF Low Energy Frame
6. STE Short Time Energy
7. LSTER Low Short Time Energy Ratio
8. VLER Variance of Low-Band Energy Ratio
9. Etc....

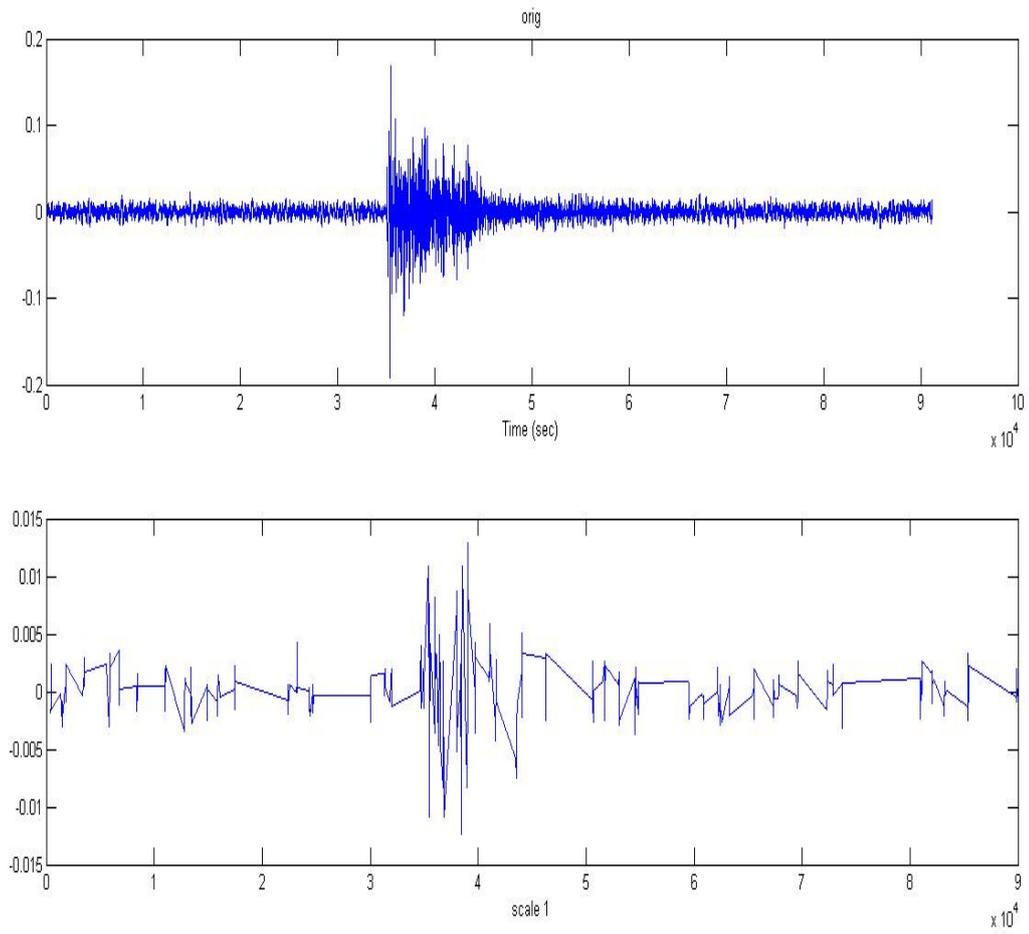
La versione approssimata del segnale, che potrebbe essere utilizzata per la loro estrazione, verrebbe ottenuta prendendo in considerazione la scala più "fine" della trasformata wavelet interpolando i punti di "localmax" dei coefficienti.

In Figura 37 viene visualizzato, a partire dall'alto verso il basso, a) un segnale audio di uno sparo, b) l'immagine ottenuta dai coefficienti della trasformata wavelet (colori chiari = valori alti dei coefficienti) c) i punti di "local max" dei coefficienti per ciascuna scala.

Per dare un'idea su come sia il segnale approssimato, in Figura 38, il segnale viene approssimato semplicemente unendo i punti di "local max" dei coefficienti della scala  $s=1$ .



**Figura 37: a) segnale audio b) trasformata wavelet c) local max**



**Figura 38 esempio di approssimazione del segnale (senza interpolazione)**

## 9. BIBLIOGRAFIA

- [1] Brian M Sadler, Tien Pham, and Laurel C Sadler, "Optimal and wavelet-based shock wave detection and estimation," *The Journal of the Acoustical Society of America*, vol. 104, no. 2, pp. 955–963, 1998.
- [2] Stephane Mallat and Sifen Zhong, "Characterization of signals from multiscale edges," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no.7, pp. 710–732, 1992.
- [3] A. Rosenfeld and M. Thurston, "Edge and curve detection for visual scene analysis," *IEEE Trans. Comput.* 20, 562–569 (1971).
- [4] Y. Xu, J. Weaver, D. Healy, and J. Lu, "Wavelet transform domain filters: A spatially selective noise filtration technique," *IEEE Trans. Image Process.* 3, 747–758 (1994).
- [5] W. Xie, M. Bao, Xi. Li, X. Zhang "an improved wavelet based shockwave detector" (2013)
- [6] R. Maher, "Modeling and signal processing of acoustic gunshot recordings," in *Proc. 2nd Signal Process. Educ. Workshop, 4th Digital Signal Process. Workshop, 2006*, pp. 257–261
- [7] R. Maher, "Acoustical characterization of gunshots," in *Proc. IEEE Workshop Signal Process. Appl. Public Security Forensics, SAFE, 2007*, pp. 1–5.
- [8] P.Weissler and M. T. Kobal, "Noise of police firearms," *J. Acoust. Soc. Amer.*, vol. 56, no. 5, pp. 1515–1522, Nov. 1974.
- [9] R. Stoughton, "Measurements of small-caliber ballistic shock waves in air," *J. Acoust. Soc. Amer.*, vol. 102, no. 2, pp. 781–787, Aug. 1997
- [10] A. Dufaux, L. Besacier, M. Ansorge, and F. Pellandini, "Automatic sound detection and recognitions for noisy environment," in *Proc. X EUSIPCO, 1997*, p. 1889
- [11] K. Molnar, A. Lédeczi, L. Sujbert, and G. Peceli, "Muzzle blast detection via short time Fourier transform," in *Proc. 12th PhD Mini-Symp.*, 2005, pp. 16–17
- [12] B. Sadler, L. Sadler, and T. Pham, "Optimal and robust shockwave detection and estimation," in *Proc. IEEE ICASSP, 1997*, vol. 3, pp. 1889–1892
- [13] B. Ferguson, L. Criswick, and K. Lo, "Locating far-field impulsive sound sources in air by triangulation," *J. Acoust. Soc. Amer.*, vol. 111, no. 1, pt. 1, pp. 104–116, Jan. 2002.
- [14] G. Duckworth, J. Barger, S. Carlson, D. Gilbert, M. Knack, J. Korn, and R. Mullen, "Fixed and wearable acoustic counter-sniper systems for law enforcement," in *Proc. SPIE Int. Symp. Enabling Technologies for Law*

Enforcement and Security Sensors, C3I, Information, and Training Technologies for Law Enforcement, 1998, pp. 3575–3577.

- [15] A. Lédeczi, P. Volgyesi, M. Maroti, G. Simon, G. Balogh, A. Nodas, B. Kusy, S. Dora, and G. Pap, "Multiple simultaneous acoustic source localization in urban terrain," in Proc. 4th Int. Symp. Inf. Process. Sensor Netw., Los Angeles, CA, 2005, p. 69.
- [16] D. Istrate, E. Castelli, M. Vacher, L. Besacier, and J. F. Serignat, "Information extraction from sound for medical telemonitoring," IEEE Trans. Inf. Technol. Biomed., vol. 10, no. 2, pp. 264–274, Apr. 2006
- [17] A. I. T. Fernández, "Propagación del sonido en bosques. Análisis comparative de las medidas in situ, en laboratorio y de los valores predichos por un modelo," Ph.D. dissertation, Universidad de Valladolid, Valladolid, España, 2002
- [18] Van Trees, H.L.: Optimum Array Processing. Wiley-Interscience (2002)
- [19] M. Brandstein and D. Ward, editors. Microphone Arrays: Signal Processing Techniques and Applications. Springer, 2001.
- [20] B.D. Van Veen and K.M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," IEEE ASSP Magazine, Vol. 5, No. 2, Apr. 1988, pp. 4-24.
- [21] M. Basseville and I. V. Nikiforov, Detection of Abrupt Changes, Theory and Application ~Prentice-Hall, Englewood Cliffs, NJ, 1993.
- [22] D. Marr and E. Hildreth, "Theory of edge detection," Proc. Royal Sot.London. vol. 207. DD. 187-217. 1980
- [23] Tzanetakis, G., G.Essl, and P. R. Cook. "Audio Analysis Using the Discrete Wavelet Transform," In Proceedings of of WSES International Conference, Acoustics and Music: Theory and Applications (AMTA), Skiathos, Greece, 2001.
- [24] S. Mallat W. Liang Hwang, "Singularity detection and processing with wavelets" IEEE Transactions on information theory, vol.38, no. 2, march 1992
- [25] J. Canny, "A computational approach to edge detection," IEEE Trans. Putt. Anal. Machine Intell., vol. PAMI-X, pp. 679-698, 1986.