
**UNIVERSITÀ DEGLI STUDI DI
BOLOGNA**

Facoltà di Ingegneria

Corso di Laurea in Ingegneria Elettronica

Tesi di Laurea
**Trasformata di Hilbert-Huang per la
rilevazione di segnali non stazionari
nel progetto SETI**

Relatore:
prof. Tullio Salmon Cinotti

Presentata da:
Leonardo Amico
Anno accademico 2007-2008

Incontrare è fondamentale,
noi siamo interamente determinati
non da ciò che siamo,
ma da ciò che incontriamo.

ALAIN BADIOU

Ringraziamenti

Vorrei ringraziare profondamente quanti, in un modo o nell'altro, hanno reso possibile questa tesi. Innanzitutto l'Ing. Stelio Montebugnoli e tutte le persone al Radiotelescopio di Medicina che mi hanno aiutato e seguito, oltre che avermi gentilmente ospitato nel loro osservatorio. Poi il professore Tullio Salmon Cinotti per esser stato così disponibile. Ringrazio Mattia Tani, con cui ho collaborato per i lavori preliminari della tesi e tutti i miei amici che mi hanno dato una mano, in particolare Francesco, Julia, Marcello, Federico, Gabriel e Filippo. Infine ringrazio la mia famiglia, senza la quale non sarei arrivato a questo punto.

Grazie.

Indice

1-Introduzione.....	6
2-Diversi tipi di segnale e tecniche di analisi.....	10
2.1 L'inefficienza dell'analisi di Fourier per processi non lineari e non stazionari.....	10
2.2 Alcune tecniche alternative per l'analisi di processi non stazionari e non lineari.....	12
2.2.1 Spettrogramma.....	12
2.2.2 KLT.....	12
3-La trasformata di Hilbert-Huang.....	14
3.1 La frequenza istantanea.....	14
3.2 Intrinsic Mode Functions.....	18
3.2.1 La normalizzazione delle IMF.....	18
3.3 Empirical Mode Decomposition.....	20
4-Algorithmo al calcolatore e test sull'HHT.....	25
4.1 Implementazione dell'EMD.....	25
4.1.1 Problemi legati alla spline e ai bordi.....	28
4.1.2 Sifting Ulteriori e convergenza.....	31
4.2 Implementazione della trasformata di Hilbert e calcolo della frequenza istantanea.....	31
4.2.1 Valori di frequenza oltre Nyquist.....	34
4.2.2 Normalizzazione.....	35

4.2.3 Problemi introdotti dal teorema di Nuttal.....	36
4.3 Hilbert Spectrum.....	37
4.4 Test e confronto tra HHT e Spettrogramma.....	39
4.4.1 Segnale sintetizzato e rumorosità.....	39
4.4.2 Test su segnali rilevati dal radio osservatorio.....	44
5-Conclusioni.....	47
Bibliografia.....	49

1-Introduzione

Sin dal 1895, quando Guglielmo Marconi trasmise il primo segnale radio nella storia sono state inviate nello spazio onde elettromagnetiche a diverse frequenze e potenze sempre maggiori con l'aumentare della tecnologia. Oggi il nostro sistema di comunicazione è interamente basato su trasmissione e ricezione di onde elettromagnetiche. Si è partiti da segnali radio fino a trasmissioni televisive, cellulari e i moderni sistemi a banda larga.

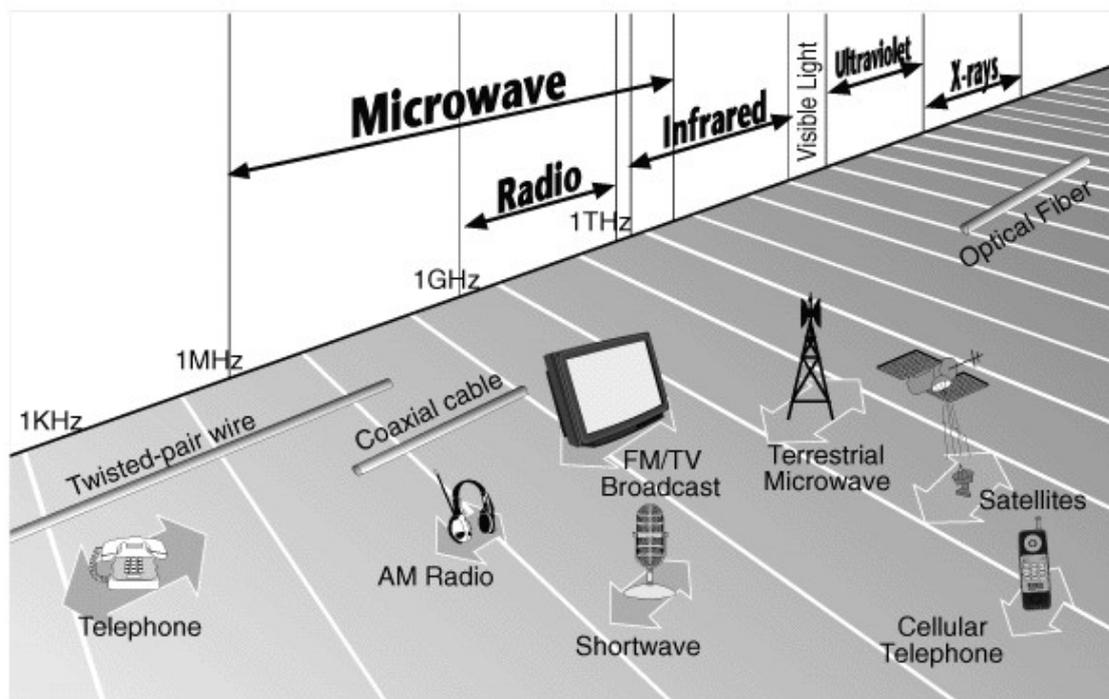


Illustrazione 1.1: Spettro EM e utilizzo delle bande di frequenza. *Informit.com*

Alcune porzioni dello spettro, non utilizzate dalle comunicazioni terrestri, sono riservate alla radioastronomia, come lo studio di corpi celesti non rilevabili con telescopi ottici (pulsar, quasar, radio galassie) oppure il monitoraggio di detriti spaziali (satelliti non più funzionanti,

parti di razzi..) pericolosi per il rischio di collisione con altri oggetti o infine la bioastronomia ovvero lo studio di vita nell'universo.

Il progetto SETI (acronimo di Search for Extra Terrestrial Intelligence) nacque nel 1959 ad opera di due fisici Cocconi e Morrison i quali per primi misero in evidenza il potenziale rappresentato dall'uso delle radioonde per la comunicazione interstellare. Qualche mese più tardi l'astronomo della Cornell University Frank Drake diede luogo al primo moderno esperimento SETI con il progetto "Ozma". Egli puntò il telescopio Tatal, 26 metri di diametro situato a Green Bank in West Virginia, sulla frequenza di 1.42 Ghz verso le stelle Tau Ceti ed Epsilon Eridani: le due più vicine alla terra, fra quelle di tipo solare e con probabili pianeti. L'esperimento non portò a nessun risultato, ma da quel giorno la pratica di sondare lo spazio in cerca di eventuali segnali provenienti da radio-sorgenti di origine "intelligente" venne applicata più volte. Nel 1971 la NASA fondò un suo progetto SETI coinvolgendo lo stesso Drake e la società di calcolatori Hewlett-Packard. Si progettò di costruire un enorme radiotelescopio dedicato alla ricerca bioastronomica, che non venne mai fatto, ma gli studi effettuati in quell'occasione fondarono la base del SETI per gli anni successivi. Nel 1974 dall'osservatorio di Arceibo, in Puerto Rico fu mandato, molto simbolicamente, un messaggio verso altre civiltà. L'unica volta che la ricerca diede qualche frutto fu nel 1977. Il Dr, Jerry R. Ehman, lavorando come volontario al progetto SETI al radio telescopio dell'università dell'Ohio, testimoniò la ricezione di un segnale di natura incerta. Un segnale narrow band di 72 secondi, di potenziale origine non-terrestre e non-solare che però non venne più ricevuto. Oggi il governo degli Stati Uniti e la NASA non si interessano più del progetto, che viene finanziato principalmente da privati o portato avanti da volontari nei radio osservatori. Questo è il caso del programma SETI-Italia, che grazie all'utilizzo in parallelo ad altre attività dei radiotelescopi di Medicina(BO) riesce a continuare l'indagine dello spazio in cerca di segnali extraterrestri.



Illustrazione 1.2: I due radiotelescolpi dell'osservatorio di Medicina(BO):la Parabola da 32 metri(in alto) e la Croce del Nord(in basso).

L'idea originaria, che ha dettato la linea teorica di tutto il progetto SETI fino a pochi anni fa, consisteva nella ricerca di segnali monocromatici in precise frequenze. Monocromatici (ovvero sinusoidi pure) perché un segnale di questo tipo è quasi sicuramente artificiale, e per quanto riguarda la banda in cui cercare il segnale, un'ipotesi fatta fu intorno alla frequenza 1.42GHz. Questa corrisponde all'emissione spontanea di radiazione degli atomi di idrogeno neutro interstellare. Tale elemento è quello più diffuso nell'universo: si ipotizza quindi che l'eventuale civiltà aliena sufficientemente avanzata conoscesse questa frequenza, e la si considerò come la più adatta per la comunicazione interstellare. Inoltre il sistema di analisi di dati più usato (allora come oggi, la trasformata di

Fourier) è particolarmente efficiente nella rilevazione di segnali monocromatici, e questo fu un ulteriore motivo per cui per i 30 anni successivi la ricerca SETI non subì sostanziali variazioni.

Più recentemente Biraud avanzò dubbi circa le linee guida della ricerca SETI. Egli argomentava che nella nostra terra i sistemi di telecomunicazioni terrestri vanno sempre più verso l'utilizzo di wide-band(banda larga) piuttosto che narrow-band(banda stretta) quindi, operando un semplice parallelismo, ricercare segnali extraterrestri supponendoli limitati intorno a specifiche frequenze è una forzatura. Considerazioni simili possono venir fatte dai risultati sui recenti studi sui sistemi di comunicazione di mammiferi marini. Mentre l'analisi di Fourier ha mostrato i suoi limiti dimostrandosi inadatta allo studio di questi segnali di breve durata e di larghe bande di frequenza, altre trasformate non soggette ai vincoli di stazionarietà e linearità hanno dato invece buoni risultati. Un altro caso in cui la trasformata di Fourier si è dimostrata inadatta è nel campo delle neuroscienze, dove ancora una volta ricorrere a metodi di analisi alternativi può essere determinante. Tutti diversi esempi di scambio di informazioni che portano a concludere che ogni ipotesi a priori sul sistema di comunicazione di una civiltà aliena sconosciuta sia non consistente. E quindi che il migliore approccio per la ricerca SETI è utilizzare tecniche di analisi del segnale valide sotto condizioni più generali possibili.

La Hilbert-Huang Trasform, al contrario di quasi tutti i sistemi di analisi di segnale oggi disponibili, risulta valida anche per sistemi non stazionari e non lineari. Queste potenzialità la rendono particolarmente adatta alle esigenze dell'analisi SETI ed è del suo utilizzo per questo progetto che tratta la tesi presentata.

2-Diversi tipi di segnale e tecniche di analisi

2.1 L'inefficienza dell'analisi di Fourier per processi non lineari e non stazionari

Lo strumento privilegiato per l'analisi dei dati, sin dalla sua introduzione, è l'analisi spettrale di Fourier, la quale fornisce una rappresentazione del segnale come distribuzione di energia su frequenza. Questa predominanza è dovuta sia alla semplicità che all'efficienza della trasformata di Fourier, grazie anche all'algoritmo FFT che permette di effettuare tale analisi con ancora maggiore velocità.

Nonostante questo metodo sia valido sotto condizioni estremamente generali ci sono alcune restrizioni cruciali. Il sistema che genera i dati dev'essere lineare e i dati analizzati devono essere strettamente periodici o stazionari, oltre che il segnale debba essere scomponibile come somma di sinusoidi.

Un segnale temporale $X(t)$ è definito stazionario in senso lato se verifica le seguenti condizioni:

$$E(|X(t)|^2) < \infty$$

$$E(X(t)) = m$$

$$C(X(t_1), X(t_2)) = C(X(t_1 + \tau), X(t_2 + \tau)) = C(t_1 - t_2)$$

dove $E(X(t))$ è il valore atteso e $C(t)$ la funzione di covarianza.

Raramente segnali (sia provenienti da fenomeni naturali che da sorgenti artificiali) soddisfano queste condizioni, essendo spesso risultato di modulazioni di frequenza o ampiezza.

Oltre alla stazionarietà, l'analisi spettrale di Fourier richiede anche che il segnale sia lineare. Altra condizione che difficilmente è rispettata dei sistemi analizzati, tanto che per la maggior parte dei casi si fanno approssimazioni per poter applicare metodi di analisi che appunto richiedono questa caratteristica. Inoltre i sistemi di acquisizione dati possono introdurre non linearità nel rilevare un fenomeno che invece lo è, complicando ulteriormente cose.

Per queste ragioni il campo d'applicazione dell'analisi di Fourier è limitato. E il suo utilizzo incondizionato per processi che non soddisfano le condizioni di linearità e stazionarietà può dare risultati fuorvianti. Ad esempio l'introduzione di molte componenti armoniche per rappresentare set dati non stazionari, quindi non uniformi globalmente. Come risultato l'energia si distribuisce in una banda molto ampia, per poter rispettare correttamente il principio di conservazione dell'energia. Una rappresentazione di questo tipo risulta corretta matematicamente, ma la sua informazione è inutile dal punto di vista fisico. Altri problemi si incontrano nell'analisi di sistemi non lineari. Tipicamente si hanno forme d'onda distorte quindi per rappresentarle l'analisi di Fourier introduce armoniche spurie che ancora una volta possono disperdere su un ampio spettro di frequenze l'energia del segnale.

2.2 Alcune tecniche alternative per l'analisi di processi non stazionari e non lineari

2.2.1 Spettrogramma

Questo è il metodo più elementare, consiste nel fare un'analisi di Fourier in una finestra limitata temporalmente e scorrere lungo tutto l'asse del tempo. Con questo metodo si produce una rappresentazione tempo-frequenza di tutto il set dati. Essendo però basata sull'analisi di Fourier, restano valide le restrizioni già presentate. Devono essere rispettate le condizioni di stazionarietà e linearità entro la finestra temporale. La scelta delle dimensioni della finestra diventa quindi determinante. Anche se assumiamo localmente stazionario il set dati analizzato, non sappiamo se le dimensioni scelte coincidano sempre con l'intervallo di stazionarietà del segnale, che può essere variabile.

2.2.2 KLT

La trasformata di Karhunen-Loeve è un nuovo strumento e ha dato ottimi risultati nel campo della compressione di dati, inoltre recentemente sono stati fatti studi sul suo utilizzo nell'ambito della radioastronomia e in particolare nel progetto SETI. Questa tecnica è basata su una rappresentazione del segnale come combinazione lineare di una base ortogonale, come per l'analisi di Fourier.

$$X(t) = \sum_{k=1}^{\infty} Z_k f_k(t)$$

dove le $f_k(t)$ sono le autofunzioni ortogonali che risolvono la seguente equazione

$$\int_0^T R_X(t_1, t_2) f_k(t_2) dt_2 = \lambda_k f_k(t_1)$$

con

$R_X(t_1, t_2)$ funzione di autocorrelazione di $X(t)$.

Quindi a differenza dell'analisi di Fourier in cui è predefinita la base mediante cui sviluppare il segnale (le funzioni trigonometriche) si ha a che fare con una base del tutto adattativa, calcolata a posteriori. Sono così scavalcate le restrizioni imposte dall'analisi di Fourier sulla stazionarietà e linearità, essendo il set dati stesso che determina qual'è la forma d'onda più adatta a rappresentare il segnale analizzato.

Nonostante i problemi iniziali sull'uso di questo strumento relativi al suo elevato costo computazionale siano parzialmente venuti meno con l'avvento di calcolatori sempre più potenti, restano alcuni dubbi sulle applicazioni della KLT nell'ambito della radioastronomia.

L'autocorrelazione rappresenta in termini numerici il grado di verosimiglianza tra un vettore e una sua versione shiftata. Un segnale casuale per sua natura presenta un basso grado di correlazione con se stesso ed è su questa osservazione che si basano le virtù della trasformata. Essa scompone il segnale mediante delle basi ortonormali che rappresentano diversi modi di correlazione indipendenti del segnale. Soltanto gli autovettori maggiori contengono informazione mentre gli altri sono generati soltanto per rappresentare le componenti rumorose, quindi casuali del segnale. Questo fa della KLT un buon metodo di rilevazione di segnali immersi nel rumore. Quando un autovettore è sensibilmente più grande degli altri allora il segnale rumoroso contiene dell'informazione, altrimenti no. Se invece si vuole avere un'informazione qualitativa sul segnale questo metodo di analisi non è appropriato. Gli autovettori non rappresentano caratteristiche fisiche del segnale, e questa analisi non fornisce informazioni di frequenza che possano essere utilizzate per localizzare la distribuzione di energia.

3-La trasformata di Hilbert-Huang

Il metodo di analisi proposto nel 1998 da Norden E. Huang, fornisce una rappresentazione tempo-frequenza del segnale, che garantisce quindi significato fisico al risultato e allo stesso momento la sua natura adattativa evita le restrizioni di stazionarietà e linearità imposte dalla Trasformata di Fourier. La HHT è composta da due fasi, per primo il segnale è processato tramite l'EMD (Empirical Mode Decomposition) che restituisce una collezione di IMF (Intrinsic Mode Function). Successivamente di ogni IMF viene calcolata la frequenza istantanea tramite la trasformata di Hilbert.

3.1 La frequenza istantanea

Il concetto di frequenza istantanea, cruciale per la HHT, è ancora fortemente controverso in particolare per due motivi. Innanzitutto la definizione di frequenza è molto influenzata dall'analisi spettrale di Fourier, per cui essa rappresenta l'inverso del periodo delle sinusoidi ad ampiezza costante con cui decompongo il segnale studiato. Quindi è necessaria un'oscillazione completa della sinusoide per una definizione locale di frequenza, che perciò non potrà identificare le variazioni istantanee tipiche di un segnale non stazionario. Il secondo problema deriva dal modo non unico con cui è possibile definire la frequenza istantanea. Difficoltà quasi completamente superata se si rende analitico il segnale mediante la trasformata di Hilbert.

Per una serie temporale $X(t)$ di classe L^p la trasformata di Hilbert $Y(t)$ è definita come:

$$Y(t) = \frac{1}{\pi} \text{P.V.} \int_{-\infty}^{\infty} \frac{X(\tau)}{t-\tau} d\tau$$

Ovvero la convoluzione tra $X(t)$ e $\frac{1}{t}$, dove P.V. indica che l'integrale va inteso come valore principale di Cauchy, dato che non è definito ovunque. Quindi ora è possibile scrivere il segnale analitico $Z(t)$ come:

$$Z(t) = X(t) + iY(t)$$

dove

$$a(t) = \sqrt{X(t)^2 + Y(t)^2}, \theta(t) = \arctan\left(\frac{Y(t)}{X(t)}\right).$$

A questo punto, dato che la trasformata di Hilbert fornisce un unico modo di definire la parte immaginaria del segnale analitico, la frequenza istantanea risulta:

$$f = \frac{1}{2\pi} \frac{d\theta(t)}{dt}.$$

Ma restano ancora alcune controversie relative al significato di questa grandezza. Senza approfondire oltre Cohen suggerì che tale definizione fosse valida soltanto per specifici segnali, che lui chiamò monocomponenti, ovvero con un' unica frequenza istantanea per ogni istante di tempo. Per illustrare il problema consideriamo ad esempio la funzione

$$X(t) = \sin(t)$$

la sua trasformata di Hilbert è $Y(t) = \cos(t)$ e nel piano di Gauss il segnale analitico è una circonferenza unitaria con centro nell'origine (figura 3.1a). La sua fase quindi è una retta e la sua derivata, la frequenza istantanea, una costante. Se invece aggiungiamo una componente continua α alla sinusoidale

$$X(t) = \sin(t) + \alpha$$

il grafico nel piano complesso sarà ancora una circonferenza, ma traslata nell'asse reale come mostrato ancora in figura 3.1a. Se $\alpha < 1$

l'origine sarà ancora dentro la circonferenza ma, come mostrato nelle figure 3.1b, 3.1c, la fase e la frequenza istantanea sono molto differenti dal caso precedente. Se invece si ha $\alpha > 1$ l'origine è esterna alla circonferenza, quindi il vettore che descrive il segnale analitico cambia verso di rotazione per percorrere la circonferenza con il risultato che si hanno valori di frequenza negativi, senza senso fisico.

Quindi generalizzando il ragionamento possiamo dedurre che ogni volta che abbiamo a che fare con segnali non simmetrici localmente non possiamo definire la frequenza istantanea correttamente. Si pensi ad esempio a una somma di due sinusoidi con frequenza diversa: si avrebbe l'oscillazione dell'onda a frequenza maggiore lungo la sinusoide a frequenza minore, con il risultato di una variazione nel tempo della media del segnale risultante. Si ricadrebbe in casi analoghi a quelli con $\alpha \neq 0$. Queste restrizioni locali suggeriscono un modo di definire una nuova classe di funzioni, indicate come

IMF (Intrinsic Mode Functions) per le quali la frequenza istantanea è sempre fisicamente accettabile.

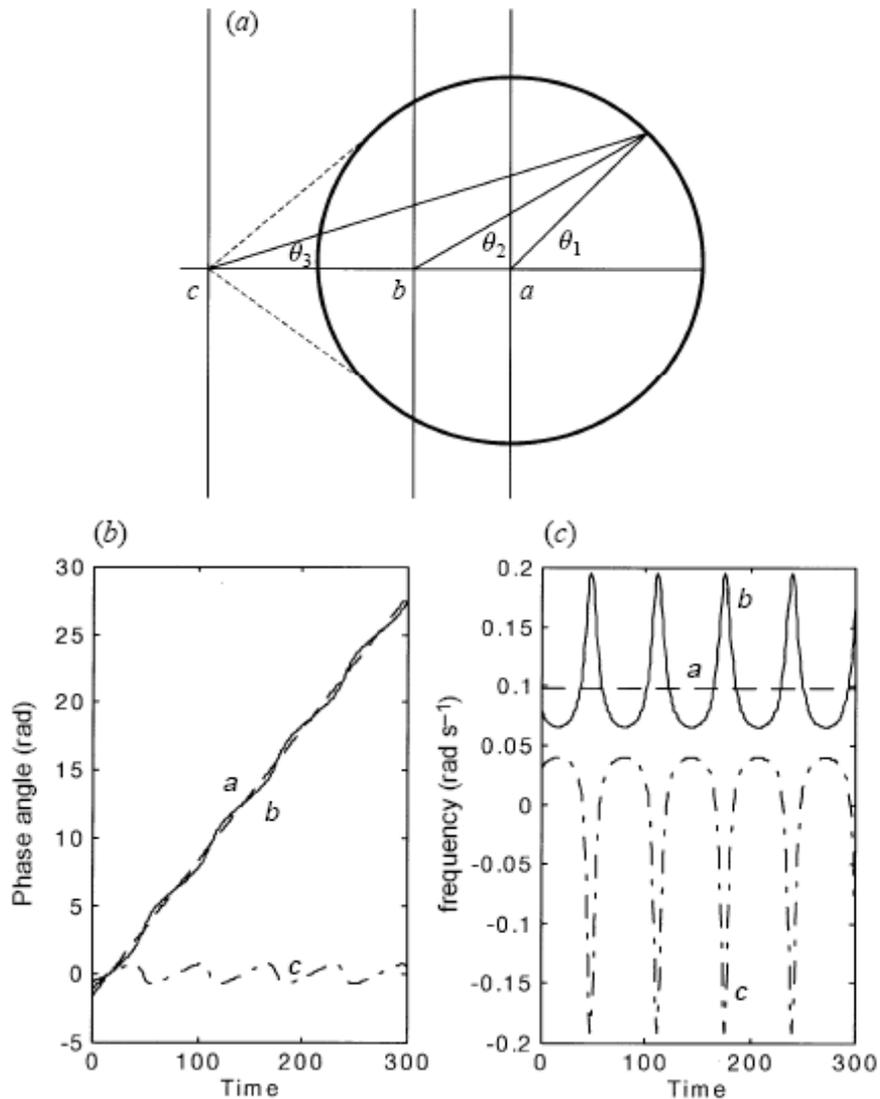


Illustrazione 3.1: Interpretazione della frequenza istantanea della funzione $X(t)=\sin(t)+\alpha$. (a) Piano di gauss con assi traslati. a, per $\alpha=0$; b, per $\alpha<1$, c per $\alpha>1$. (b) grafico della fase $\theta(t)$. (c) grafico della frequenza istantanea.

3.2 Intrinsic Mode Functions

Huang nel definire la classe di segnali IMF propone due condizioni necessarie:

- che siano simmetrici, ovvero con media temporale localmente nulla
- che abbiano lo stesso numero di zero crossing ed estremanti (o che differiscano al massimo di 1) quindi che tutti i massimi siano sopra l'asse temporale e tutti i minimi sotto.

Quest'ultima condizione è direttamente legata alla necessità che il segnale sia “monocomponente”, ovvero che non presenti onde che “salgono” o “scendono” come nel caso della somma di sinusoidi, mentre la prima richiesta necessita di giustificazioni. Per calcolare la media locale di un segnale non stazionario sarebbe necessario introdurre una scala temporale locale, e questo non è possibile per la natura del segnale stesso. L'ostacolo viene scavalcato spostando la richiesta sulla media degli involucri superiore e inferiore del segnale. Anche se non rigorosa, prove sperimentali garantiscono (per gran parte) la validità di questo stratagemma, e la frequenza istantanea di segnali IMF così definiti risulta consistente.

3.2.1 La normalizzazione delle IMF

Recenti studi hanno però messo in luce come la condizione di IMF non sia sufficiente per avere una frequenza istantanea sensata, in particolare per la limitazione introdotta dal teorema di Bedrosian (1963). Questo dimostra che la trasformata di Hilbert del prodotto di due funzioni $f(t)$ e $g(t)$ può essere scritto come

$$H[f(t)g(t)] = f(t)H[g(t)]$$

solo quando gli spettri di Fourier delle due funzioni sono completamente disgiunte nello spazio delle frequenze, e la porzione di

frequenza di $g(t)$ è superiore a quella di $f(t)$. Questa limitazione è critica. Prendiamo ad esempio il segnale analitico di una funzione IMF

$$Z(t) = a(t)e^{i\theta(t)}$$

la cui parte reale, è il segnale analizzato, e poiché $a(t)$ reale si ha

$$X(t) = \Re[Z(t)] = a(t) \cos(\theta(t)) .$$

Quindi possiamo interpretare il segnale come prodotto tra una portante(carrier) $\cos(\theta(t))$ e una modulante $a(t)$.

Poi per definizione della HHT possiamo scrivere la parte immaginaria come

$$Y(t) = H[X(t)] = H[a(t) \cos(\theta(t))] \quad (3.1)$$

che dal segnale analitico, poiché $a(t)$ reale risulta anche come

$$Y(t) = \Im[Z(t)] = a(t) \Im[e^{i\theta(t)}] = a(t) H[X(t)] \quad (3.2)$$

Ma per il teorema di Bedrosian l'uguaglianza tra la 3.1 e la 3.2 è vera soltanto se $\cos(\theta(t))$ e $a(t)$ sono completamente separati nello spazio delle frequenze e $\cos(t)$ occupa sempre una banda di frequenza superiore a quella di $a(t)$. Per soddisfare questo requisito e rendere l'uguaglianza tra la 3.1 e la 3.2 vera sempre, Huang e Long (2003) proposero la normalizzazione delle IMF prima di calcolare il segnale analitico, quindi la frequenza istantanea. Il metodo consiste nel trovare tutti i massimi della funzione, poi ricavare un involuppo $E(t)$ mediante una spline passante per essi. A questo punto normalizzare il set dati dividendo l'IMF (qui indicata come $X(t)$) per $E(t)$.

$$X_{\text{norm}}(t) = \frac{X(t)}{E(t)} ,$$

dove $X_{\text{norm}}(t)$ dovrebbe essere l'onda portante con tutti i massimi locali uguali a uno.

3.3 Empirical Mode Decomposition

Si è visto che la trasformata di Hilbert ha un buon comportamento solo per una specifica classe di funzioni, le IMF. Sfortunatamente la maggior parte dei segnali non vi appartengono. Questi segnali sono come caratterizzati da differenti modi di oscillazione, ognuno dei quali con una specifica frequenza istantanea. Per poter fare un'analisi mediante HHT c'è quindi bisogno di scomporre il segnale originario in IMF. Huang propone un metodo per farlo, detto Empirical Mode Decomposition (EMD).

Perché il segnale rispetti la definizione di Intrinsic Mode Function, una richiesta è che la sua media locale sia nulla, come visto in precedenza. Quindi il primo passo per l'estrazione di un IMF dal set dati consiste nel calcolare l'involuppo superiore $\maxSpline(t)$ (passante per i massimi) e inferiore $\minSpline(t)$ (spline passante per i minimi) e calcolare la media dei due.

$$m1(t) = (\maxSpline(t) + \minSpline(t)) / 2$$

A questo punto si sottrae la media $m1(t)$ al segnale originario $X(t)$

$$h1(t) = X(t) - m1(t) .$$

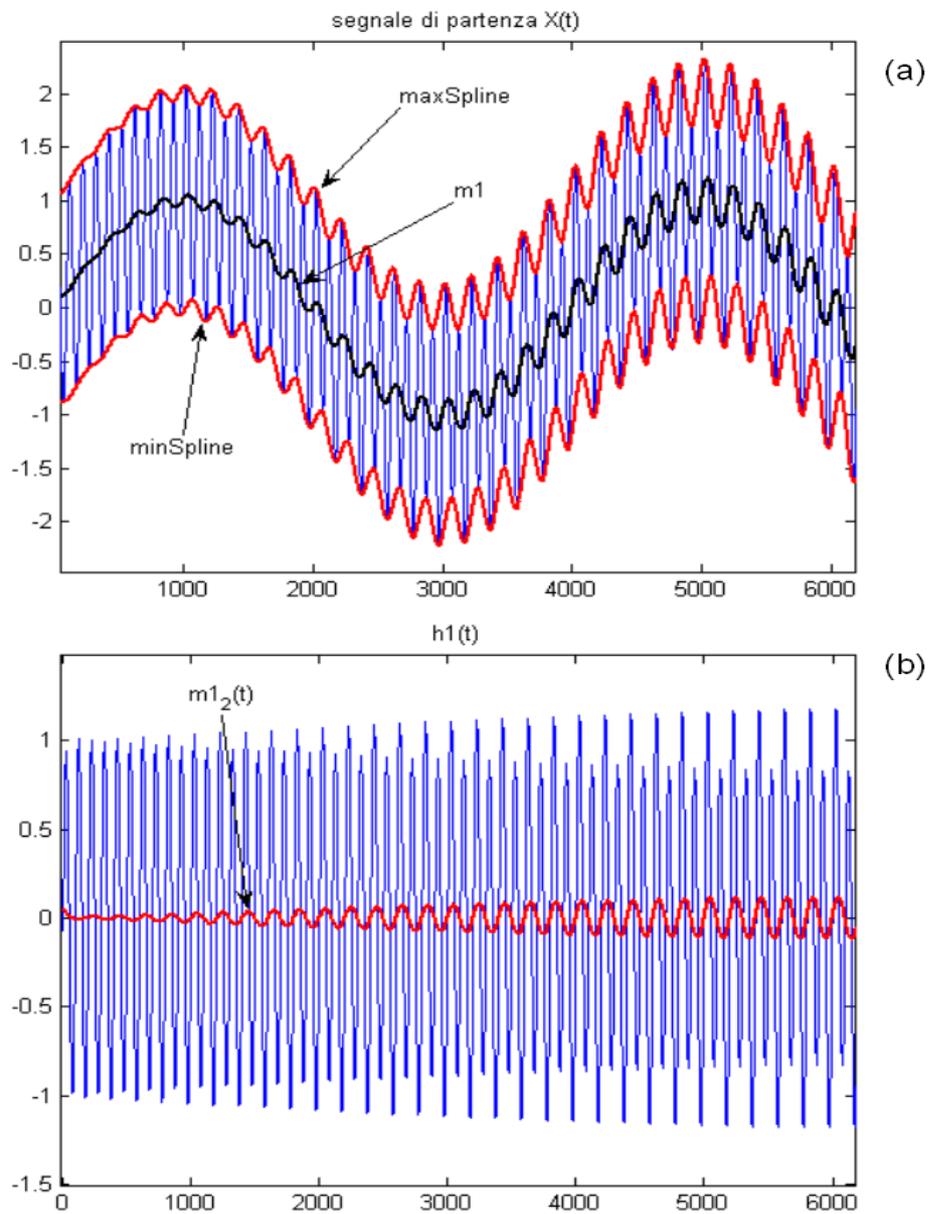


Illustrazione 3.2: Segnale di partenza(a) e prima candidata IMF(b)

Idealmente la funzione $h1(t)$ dovrebbe già essere IMF, per come è stata costruita. Ma in realtà il segnale risultante avrà ancora massimi negativi e minimi positivi. Questo perché la media calcolata con gli involucri non corrisponde perfettamente alla media reale, di conseguenza si possono generare nuovi massimi e alcuni già esistenti possono cambiare posizione o valore. Il processo di sifting viene quindi reiterato per ogni nuovo candidato IMF finché il k -esimo non rispetta tutte le condizioni

$$h1_k(t) = h1_{(k-1)}(t) - m1_k(t) = c1(t)$$

dove $m1_k(t)$ è la media calcolata dagli involucri di $h1_{(k-1)}(t)$ ($k-1$)esimo candidato IMF e $c1(t)$ è la prima IMF estratta.

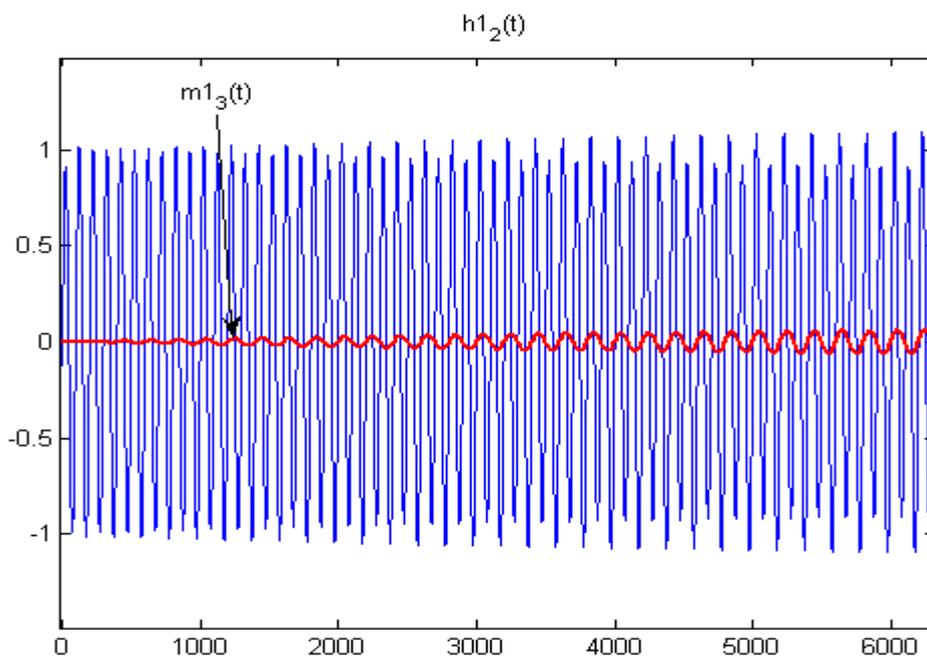


Illustrazione 3.3: seconda candidata IMF

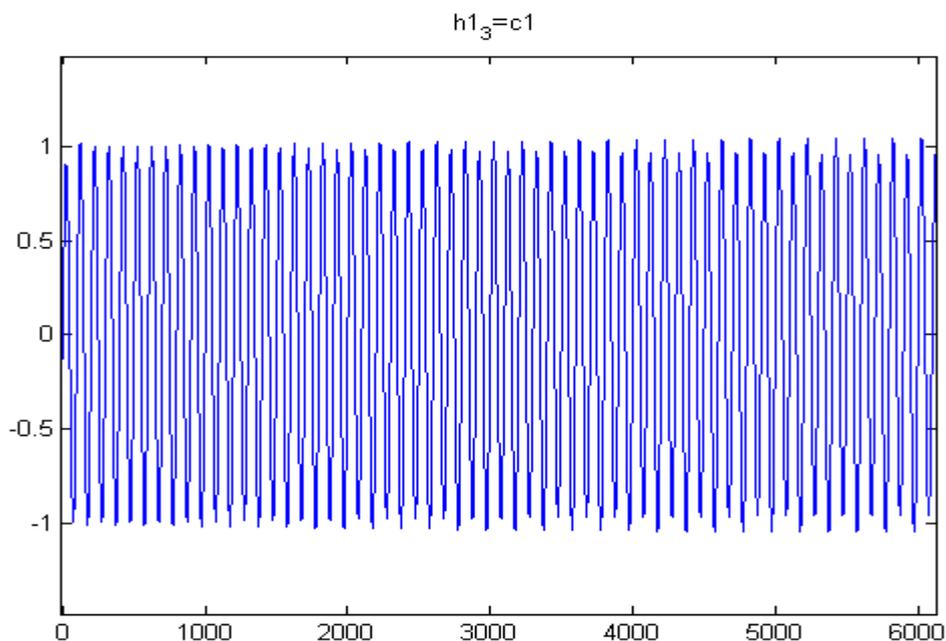


Illustrazione 3.4: prima IMF estratta

L'IMF verrà sottratta dal set dati iniziale ricavando un residuo

$$r1(t) = X(t) - c1(t)$$

E da $r1(t)$ si procede con una nuova operazione di sifting, quindi una nuova IMF. Il processo viene reiterato fino a scomporre completamente il segnale originario.

$$r2(t) = r1(t) - c2(t)$$

$$r3(t) = r2(t) - c3(t)$$

.

.

$$rn(t) = r(n-1)(t) - cn(t)$$

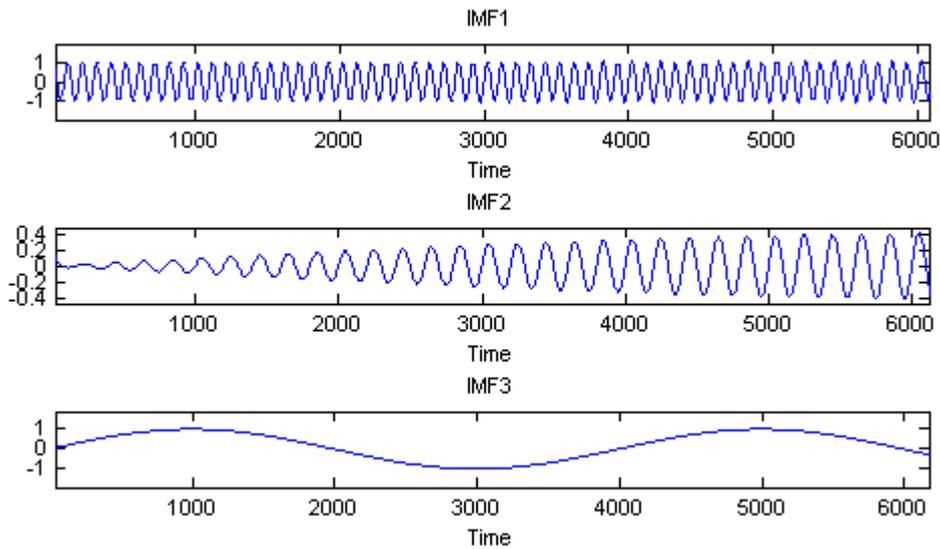


Illustrazione 3.5: Tutte le IMF estratte dal segnale in figura 3.2(a)

La prima IMF necessariamente sarà quella con oscillazioni più veloci, mentre le successive avranno frequenze sempre minori. L'EMD viene interrotta quando l'ultimo residuo contiene meno di due massimi e due minimi, necessari per il calcolo della spline, e a questo punto avremo scomposto tutto il segnale.

$$X(t) = \sum_{k=1} c_k + r_k.$$

4-Algorithmo al calcolatore e test sull'HHT

Allo scopo di esplorare le potenzialità della trasformata HHT e il suo possibile utilizzo nell'ambito della radioastronomia, in particolare per la rilevazione di segnali di natura sconosciuta per il progetto SETI, è stato prodotto un algoritmo, quindi un codice per l'ambiente Matlab, e sono stati fatti dei test. I segnali utilizzati sono sia di sintesi che acquisiti dal radio osservatorio, per testare il comportamento della trasformata in condizioni reali.

4.1 Implementazione dell'EMD

L'algoritmo che calcola la HHT si compone di tre parti. Per primo dai set dati di partenza vengono estratte le IMF mediante l'Empirical Mode Decomposition, poi di ognuna di esse viene calcolata la frequenza istantanea mediante la trasformata di Hilbert. Infine viene visualizzato l'Hilbert Spectrum di ogni IMF, ovvero un'immagine in coordinate tempo-frequenza in cui le diverse gradazioni di colore rappresentano zone di maggiore o minore energia. In questo paragrafo e nei due seguenti verranno presentate singolarmente le tre fasi, insieme ad alcuni test su segnali sintetizzati per giustificare le diverse scelte implementative.

Nella figura seguente viene presentato il diagramma a blocchi della prima parte dell'algoritmo per la HHT, l'Empirical Mode Decomposition.

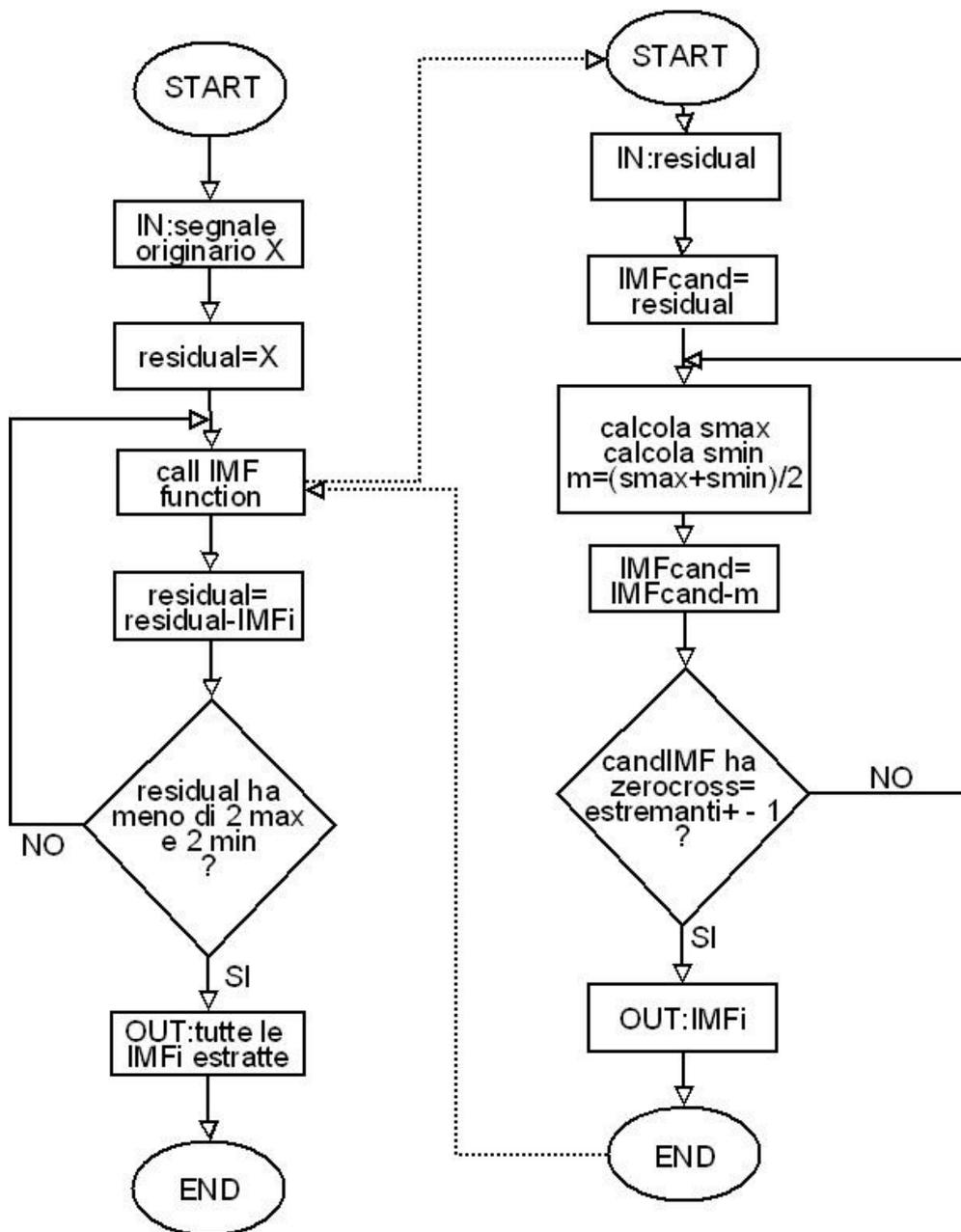


Illustrazione 4.1: Flow Chart dell'algoritmo implementato per l'EMD

L'albero di sinistra rappresenta il processo EMD, che per estrarre le successive IMF dai dati di partenza utilizza un'ulteriore funzione, appunto chiamata IMF (albero di destra). In quest'ultima vengono fatti i sifting successivi finché il segnale denominato IMF_{cand} non soddisfa la condizione sul numero di zero-crossing ed estremanti, e un'ulteriore richiesta che verrà presentata in seguito. A questo punto la funzione estratta viene restituita al processo EMD, che la sottrae all'attuale residuo (ovvero il segnale originario al quale sono state sottratte le IMF estratte precedentemente). Il risultato dell'operazione è un nuovo residuo che viene analizzato per decretare o meno la fine della decomposizione. Se questo ha meno di due massimi o meno di due minimi il processo si interrompe, se invece la condizione non è rispettata viene chiamata nuovamente la funzione che realizza i sifting, per estrarre una nuova IMF.

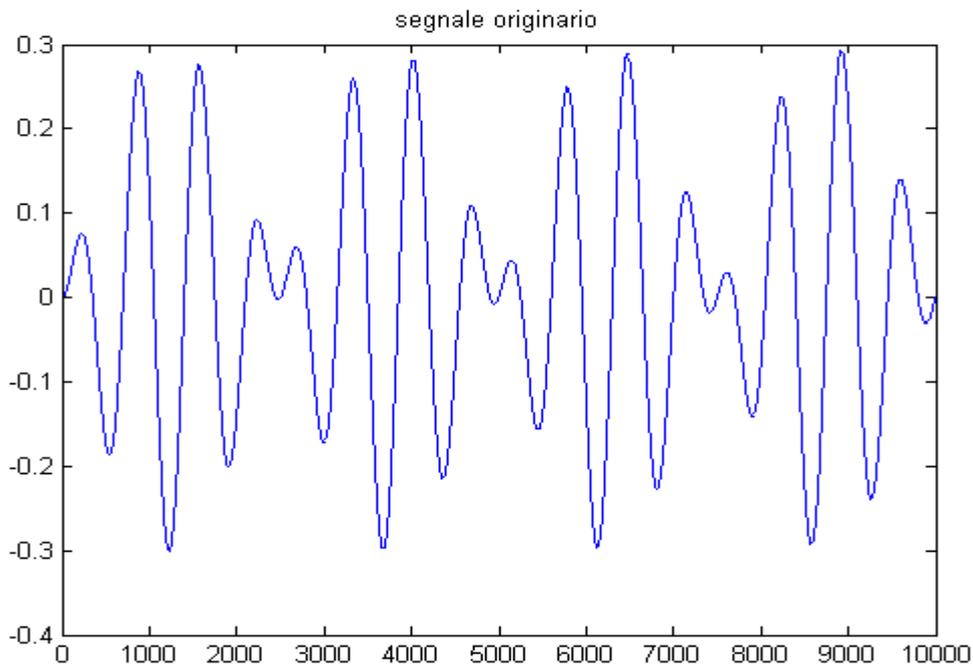


Illustrazione 4.2

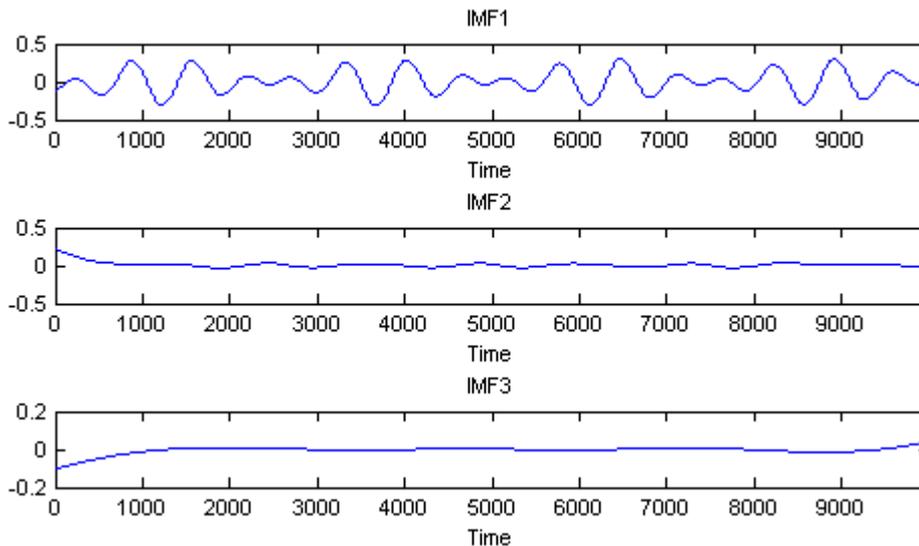


Illustrazione 4.3: IMF estratte dalla decomposizione del segnale in figura 4.2.

Guardando la figura 4.2 del segnale originale e la sua decomposizione (figura 4.3) si può subito notare che il grafico dell'ultima IMF estratta non rappresenta un modo di oscillazione dei dati, in particolare ai bordi. Il processo di sifting è responsabile di questo, per i due motivi presentati di seguito.

4.1.1 Problemi legati alla spline e ai bordi

Come già illustrato il processo di estrazione di IMF si basa sulla considerazione che un segnale è simmetrico in buona approssimazione se la media degli involucri superiore e inferiore risulta nulla. Per il calcolo della spline superiore sono stati individuati tutti i punti di massimo relativo (o di minimo relativo nel caso della spline inferiore) e sono stati interpolati mediante una funzione cubica. Anche in segnali costruiti simmetrici, è inevitabile che le due spline diano una media non nulla. Inoltre la spline cubica, per sua natura, tende ad interpolare i punti mantenendo una certa fluidità. Quindi nel caso di zone con basse frequenze c'è il rischio che l'involucro sovrastimi o sottostimi il reale

andamento del segnale. Innescando così l'estrazione di IMF che non rappresentano modi reali di oscillazione del segnale originario.

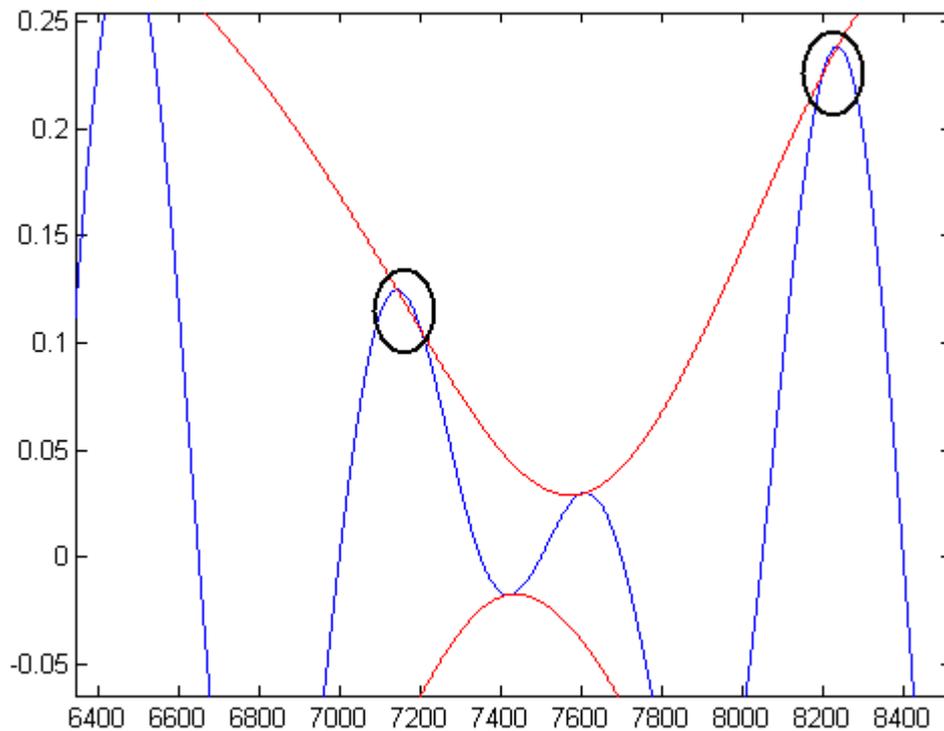


Illustrazione 4.4: Sono indicati con le circonferenze i punti in cui la spline interseca il segnale

Un ulteriore problema, ancora legato al processo di sifting, ma non direttamente al calcolo delle spline, è dovuto alla prima e ultima parte della serie temporale. Questo è provocato dalla mancanza di punti in cui calcolare la spline. Soprattutto quando la frequenza del segnale da cui estrarre IMF è bassa, verso i bordi gli involucri non sono vincolati da estremanti, quindi tendono a divergere seguendo l'andamento di cubica. Anche nell'analisi di Fourier si ha un problema analogo, chiamato fenomeno di Gibbs. Ai bordi del segnale nette

variazioni di ampiezza possono introdurre frequenze non reali, perciò esso viene moltiplicato con un particolare tipo di finestra che riduce a zero l'ampiezza ai in queste zone, eliminando così l'end-effect. La soluzione adottata per l'analisi di Fourier potrebbe essere utilizzata per forzare a zero la spline agli estremi. Ma nel caso di segnali non stazionari, la finestra eliminerebbe dell'informazione utile oltre a inserire una modulazione non reale.

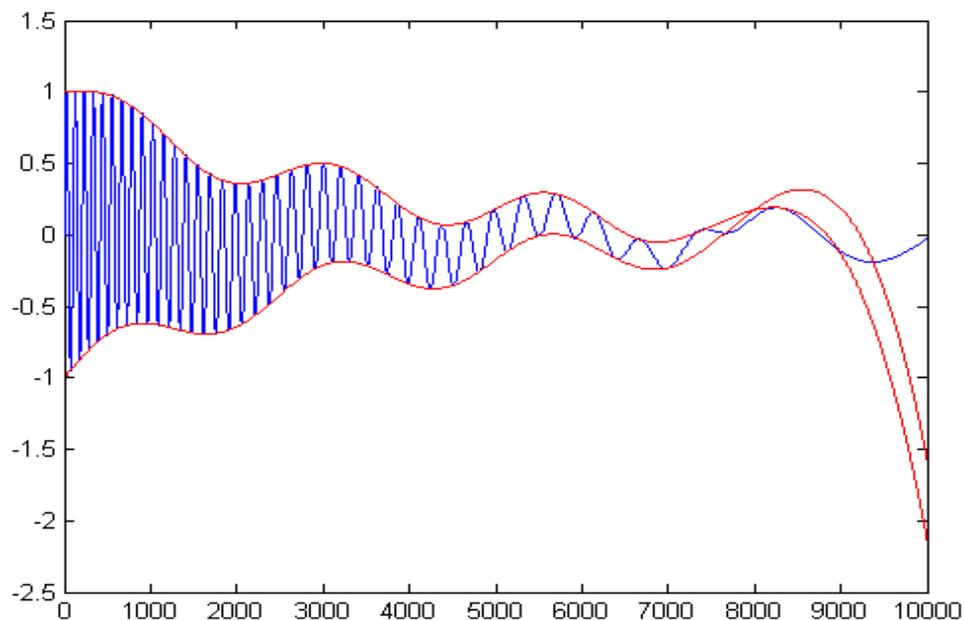


Illustrazione 4.5: è ben visibile come nell'ultima parte del segnale le spline non seguono più il suo involuppo ma divergono verso il basso

4.1.2 Sifting Ulteriori e convergenza dell'EMD

Per una maggiore sicurezza sulla simmetricità della funzione estratta, come suggerito in HHT-DSP([.5]), nell'algoritmo può essere richiesto che il numero di zero crossing ed estremanti resti costante per una quantità arbitraria di sifting dopo aver raggiunto la condizione richiesta. Come presentato, affinché il segnale venga considerato IMF, si chiede soltanto che il numero di zero-crossing ed estremanti differiscano al massimo di uno, non che la media degli involuppi sia zero. Una condizione di questo tipo può mettere a serio rischio la convergenza dell'algoritmo, o comunque ritardarla molto, visto il comportamento delle spline. Perciò il metodo proposto è una soluzione per aumentare la simmetricità del segnale senza pesare sui calcoli. Infatti, la convergenza della decomposizione non è provata matematicamente, e nei test effettuati il tempo di calcolo per la decomposizione di diversi segnali è variato molto. Crescendo in relazione con la sua complessità e la presenza di rumore.

4.2 Implementazione della trasformata di Hilbert e calcolo della frequenza istantanea

Dopo aver estratto tutte le IMF dal segnale analizzato, il passo successivo consiste nel calcolare di ognuna di esse la frequenza istantanea. Come illustrato precedentemente questo viene fatto mediante la trasformata di Hilbert. L'ambiente Matlab fornisce già una funzione per implementarla, quindi non si è scritto del codice per essa, anche se l'utilizzo di tale funzione primitiva porterà a delle complicazioni matematiche che discuteremo in seguito. A questo punto si dispone del segnale analitico della IMF

analizzata, quindi da esso si può calcolare la fase mediante l'arcotangente del rapporto tra la parte immaginaria e la parte reale del segnale analitico. La funzione `angle` che calcola l'arcotangente, anch'essa primitiva in Matlab, è limitata nell'intervallo $[-\pi, +\pi]$. Per avere quindi una fase continua, quindi derivabile per tutta la serie temporale è stato necessario intervenire sul suo valore. Si è assunto (come le caratteristiche di funzione IMF dovrebbero garantire) che la fase sia monotona crescente. Quindi ogni volta che il suo valore risultasse minore del precedente gli è stato sommato 2π .

```
r(1)=angle(z(1));
    for k=2:n
        r(k)=angle(z(k));
        while r(k)<r(k-1);
            r(k)=r(k)+2*pi;
        end
    end
```

I valori di fase calcolati vengono quindi derivati

```
f(1)=r(2)-r(1);
f(n)=r(n)-r(n-1);
f(2)=(r(3)-r(1))/2;
f(n-1)=(r(n)-r(n-2))/2;
for k=3:(n-2)
    f(k)=(r(k+1)-r(k-1))/2+(r(k+2)-r(k-2))/4;
end
```

e per ottenere una pulsazione che risenta meno delle discontinuità locali della fase si sono utilizzati 4 punti. A questo punto la frequenza si ottiene semplicemente dividendo per 2π i valori trovati.

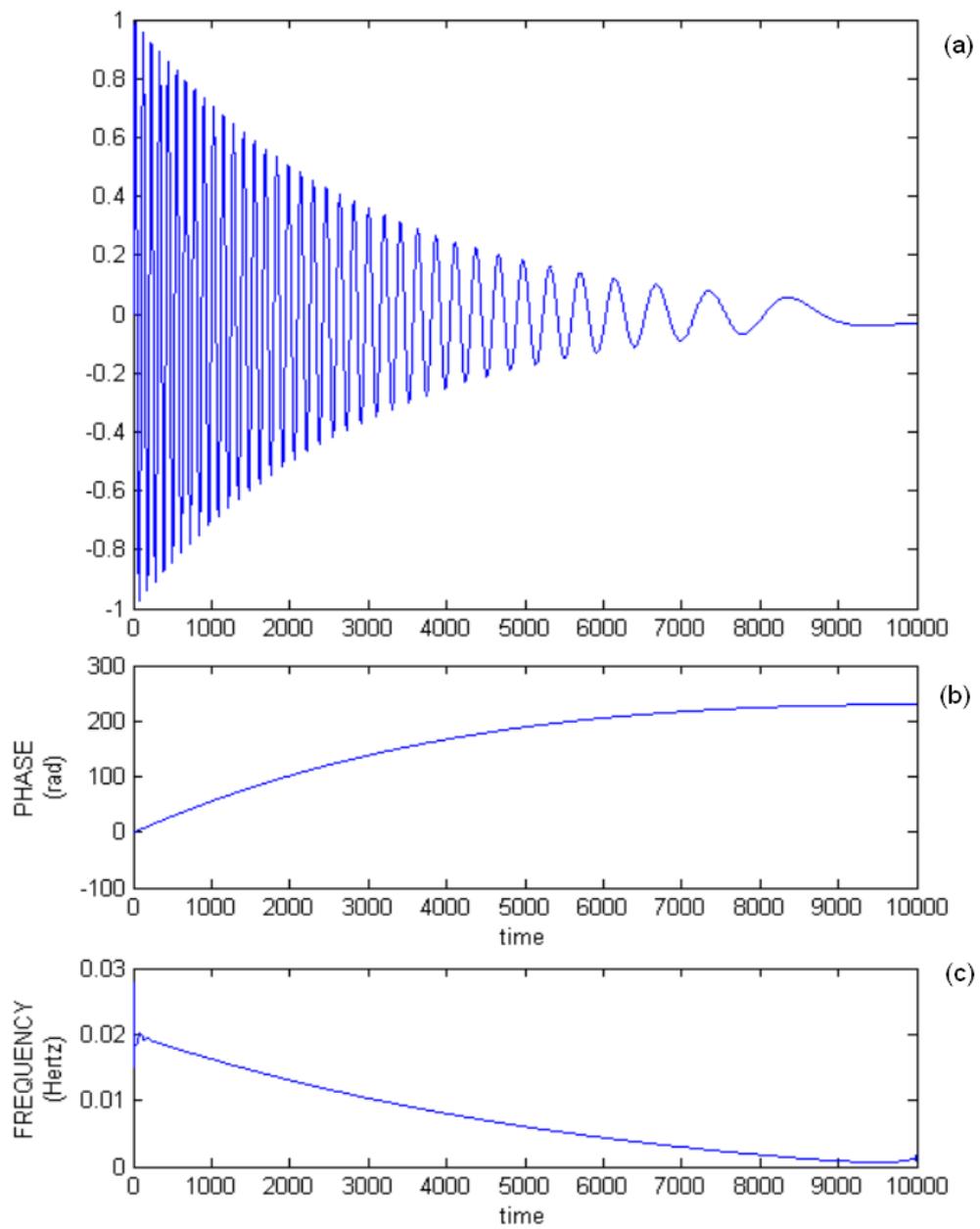
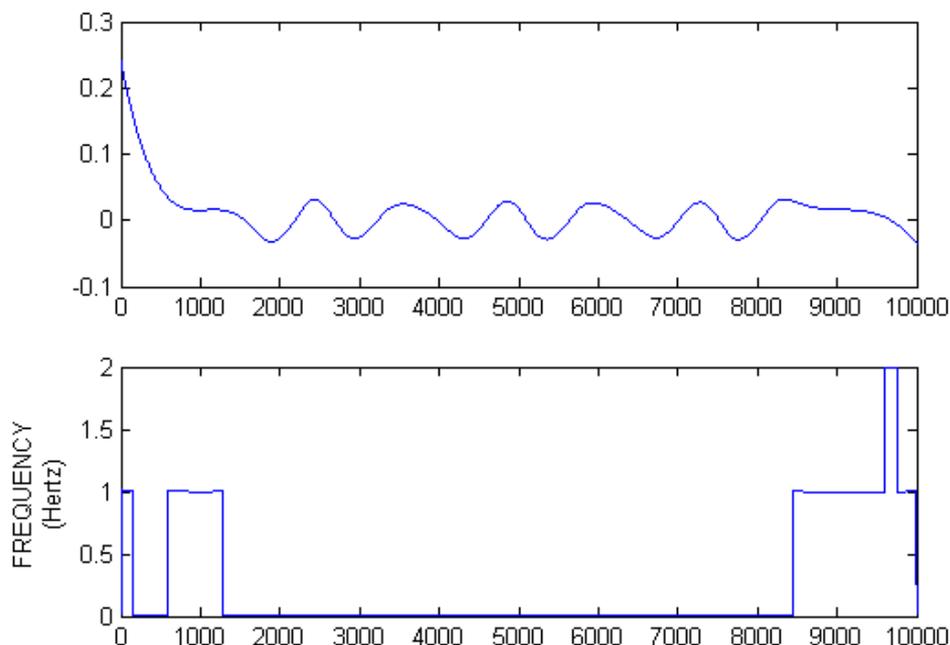


Illustrazione 4.6: in alto(a) l'IMF studiata, (b) la sua fase, (c) la frequenza.

4.2.1 Valori di frequenza oltre Nyquist

Il procedimento mediante il quale viene calcolata la frequenza istantanea può portare a valori maggiori della metà della frequenza di campionamento (frequenza di Nyquist). Questo perché la derivata della fase non è direttamente legata alla scala temporale locale del segnale,



campionamento f_s pari a 1sample/secondo, e sotto la corrispondente frequenza istantanea. Come si vede essa supera la frequenza di Nyquist pari a $f_s/2$.

naturalmente tali valori di frequenza non costituiscono informazione reale, poiché violano i limiti fisici della rappresentazione discreta del segnale.

4.2.2 Normalizzazione

Negli esempi finora analizzati non si è considerato il problema introdotto dal teorema di Bedrosian, che come abbiamo visto precedentemente(3.2.1) è stato risolto dal metodo della normalizzazione dell'IMF. Ma da test effettuati, come mostrato dalla figura seguente, la normalizzazione proposta sembra stravolgere il calcolo della frequenza istantanea. L'andamento della fase resta lo stesso, ma ha un andamento di spezzata, che compromette la sua derivazione. Per questo motivo, nei successivi test si è scelto di evitare la normalizzazione, ottenendo così dei risultati più soddisfacenti, anche se non supportati appieno dalla teoria matematica.

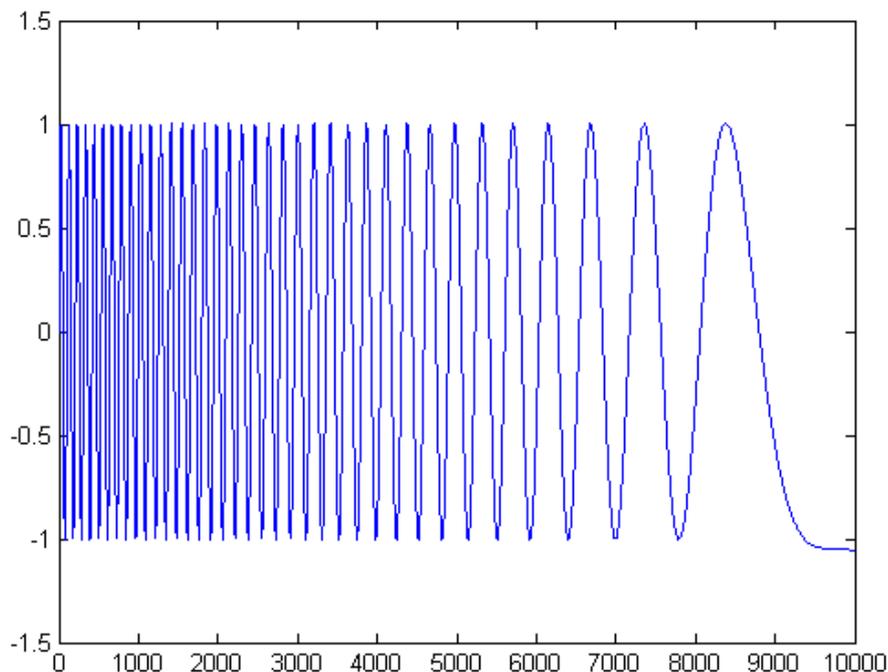


Illustrazione 4.8: Stessa IMF della figura 4f, normalizzata.

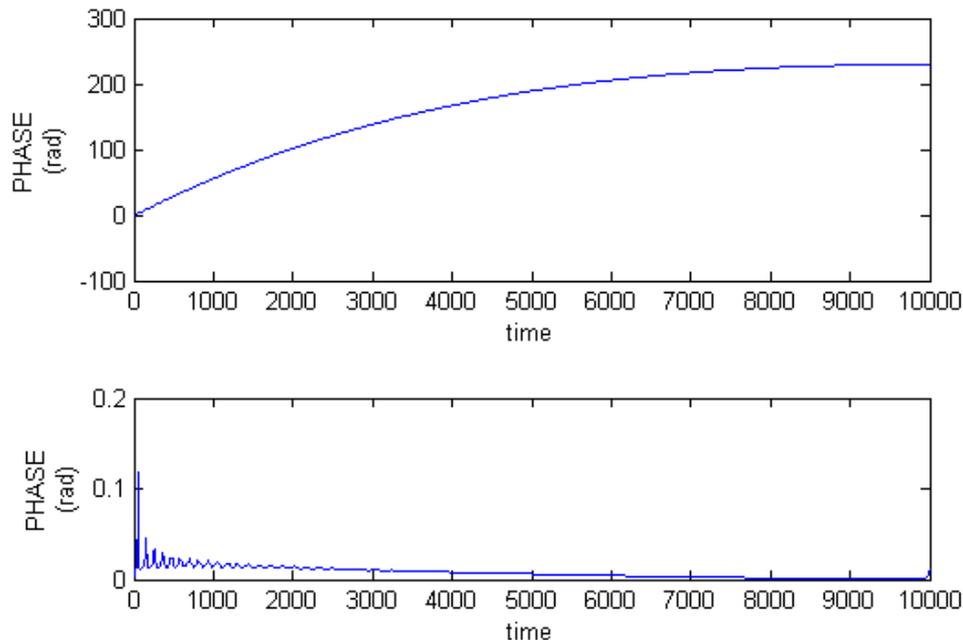


Illustrazione 4.9: In alto fase estratta dall'IMF normalizzata di figura 4h, in basso la frequenza istantanea corrispondente.

4.2.3 Problemi introdotti dal teorema di Nuttall

Come anticipato vi è un ulteriore problema nel calcolo della frequenza istantanea dell' algoritmo presentato, dovuto alla funzione che implementa la trasformata di Hilbert dell' ambiente Matlab. Il teorema di Nuttall dimostra che tale trasformazione non può venire sempre considerata come una traslazione di fase di 90gradi, ma che vi è un errore tra i risultati delle due operazioni. La funzione di Matlab parte invece proprio da questo assunto per implementare la sua trasformata di Hilbert, che viene poi realizzata utilizzando l' efficiente algoritmo della FFT. Considerato il discostamento minimo introdotto da questa approssimazione si è comunque scelto di utilizzare tale funzione, invece di scriverne

una nuova che, senza tali ipotesi, avrebbe comportato inevitabilmente un aumento del costo computazionale.

4.3 Hilbert Spectrum

Per un'analisi qualitativa del segnale, e quindi saggiare il comportamento della trasformata HHT, dai dati estratti nei processi precedentemente descritti si ricava una rappresentazione grafica indicata come Hilbert Spectrum. Questa consiste in una rappresentazione tempo-frequenza dell'energia di ogni IMF con cui viene scomposto il segnale studiato.

L'immagine prodotta ha diverse gradazioni di colore in base ai diversi valori di energia del segnale. La funzione realizzata chiede che venga specificata la risoluzione dell'immagine ovvero il massimo numero di celle per la frequenza (y_{max}) e per il tempo (x_{max}), e in base a questa vengono riscalati i valori di tempo e frequenza

```
k_res=round((k/n)*(x_max-1))+1;  
f_res=round((fi(j,k)/f_max)*(y_max-1))+1;
```

dove k_{res} è il valore di tempo k , riscalato secondo le dimensioni dell'ascissa dell'immagine, e la lunghezza del set dati. f_{res} è il valore di frequenza $fi(j,k)$ (ovvero per l'istante di tempo k e la j -esima IMF) riscalato secondo le dimensioni dell'ordinata dell'immagine e il massimo valore di frequenza. Quest'ultimo è la frequenza di Nyquist nel caso che il massimo valore di frequenza calcolato dai dati lo ecceda. La matrice che rappresenta l'immagine, di dimensioni x_{max} per y_{max} viene poi costruita come

```
A(f_res,k_res)=A(f_res,k_res)+(imf(j,k))^2;
```

ovvero: per ogni valore temporale si scrive all'indice dei corrispondenti valori riscalati di frequenza e tempo, l'ampiezza al quadrato della j -esima IMF. A questo punto sono state fatte delle operazioni

all'immagine allo scopo di una migliore visualizzazione. Per prima cosa un enhancement logaritmico, normalizzando ad uno il valore dell'immagine

```
imA=20*log10(A+1)/(max(max(20*log10(A+1)+eps)));
```

si è sommato all'interno del logaritmo per impedire il calcolo di $\log(0)$, e il valore "eps" (minimo valore rappresentabile dall'ambiente matlab) al denominatore evita divisioni per zero. Poi è stato fatto uno smoothing per evidenziare l'andamento dello spettro.

```
mask=ones(9)/81;  
imA=conv2(imA,mask,'valid');
```

convolvendo una maschera che implementa una semplice media dei dati. Vediamo ad esempio lo spettro seguente.

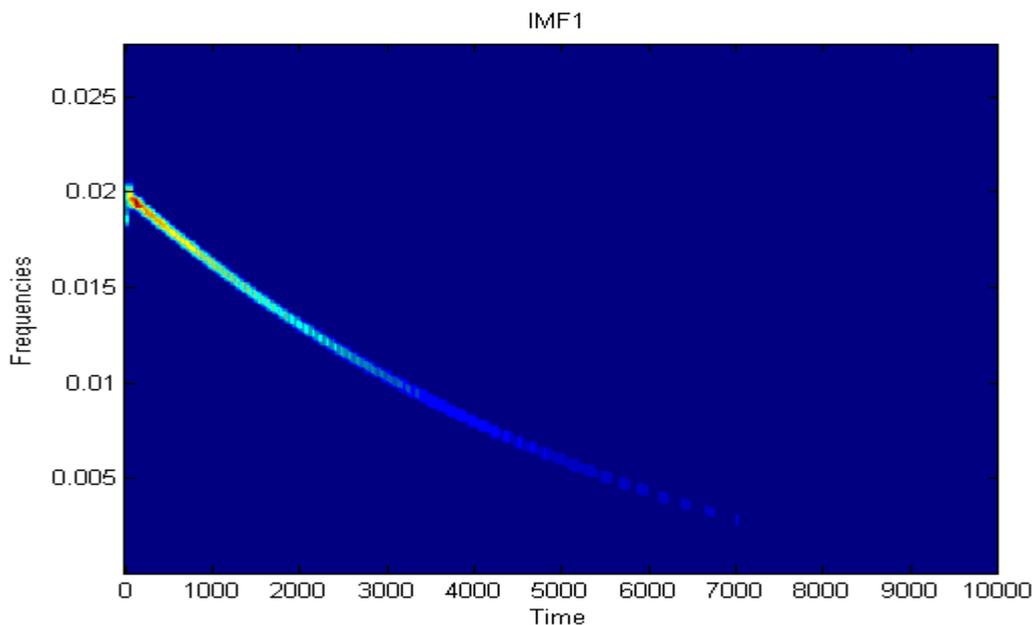


Illustrazione 4.10: Spettro di Hilbert dell'IMF in figura 4.6

Tralasciando le imperfezioni all'inizio, come già spiegato dovute ai problemi ai bordi, l'immagine rappresenta bene l'andamento del

segnale in questione. Esso è stato infatti sintetizzato come sinusoidale modulata in frequenza e in ampiezza con un esponenziale negativo(come si vede nell'illustrazione 4.6).

4.4 Test e confronti tra HHT e Spettrogramma.

In questo paragrafo sono illustrati diversi test confrontando l'analisi fatta mediante HHT e Spettrogramma, effettuati prima su segnali sintetizzati, anche in presenza di rumore, poi su segnali reali rilevati dal radiotelescopio.

4.4.1 Segnale sintetizzato e rumorosità

Le seguenti immagini sono un confronto tra lo spettro di Hilbert e lo spettrogramma di un segnale sintetizzato con frequenza di campionamento di 0.001 samples/sec la cui espressione è $\sin(2 * \pi * 200 * t * e^{(-t * 0.08)})$ a cui è stato aggiunto del rumore gaussiano bianco, in diversi valori di SNR(rapporto al tra potenza del segnale e potenza del rumore). Per il primo segnale verranno anche raffigurate le IMF e le frequenze istantanee. Sono stati tolti dei campioni all'inizio e alla fine delle serie temporali, poiché in quelle zone, per il problema trattato nel paragrafo 4.4.1, si può superare di molto l'energia media del segnale, quindi rendere invisibili nello spettro le altre parti(naturalmente in questo modo si eliminano delle informazioni, questa non è una soluzione al problema).

- SNR=1

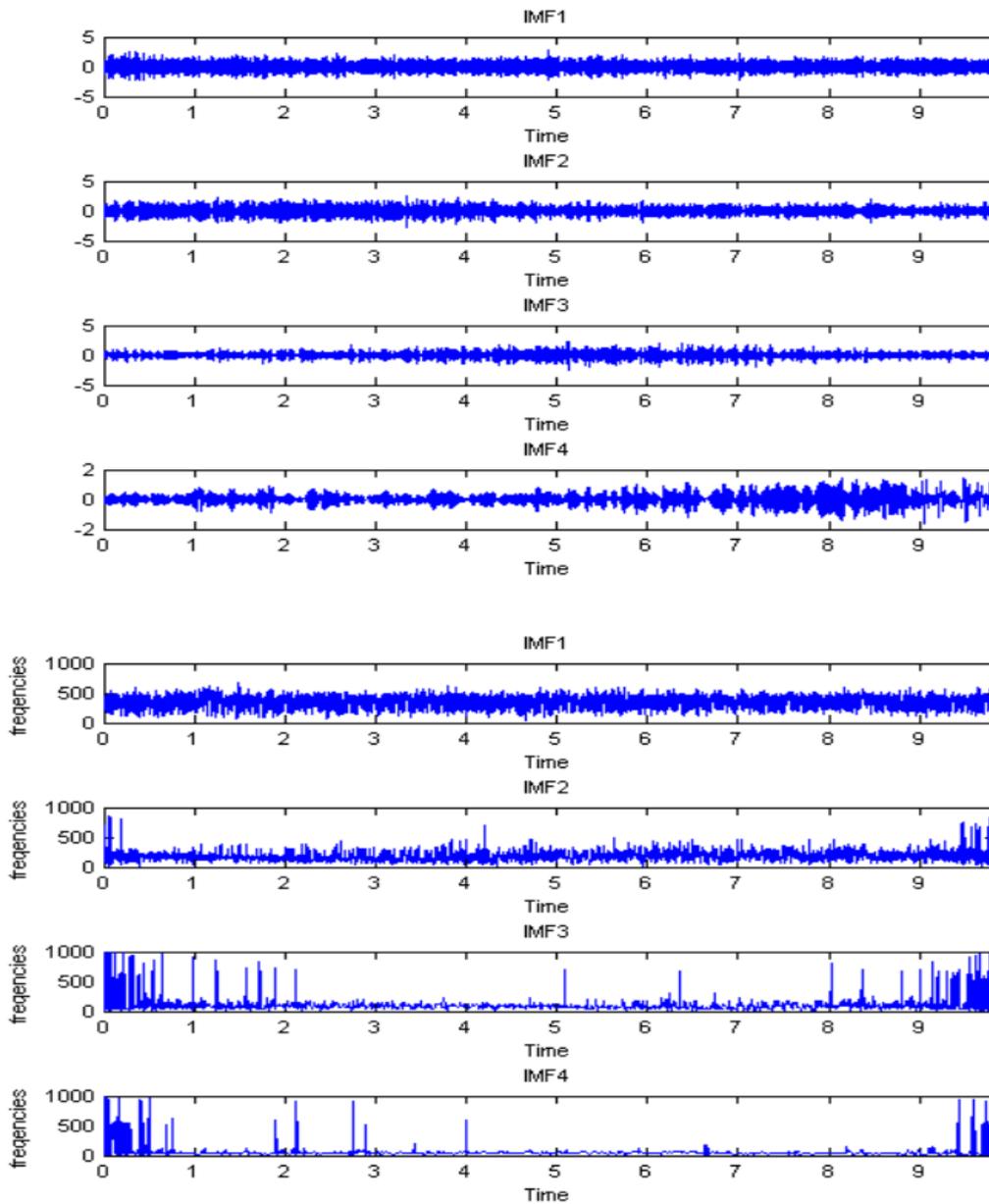


Illustrazione 4.11: IMF e frequenze istantanee del segnale rumoroso SNR=1

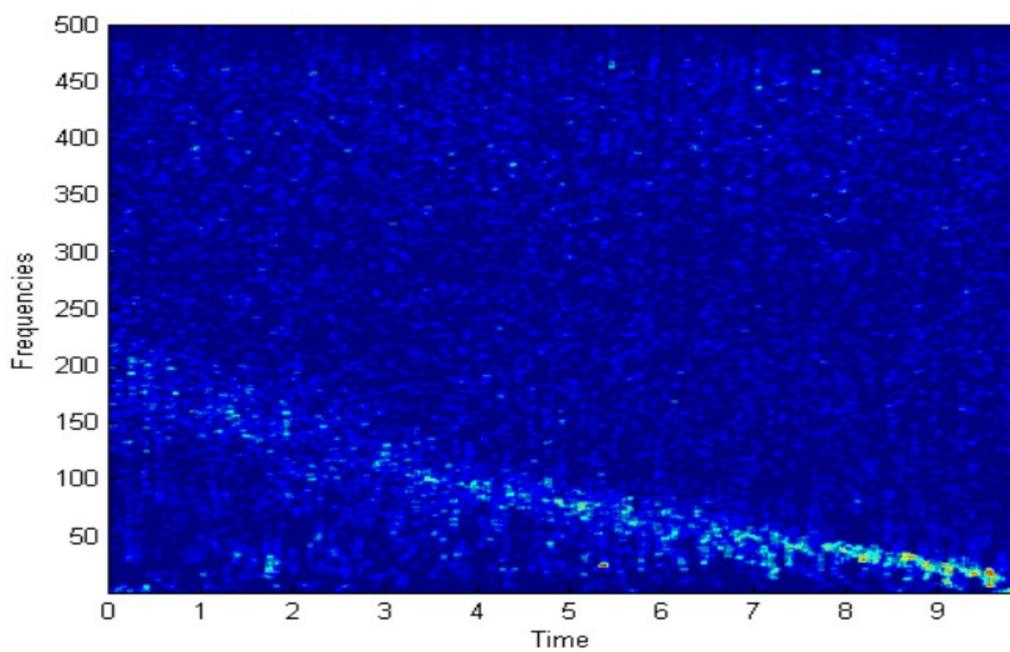
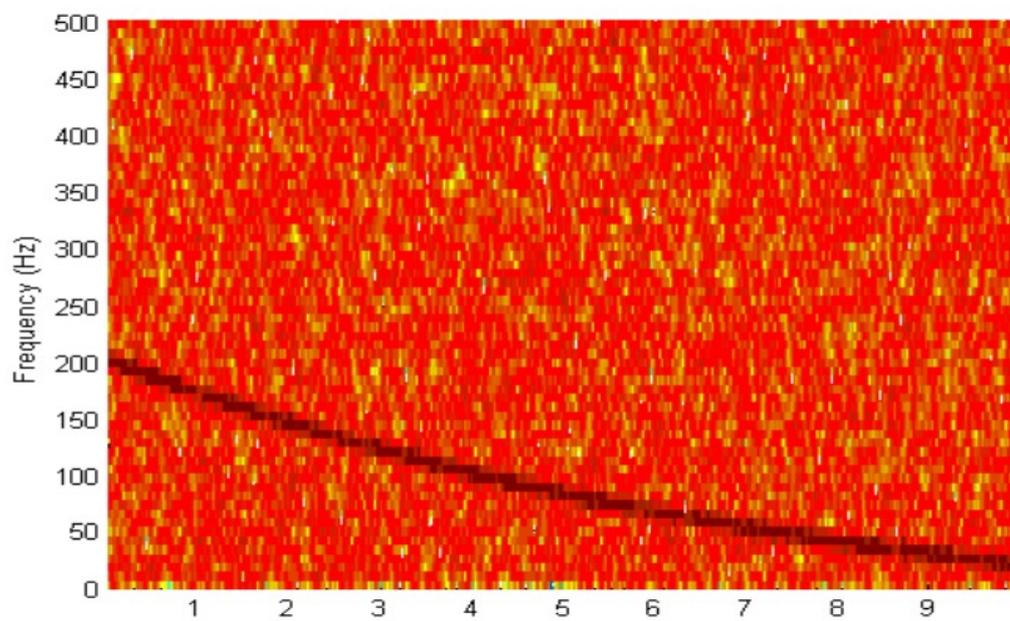


Illustrazione 4.12

-
- SNR=0.5

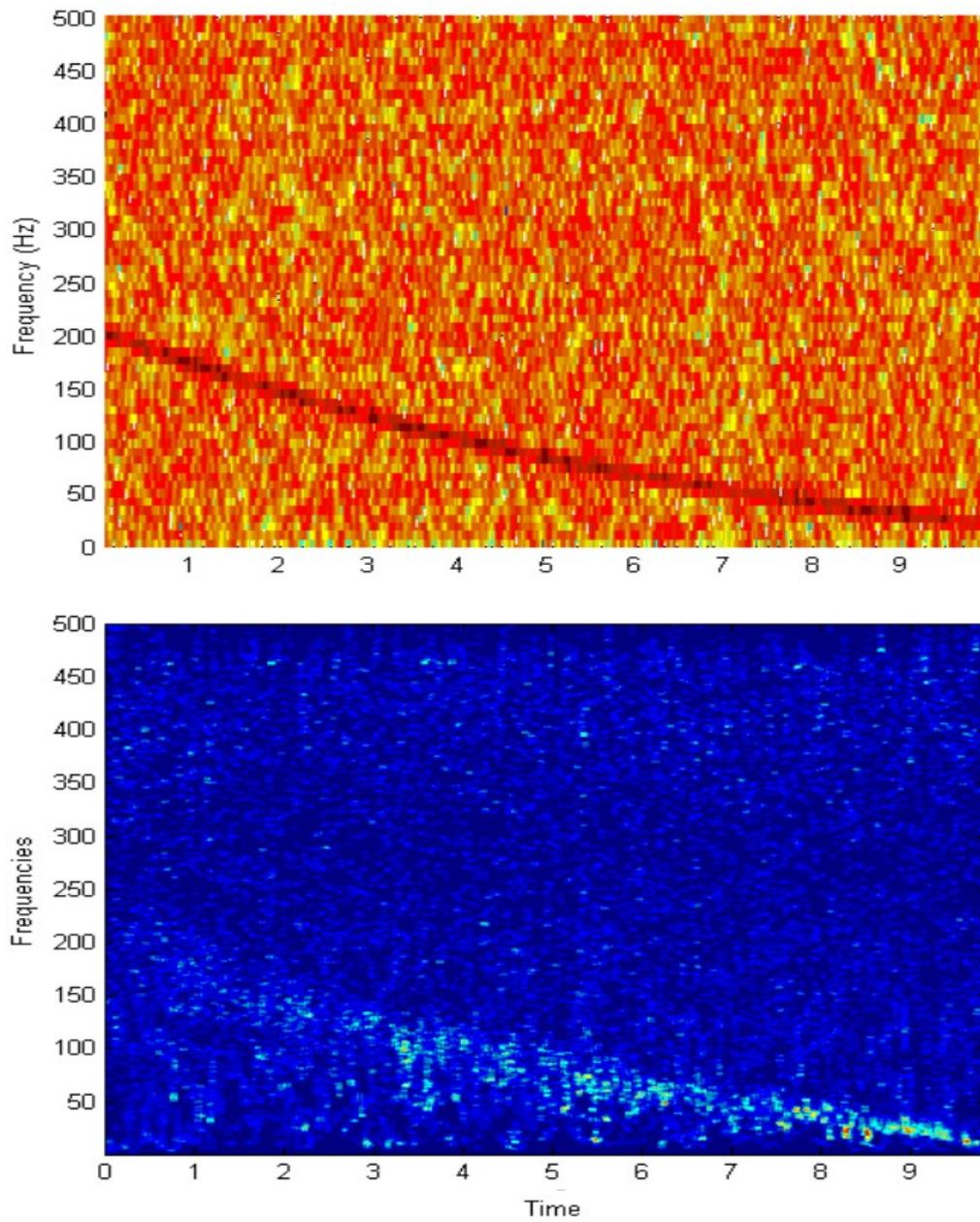


Illustrazione 4.13

-
- SNR=0.1

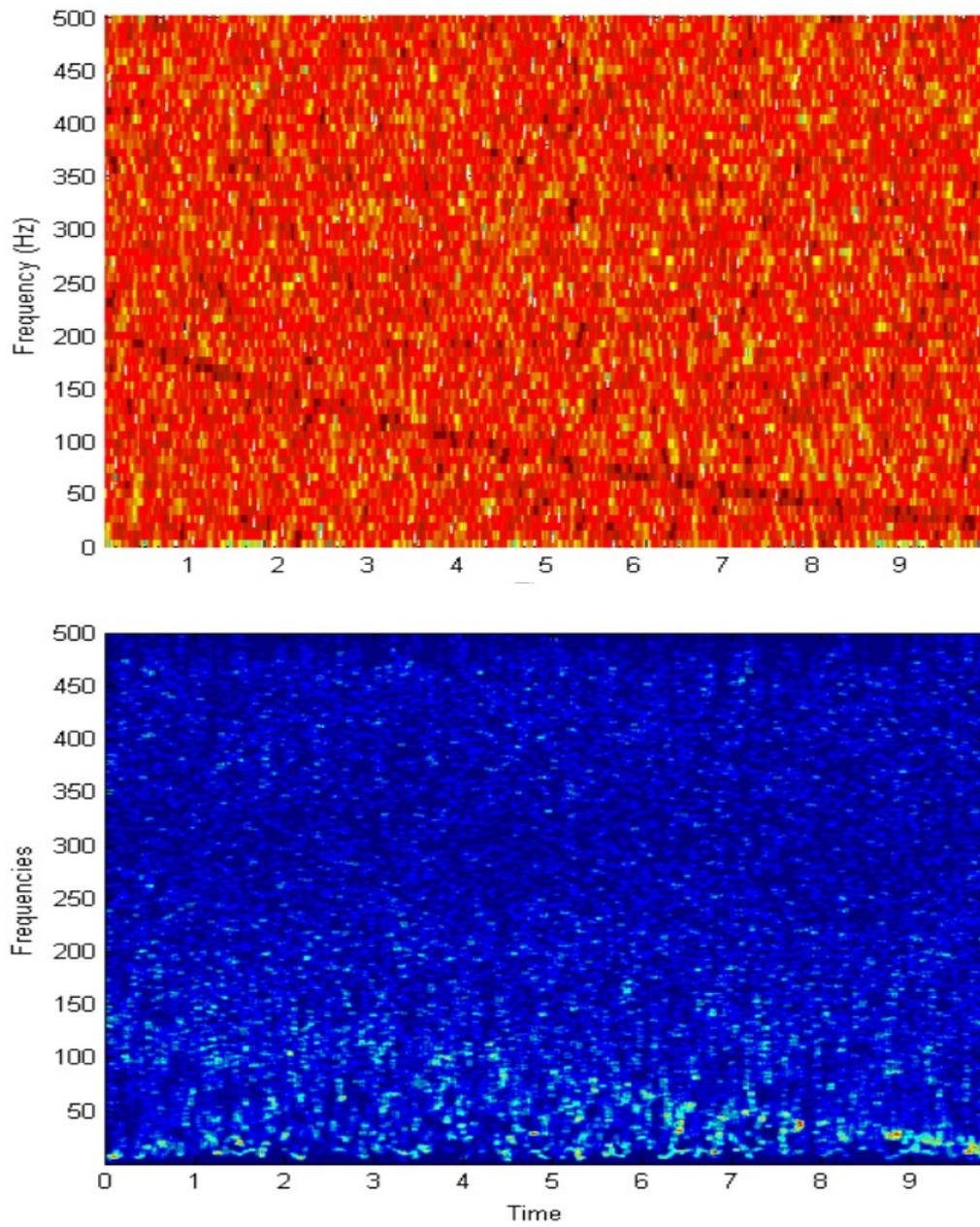


Illustrazione 4.14

Sono state visualizzate soltanto 4 delle IMF, poiché le altre non contengono ulteriori informazioni sul segnale. Lo spettro di Hilbert visualizzato è la somma degli spettri di queste IMF. Lo spettrogramma mostra più chiaramente l'andamento del segnale, che comunque è visibile anche con lo spettro di Hilbert, per i primi due segnali rumorosi. Invece, nel terzo, quando si ha $SNR=0.1$, ovvero un segnale molto rumoroso, soltanto lo spettrogramma, a fatica, riesce a rilevare l'informazione. In queste condizioni, quindi, lo spettrogramma si dimostra appena più sensibile.

4.4.2 Test su segnali rilevati dal radio osservatorio

I seguenti spettri sono stati calcolati su due segnali reali rilevati dal radio osservatorio di Medicina. Nel primo dei due lo spettro di Hilbert mostra una modulazione in frequenza che lo spettrogramma non rileva (la frequenza sembra salire fino a metà del tempo, poi scendere) mentre per l'altro i due spettri mostrano circa lo stesso andamento, anche se lo spettro di Hilbert si dimostra nuovamente meno sensibile, in presenza di rumore, dello spettrogramma.

-
- Primo segnale

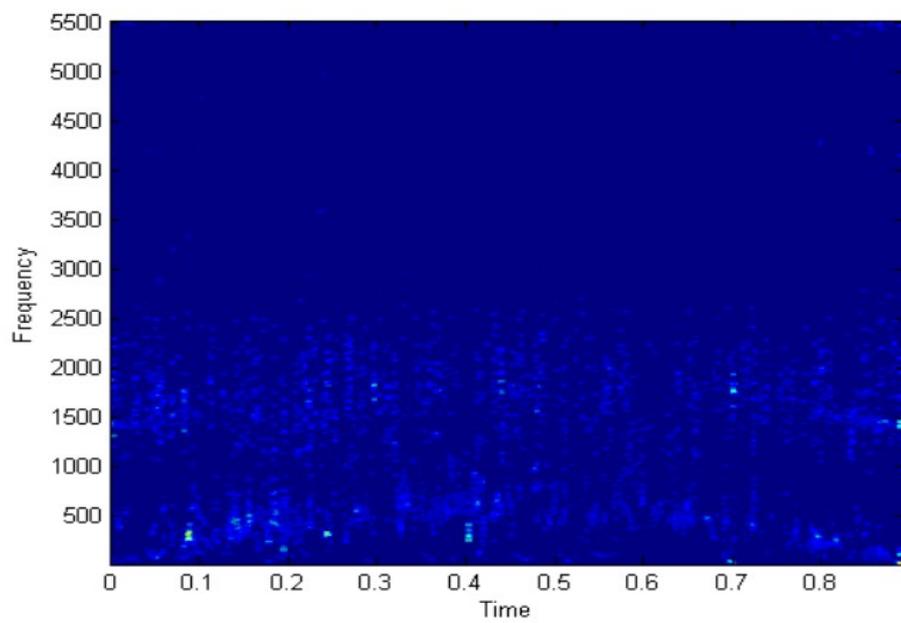
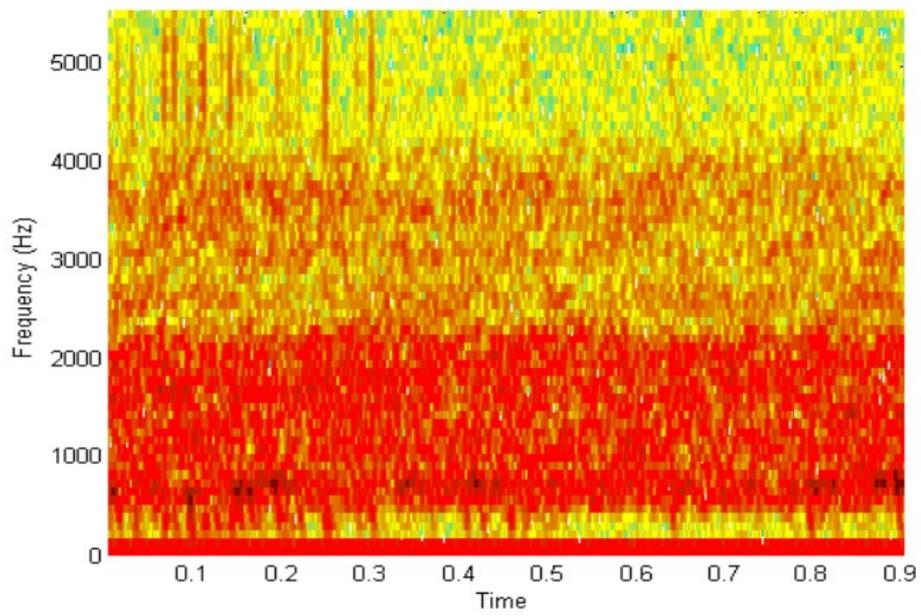


Illustrazione 4.15

- **Secondo segnale**

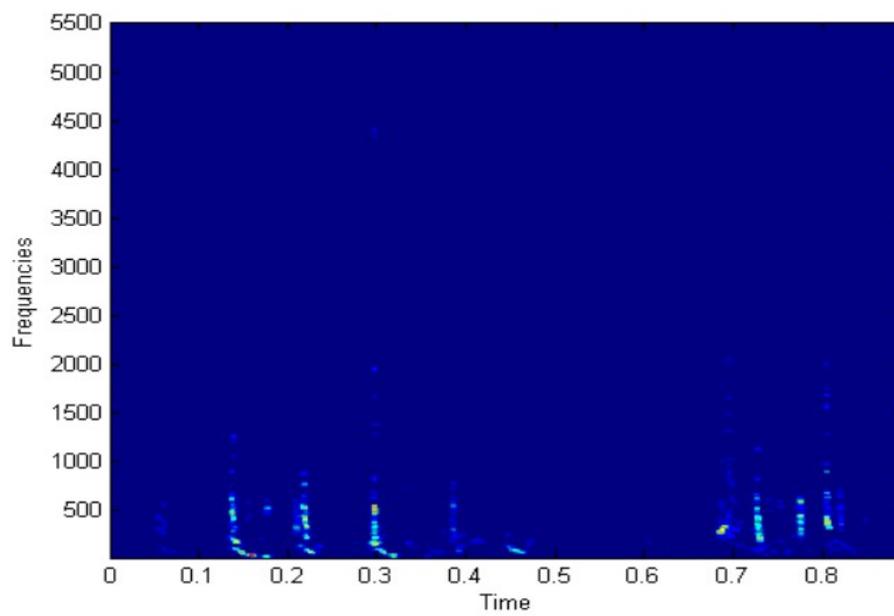
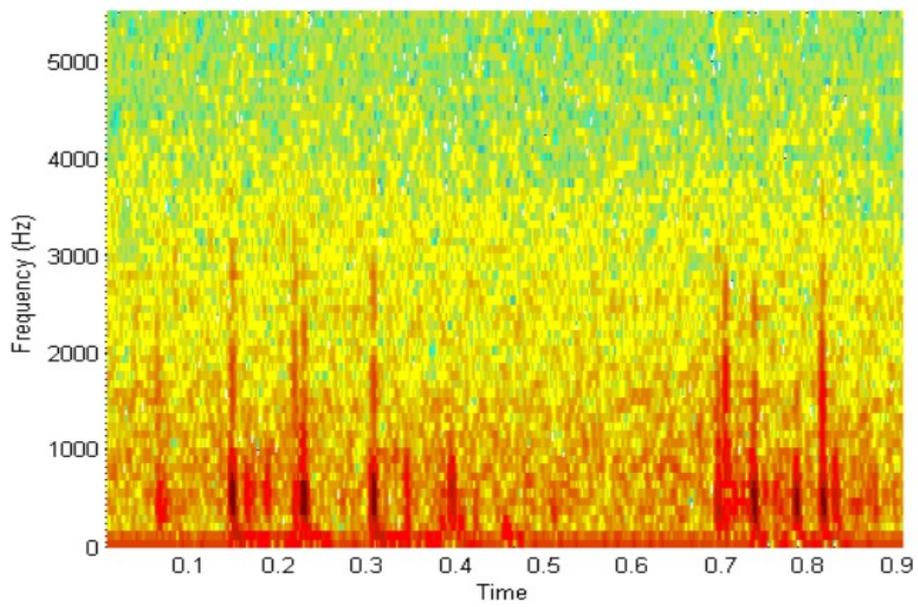


Illustrazione 4.16

. 5-Conclusioni

In questa tesi si è fatto uno studio preliminare dell'HHT come strumento alternativo alla trasformata di Fourier per l'analisi radioastronomica. La possibilità fornita da questa tecnica di lavorare in condizioni del tutto generali, senza vincoli di stazionarietà è stato il principale motivo del lavoro svolto. La trasformata si è dimostrata un valido strumento per l'analisi di segnali modulati, e in condizioni di bassa rumorosità è comparabile con lo spettrogramma. Tuttavia nel caso di $SNR \ll 1$ (con rumore molto oltre il livello dell'informazione) l'HHT non riesce a portare miglioramenti, anzi la trasformata di Fourier è più sensibile. In queste condizioni la decomposizione EMD è in difficoltà e le IMF estratte sono prevalentemente rumorose.

Inoltre se si volesse utilizzare in radioastronomia questo strumento, si dovrebbero considerare le difficoltà introdotte dall'implementazione real-time. Innanzitutto sarebbe necessario un metodo per eliminare il problema ai bordi della finestra temporale, che in quel caso diverrebbe molto più critico, data la necessità di effettuare un'analisi ogni pacchetto di dati ricevuti, invece di avere una lunga serie temporale, quindi solo due bordi. Poi per diversi tipi di segnale e lunghezze, la trasformata ha diversi tempi di calcolo, quindi non potrebbe essere stabilita a priori la frequenza con cui fare l'analisi real-time. E oltre a questi ostacoli implementativi resta la mancanza di una teoria matematica completa che supporti la HHT, e quindi garantisca validità ai suoi risultati. Per un uso esteso della trasformata tutti questi problemi andrebbero affrontati, quindi per ora dobbiamo limitarci ad applicazioni più modeste.

In particolare l'HHT potrebbe fornire un utile strumento, da affiancare a quelli già in uso, per la detection di segnali (rilevare con una semplice risposta sì/no la presenza di informazione in un set dati rumoroso). Nonostante la sua sensibilità non sia ottimale, la HHT non ha le

limitazioni di stazionarietà richieste per le finestre temporali dallo spettrogramma, quindi potrebbe essere consultata quando il rispetto di queste condizioni è critico. Futuri studi potrebbero quindi essere volti all'implementazione di un algoritmo che calcoli l'HHT in tempo reale, che non curi molto la precisione nei risultati forniti, quanto l'efficienza e la stabilità.

Un utilizzo di questo tipo sarebbe più che sufficiente per il progetto SETI, le cui ambizioni si limitano appunto alla semplice rilevazione. La ricerca di forme di vita extraterrestri non ha ancora fornito risultati probanti, ma questo non vanifica il lavoro svolto. Piuttosto, citando il filosofo svedese Niklas Boström, l'ipotesi di essere soli nell'universo deve renderci più responsabili nei suoi confronti.

Bibliografia

- [1] Giuseppe Cocconi, Philip Morrison, *Searching for interstellar communications*, Nature, vol 184, 1959
- [.2] Francois Biraud, *Seti at the nançai radiotelescope*, 1983
- [.3] Norden E. Huang, Zheng Shen, et al, *The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis*, 1996
- [.4] Norden E. Huang, *Introduction to the hilbert huang transform and its related mathematical problems*,
- [.5] Karin Blank, Norden Huang, et al, *HHT-DSP Documentation*, 2004
- [.6] Pierpaolo Pari *Rilevazione di segnali tramite trasformata KLT nel progetto SETI*, 2003
- [.7] Mattia Tani, *Confronto fra trasformate nell'analisi di segnali per lo spettrometro SERENDIP IV* , 2008