

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

FACOLTÀ DI SCIENZE MATEMATICHE, FISICHE E NATURALI
Corso di Laurea in Scienze di Internet

**ANALISI DI SEGNALI AUDIO
MEDIANTE FUNZIONI
WAVELET**

Tesi di Laurea in ANALISI NUMERICA

Relatore:
Chiar.ma Prof.
Carla Guerrini

Presentata da:
Cristina Occhipinti

Sessione II
2009/2010

Indice

1	Introduzione	5
1.1	Cos'è il suono?	6
1.2	Il concetto di acustica	6
1.3	L'audio digitale	9
1.4	Obiettivi della tesi	12
2	La rappresentazione digitale del suono	13
2.1	Il suono analogico e il suono digitale	13
2.2	Il campionamento	15
2.3	La quantizzazione	16
2.4	Struttura di un generico audio codificatore	17
2.5	I banchi filtri	19
2.6	Esempi di banco filtri	21
2.7	Sintesi dei problemi di progettazione	22
3	La Psicoacustica	24
3.1	Definizione	24
3.2	L'orecchio umano	25
3.3	La psicologia dell'udito e caratteristiche fisiche del suono	28
3.4	I parametri della percezione	29
3.4.1	Altezza	30
3.4.2	Timbro	30
3.4.3	Volume	31
3.5	Mascheramento uditivo	31
3.6	Rumore (noise) ed effetti del rumore	32

<i>INDICE</i>	2
4 Le funzioni Wavelet	35
4.1 Introduzione storica	35
4.2 Le funzioni wavelet	40
4.3 Funzione wavelet continua	42
4.4 Funzione wavelet discreta	43
4.5 Filter Bank Wavelet Transform	44
4.5.1 Wavelet filter banks	44
4.5.2 Collegamento tra wavelet e banchi filtro	47
4.6 Proprietà della trasformata wavelet	51
4.7 Applicazione delle wavelet	52
4.8 La WT per elaborare segnali musicali in tempo reale	52
5 Le wavelet nella codifica audio percettiva	54
5.1 La simulazione in MATLAB	55
5.2 Principali caratteristiche dell'implementazione	55
5.3 Considerazioni	57
5.4 Risultati	59
5.5 Codice MatLab	63
6 Conclusioni	69

Elenco delle figure

2.1	Quantizzazione e campionamento	15
2.2	Un generico codificatore percettivo: a)encoder b)decoder	18
2.3	Esempio di soglie di mascheramento di un segnale nel piano tempo-frequenza (a) nacchere (b) piccolo	20
3.1	Morfologia dell'orecchio	26
3.2	Rappresentazione della coclea	26
3.3	Sezione della coclea	27
3.4	Le frequenze sonore percepibili dagli esseri viventi	28
3.5	Tabella danni uditivi associati al livello di pressione esercitata	34
4.1	Rappresentazione di Fourier di un'onda quadra	36
4.2	Esempio di una wavelet	36
4.3	Rappresentazione tempo- frequenza e tempo-scala	37
4.4	Una senoide imperfetta e la sua trasformata di Fourier	38
4.5	Approssimazioni e dettagli della stessa senoide	39
4.6	Traslo l'ondina nel tempo	40
4.7	Ripeto il procedimento con una scala maggiore	41
4.8	Grafico rappresentante la matrice della soluzione della trasformata wavelet	41
4.9	Primo livello di decimazione del banco filtro wavelet	44
4.10	Terzo livello di decomposizione del banco filtro wavelet	45
4.11	Esempio di albero wavelet	45
4.12	Grafico analisi di frequenza tramite FT e tramite CWT	47
4.13	Scomposizione tramite filtri	48
4.14	Ricomposizione tramite filtri	49

4.15	Immagine riassuntiva: a) la funzione di ridimensionamento, b) la funzione wavelet, c), d), e) ed f) i coefficienti del filtro e g) le risposte in frequenza dei filtri di analisi per la Daubechies 2 wavelet	50
5.1	Diagramma a blocchi dell'implemetazione in MATLAB	57
5.2	Rilevamento dei toni mascherati in ogni frame	58
5.3	Grafico ottenuto dalla compressione di un file vocale	61
5.4	Compressione di un file con note di chitarra elettrica	62
5.5	Grafico ottenuto dalla compressione di un file con note di pianoforte	62

Capitolo 1

Introduzione

L'audiocodifica è un'applicazione che rientra nell'area generale della codifica digitale di forme d'onda. Essa ha registrato progressi significativi negli ultimi due decenni in particolare con la digitalizzazione dell'audio. Questa innovazione ha influenzato anche il rapporto domestico con la musica, uscendo dal solo ambito professionale. A mischiare ulteriormente le carte a partire dagli anni '90, è intervenuta una nuova tecnologia chiamata MP3 (MPEG LayerIII), uno standard di compressione per files audio, la cui utilità fondamentale è quella di permettere di economizzare le informazioni necessarie per tramutare in bits le onde sonore che compongono un brano musicale.

Un algoritmo di codifica audio si svolge essenzialmente su dati audio a velocità di trasmissione (bitrate) intensiva e riduce la sua richiesta di trasmissione dati (data-rate) fornendo una qualità trasparente o quasi.

Una fonte comune di formato audio è l'audio CD (Compact Disc), che offre una risoluzione di 16 bit e una frequenza di campionamento di 44,1 kHz. Ciò si traduce in un bitrate di 705,6 chilo bit al secondo (kbps) per un canale monofonico e 1.41 Mbps per un canale stereo, entrambi troppo grandi per la trasmissione su reti comuni. Ma quando compresso, ad esempio dal MPEG-2 Advanced Audio Coding (AAC), algoritmo quasi trasparente che rappresenta lo "state-of-the-art" dell'audio codifica, viene ricondotto a un segnale ad un bitrate di 128 kbps. Si tratta di un rapporto di compressione di circa 11:01 e che richiede una larghezza di banda molto più pratica. Lo sviluppo di strumenti per la codifica dei segnali audio è stato essenziale per fornire modi pratici e convenienti per memorizzare e trasmettere dati audio in una varietà di applicazioni. Per esempio,

gli audio codificatori sono ampiamente utilizzati su Internet per la trasmissione di file audio, trasmissioni radiofoniche, e per condividere musica. Le applicazioni di codifica audio continueranno a crescere poichè la domanda per questo settore è sempre crescente, ciò porterà allo sviluppo di migliori e sempre più efficienti algoritmi di codifica audio.

1.1 Cos'è il suono?

La prima cosa che occorre fare, è spiegare cos'è il suono, l'elemento alla base di tutto questo progetto.

La natura fisica del suono è di tipo ondulatorio: si tratta di onde meccaniche che trasportano energia lontano dalla sorgente sonora, che è un oggetto in vibrazione. A viaggiare non è del materiale, ma un segnale, una variazione continua di qualche parametro legato all'ambiente in cui avviene la propagazione. Tutto ha origine dalla vibrazione di un oggetto o di una sua parte. La vibrazione provoca una successione di compressioni e rarefazioni nel mezzo dell'ambiente circostante e tale disturbo comincia a propagarsi lontano dalla sorgente in tutte le direzioni (un caso visibile è quello delle onde sull'acqua). Questo disturbo o segnale (caratterizzato da alcuni parametri) pone in vibrazione ogni oggetto che si trova nella sua traiettoria. Quando l'onda è passata, ogni cosa torna al suo posto. Tra gli oggetti che si possono trovare sulla traiettoria di un'onda acustica ve ne può essere uno speciale: il timpano dell'orecchio. Inizia un processo molto complesso, che da un segnale acustico porta al riconoscimento della voce di una persona nota, alla comprensione della frase enunciata, all'identificazione di uno strumento musicale e così via.

Spesso durante la sua propagazione il segnale cambia il suo contenuto spettrale, esso arriva quindi al nostro orecchio come una combinazione di sorgente sonora e ambiente di propagazione.

Nel corso della sua evoluzione l'uomo ha imparato a sfruttare la conoscenza di come alcuni ambienti distorcono il suono per localizzare con maggior precisione gli oggetti che lo circondano sulla base delle loro emissioni sonore.

1.2 Il concetto di acustica

L'acustica (dal greco *akoustikà*, dal latino *acùstica*, «udire») è quel ramo della fisica che studia il suono, le sue cause, la sua propagazione e la sua ricezione. I primi studi sul

suono furono eseguiti dai greci a partire dal VI sec. a.C. Le loro conoscenze erano già piuttosto raffinate, come dimostrano i loro famosi teatri, tra cui il più famoso è quello di Epidauro. Uno dei primi scienziati a interessarsi all'acustica come scienza fu Pitagora, ma l'idea che il suono fosse una conseguenza di onde di pressione è stata sostenuta da Crisippo. La prima vera e propria misurazione della velocità del suono fu eseguita dopo il 1600 da Marsenne, è in quegli anni che l'acustica diventa fondamentale, anche per lo sviluppo delle arti. Per alcune di queste, in particolare nel campo delle scale musicali e degli strumenti musicali, sono state elaborate teorie esaurienti soltanto dopo anni di studio scientifico e di sperimentazione da parte dei musicisti. [fonte Wikipedia 2010: Acustica]

Dal punto di vista applicativo, l'acustica può essere suddivisa in numerosi settori, alcuni di essi sono descritti in seguito:

- l'acustica architettonica: è quella disciplina dell'acustica che tratta della produzione, propagazione e ricezione del suono all'interno degli ambienti chiusi (per esempio edifici e sale dei teatri);
- l'acustica degli strumenti musicali: studia a livello teorico e a livello applicativo, gli stimoli acustici prodotti da strumenti musicali come scale musicali e intervalli, vibrazioni delle corde e delle piastre, propagazione del suono nei tubi sonori e emissione dalle canne sonore, tecniche di costruzione degli strumenti musicali tradizionali, elettrici ed elettronici, in relazione alle caratteristiche dei suoni emessi, nonché i problemi costruttivi riguardanti gli strumenti della musica sperimentale;
- l'acustica ambientale: che si occupa dei problemi collegati al rumore;
- l'acustica subacquea: che tratta della propagazione delle onde e della loro percezione negli ambienti marini; l'acustica medica: che si occupa di sviluppare in ambito terapeutico e diagnostico metodi e strumenti basati sulla propagazione di onde acustiche all'interno del corpo umano;
- l'acustica fisica: studia le proprietà e le caratteristiche dello stimolo acustico e dei fenomeni connessi alla sua propagazione nei vari mezzi: riflessione, diffrazione, risonanza, interferenza, assorbimento, la loro amplificazione e rivelazione tramite

dispositivi o apparati di vario tipo, campo questo di specifica pertinenza dell'elettroacustica. Il campo di frequenze che interessa l'acustica fisica si è esteso al di là della cosiddetta banda dell'udibile; alle frequenze inferiori: cosiddetti infrasuoni, alle frequenze superiori (studiate dall'ultracustica) gli ultrasuoni e gli ipersuoni" (Unige 2009);

- *l'acustica fisiologica*: riguarda i due campi, distinti ma interconnessi, dell'udito e della fonazione, quindi si sovrappone in parte alle specializzazioni della medicina costituite dall'audiologia e dalla foniatria. L'acustica fisiologica differisce da questi campi della medicina perchè non si interessa ai fenomeni patologici e limita l'indagine anatomica a quanto è necessario dal lato funzionale. Dal lato uditivo, l'acustica fisiologica studia le grandezze soggettive caratteristiche dell'ascolto (altezza, sensazione sonora, soglie differenziali di frequenza e intensità) in relazione alle corrispondenti grandezze oggettive della stimolazione esterna (frequenza, pressione sonora globale e sua distribuzione spettrale). Si interessa inoltre a tutti i fenomeni caratteristici dell'ascolto (mascheramento, fatica uditiva, effetto direzionale, adattamento, battimenti mono e binaurali e altri fenomeni temporali), anche allo scopo di inquadrare l'insieme di detti fenomeni in una soddisfacente teoria dell'udito. Dal lato della fonazione, l'acustica fisiologica costituisce quella parte della fonetica che comprende lo studio del meccanismo dell'emissione vocale;
- *psicoacustica*: fa parte della psicofisica, e studia il suono inteso come rappresentazione psichica di una realtà esterna, esistente quindi solo nella nostra mente che lo attribuisce a oggetti del mondo esterno, localizzandolo in determinate posizioni dello spazio, e studia come l'effetto di onde elastiche sull'organo del senso dell'udito viene rappresentato mediante il suono, inteso come sensazione psichica. Tuttavia capire come queste onde vengano recepite e convertite in pensieri all'interno del nostro cervello non è affatto da sottovalutare: il suono è un segnale continuo che (approssimando a zero il volume delle molecole d'aria) può teoricamente portare un infinito numero di informazioni (essendoci un infinito numero di frequenze portanti, ognuna contenente informazioni relative ad ampiezza e intensità). Individuare le caratteristiche peculiari della percezione uditiva permette agli scienziati ed agli ingegneri di concentrarsi, per l'analisi e la progettazione di strumenti e apparec-

chiature acustiche, sulle componenti realmente udibili. Questo argomento verrà approfondito meglio nel secondo capitolo di questa tesi.

1.3 L'audio digitale

L'effetto uditivo (dal latino «audire») detto suono, consiste nella percezione da parte di un apposito dispositivo (orecchio di esseri viventi o microfoni artificiali) delle piccole e rapidissime vibrazioni emesse appunto da una "sorgente sonora" e propagate nell'ambiente circostante da un mezzo atto alla loro propagazione (in genere l'aria, ma anche acqua e rocce sono sede di fenomeni analoghi). Quindi il suono si genera in natura come vibrazioni meccaniche, si propaga in analogo maniera e sempre come vibrazioni viene percepito.

Fin dai primordi l'uomo ha cercato di trovare varie soluzioni per facilitare questa naturale propagazione con l'ausilio di dispositivi artificiali che aumentassero l'efficienza di questa trasmissione, in particolare cercando di far attraversare al suono spazi maggiori senza essere troppo contaminato da disturbi (detti rumore, o noise in inglese) ed eventualmente essere memorizzato su supporti di memorizzazione che ne permettessero, oltre che il trasporto nello spazio, anche quello nel tempo, cioè che ne permettessero una memorizzazione permanente. Per far ciò molti artifici tecnologici sono stati usati, fino alla definitiva invenzione del telefono (da parte dell'italiano Antonio Meucci, invenzione però brevettata per primo dall'americano Bell), che permetteva di trasformare i segnali sonori in segnali elettrici che contenevano tutte le informazioni (o quasi) del segnale sonoro originale, ma che potevano essere più facilmente (e più velocemente) trasportate nello spazio (tramite fili elettrici in un primo tempo e successivamente anche tramite onde elettromagnetiche con l'invenzione della radio di Guglielmo Marconi). Per ottenere ciò, al suono veniva associata un'altra grandezza fisica, con caratteristiche di rigorosa relazione con il suono che rappresentava. Questa relazione rigorosa con il suono di partenza permetteva all'arrivo della grandezza fisica di essere di nuovo riconvertita in suono, come ad esempio succede nel ricevitore del telefono, dove la corrente elettrica, che ha trasportato l'informazione sonora, viene all'arrivo ritrasformata in suono dal piccolo altoparlante inserito nella cornetta. Questa relazione di corrispondenza viene

detta "codifica" del suono in corrente elettrica. Analoghi criteri di codifica si hanno ad esempio nella memorizzazione dei suoni codificati in parametri magnetici nel nastro dei registratori magnetici a nastro, o nei parametri geometrici di forma e profondità dei solchi nelle vecchie registrazioni su dischi di vinile. Le moderne tecnologie informatiche, che mettono a disposizione computer capaci di elaborare grandi quantità di numeri al secondo, forniscono un'ulteriore e diversa possibilità di codifica dei suoni, associando ai parametri acustici delle onde sonore delle lunghe serie di numeri (detti anche digit, in italiano cifra), che li rappresentano piuttosto fedelmente e che possono, con elevata precisione essere riconvertite nei suoni originali.

Questo processo di codifica delle grandezze fisiche continue (analogiche) in serie numeriche di cifre digitali è detta digitalizzazione e le grandezze sono dette essere rappresentate in maniera digitale. Queste lunghe serie numeriche possono poi essere memorizzate in memorie al silicio (pen drive ad esempio) o in memorie magnetiche (hard disk di computer) o infine in memorie ottiche (CD o DVD), per essere trasportate nello spazio e nel tempo. Il miglior metodo in uso attualmente prevede che un segnale audio sia convertito (quindi codificato) in un analogo segnale elettrico (analogo per informazione contenuta, a meno dell'inevitabile rumore, anche minimo, introdotto da qualsiasi manipolazione) per mezzo di un microfono, il quale produce, quando viene colpito da una onda meccanica sonora continua, un segnale elettrico ininterrotto, i cui valori di tensione, normalmente compresi in un range (detto dinamica del segnale e compreso tra un minimo e un massimo), contengono la forma dell'onda acustica originaria. Un segnale siffatto, continuo nel tempo e che può assumere con continuità tutti i valori all'interno della sua dinamica, è detto analogico.

Una seconda conversione può essere fatta associando a questo segnale una serie numerica (quindi digitale) che codifichi con sufficiente precisione la forma d'onda elettrica analogica originaria, ottenendo così la conversione analogico-digitale, detta brevemente conversione A/D. Per far questo è necessario andare ripetutamente a leggere i valori di tensione continui della forma d'onda analogica con sufficiente frequenza temporale, cioè effettuare una lettura sufficientemente fitta di questi valori di tensione, producendo un numero di letture (e quindi di valori numerici) in genere molto alto per ogni secondo di conversione A/D. Le singole letture sono dette campioni e il teorema del campionamento afferma che se la frequenza temporale di queste letture (detta frequenza di campionamento) è sufficientemente grande, non si hanno perdite di informazione rispetto alla

forma d'onda originale. Cioè la serie di numeri prodotta contiene pressochè intatta tutta la informazione sulla forma d'onda elettrica analogica iniziale. Nei moderni standard tecnologici, in genere le frequenze di campionamento spaziano dagli 8.000 campioni al secondo (Samples per second, S/s) per la voce telefonica, fino ai 44.100 e più campioni al secondo per la qualità musicale. Queste letture di valori di tensione possono poi cadere in un qualsiasi punto della dinamica del segnale, cioè ogni singolo campione può avere un valore compreso tra il minimo e il massimo possibile. Quindi potenzialmente si possono avere infiniti valori di lettura di tensione per ogni singolo campione. Per completare l'opera di conversione del segnale da analogico in digitale, va ora suddivisa tutto il possibile range dinamico del segnale in un numero finito di intervalli e ogni singolo intervallo va codificato con un valore digitale ben determinato. Queste due operazioni si chiamano quantizzazione e codifica di sorgente. La quantizzazione in genere suddivide il range dinamico del segnale in un numero di intervalli potenza del due (2^n intervalli), in maniera tale che ogni singolo campione cadrà inevitabilmente in uno degli intervallini quantizzati e potrà così essere codificato digitalmente con n bit.

I valori più ricorrenti di digitalizzazione attualmente usati vanno da un minimo di 8 bit per campione in campo telefonico (range dinamico del segnale suddiviso in 256 intervallini), fino a 20 e più bit per campione (range dinamico del segnale suddiviso in un milione e più di intervallini). Naturalmente all'aumentare del numero dei bit per campione aumenta la fedeltà del segnale campionato alla forma d'onda originale e si riduce l'imprecisione introdotta dalla quantizzazione (rumore di quantizzazione), ma va osservato che già 8 bit per campione quasi basterebbero per soddisfare i vecchi criteri di alta fedeltà (HiFi). La serie numerica che così discende è detta segnale audio digitale e contiene in sé tutte le informazioni necessarie per ricostruire la forma elettrica originale, che a sua volta è l'immagine quasi perfetta della forma d'onda acustica che l'aveva originata. Volendo, si potrebbe quindi ora procedere alla sua conversione da digitale ad analogica con convertitori D/A, per riottenere la forma elettrica originale, che una volta inviata ad un altoparlante riproduce il suono originario. Tutto questo processo costa in termini di introduzione di rumore vario, ma con le moderne tecniche questo può facilmente essere tenuto sotto una soglia in genere accettabile.

Il segnale audio digitale prodotto dai convertitori A/D è in genere codificato con un certo numero di bit per ogni campione e così una registrazione audio di 60 secondi campionata a 44.100 campioni al secondo, con ogni campione codificato con 16 bit, dà per

risultato una sequenza di 44.100 campioni al secondo per 60 secondi, pari a 2.646.000 campioni, che vanno ora moltiplicati per 16 bit per campione, ottenendo una serie di 42.336.000 bit. Questo segnale audio digitale così codificato è detto "raw", cioè grezzo. Per questi motivi si rende necessario un ulteriore livello di codifica, che consenta di comprimere le informazioni in sequenze numeriche più corte e che occupino meno bit per ogni secondo di conversione. Con le moderne tecniche di codifica si arriva a comprimere il suono in maniera molto efficace, come ad esempio nello standard mp3, tanto usato per diffondere musica e suoni in generale. [fonte Wikipedia 2010: Audio digitale]

1.4 Obiettivi della tesi

L'obiettivo principale di questo progetto è lo studio di alcune tecniche di compressione audio che utilizzano wavelet. Per questo, ho considerato le seguenti attività:

- Rivedere e riassumere la letteratura generale delle tecniche di compressione audio.
- Introdurre e spiegare il concetto di psicoacustica e il suo collegamento con la sintesi audio.
- Includere una breve panoramica delle attuali applicazioni di wavelet nel campo della compressione audio.
- Analizzare un sistema di compressione mediante simulazione in ambiente MATLAB.

Capitolo 2

La rappresentazione digitale del suono

2.1 Il suono analogico e il suono digitale

L'onda sonora può essere rappresentata in due modi, nel dominio del tempo e nel dominio della frequenza. Dato un tipo di rappresentazione, si decide quali sono le grandezze interessate e in che modo dipendono l'una dall'altra: la rappresentazione nel dominio del tempo è una curva dell'ampiezza del suono in funzione del tempo; la rappresentazione nel dominio della frequenza è data da due curve, rispettivamente dell'ampiezza e della fase in funzione della frequenza.

Una rappresentazione analogica è una rappresentazione che lavora per analogia: in altre parole, la curva continua nel tempo delle variazioni di ampiezza viene rappresentata da una curva continua nel tempo delle variazioni di tensione elettrica, curva che è memorizzata nei solchi di un disco in vinile o nel campo magnetico di un nastro. Per essere più precisi, il tracciato dei solchi del vinile e l'intensità del campo magnetico sul nastro seguono l'andamento della curva dell'ampiezza, dopo che questa è stata convertita in un segnale elettrico.

Una rappresentazione digitale non cerca di «imitare» la curva continua di ampiezza con una curva analoga ad essa, ma piuttosto assegna dei numeri che rappresentano di volta in volta il valore dell'ampiezza in istanti successivi di tempo. Sarà la successione di numeri a rappresentare l'andamento della curva.

La differenza è evidente: da una parte si ha una curva continua che approssima una curva continua; dall'altra si ha una successione di numeri che approssima una curva

continua. La rappresentazione digitale non è continua, ma discreta; cioè esistono degli eventi ben definiti che sono i valori dell'ampiezza in precisi istanti di tempo. La rappresentazione digitale si avvale di un codice simbolico preciso dato dalle cifre (digits) che compongono i numeri: fissate le cifre a disposizione e la lunghezza dei numeri, la rappresentazione avrà a disposizione un insieme finito di simboli. Ad esempio, nel sistema decimale le cifre a disposizione sono 10, da 0 a 9. Se limitiamo a tre cifre la lunghezza del numero, si ha che il codice simbolico è composto da 1000 elementi, tutti i numeri da 0 a 999. Nel caso analogico, invece, la curva continua rappresentata non è l'espressione di un codice preciso e ben definito, e quindi non è possibile conoscere le limitazioni intrinseche del sistema.

I vantaggi di un codice simbolico sono molteplici: dalle operazioni di copia del segnale, a quelle di manipolazione. Ma un aspetto totalmente nuovo introdotto della rappresentazione digitale è la possibilità di correzione degli errori introdotti dai supporti per la memorizzazione. Per errore si intende che qualcuno dei numeri che rappresentano il segnale viene letto in maniera differente da come era stato memorizzato o trasmesso.

La rappresentazione digitale ha portato inevitabilmente anche svantaggi. Uno di questi è l'eccessivo costo dovuto all'utilizzo di apparati con grandi capacità di memoria e grande velocità di trasmissione per la memorizzazione e il trasferimento del segnale. Questo svantaggio ha portato allo sviluppo di tecniche di compressione, che hanno permesso in tempi relativamente recenti l'impiego dell'audio nella multimedialità e la trasmissione della musica via Internet.

Il passaggio da analogico a digitale è caratterizzato da due operazioni di discretizzazione, in cui delle grandezze continue vengono ad assumere valori discreti: una è il tempo, dal momento che l'orologio dei convertitori stabilisce quali sono gli istanti da considerare per il campionamento, l'altra è l'ampiezza, dal momento che i numeri che è possibile usare nella rappresentazione appartengono a un codice simbolico definito e preciso. L'operazione che discretizza il tempo è detta campionamento; l'operazione che discretizza l'ampiezza è detta quantizzazione. Nella figura 2.1 si possono osservare gli effetti separati e combinati delle due operazioni: in (a), si ha il segnale analogico di partenza; in (b) il segnale è quantizzato, cioè l'ampiezza assume solo alcuni valori e non tutte le infinite possibilità offerte mentre il tempo è continuo; in (c) il tempo esiste solo in istanti discreti, quindi anche il segnale esiste solo in quegli istanti; in (d) il segnale è discreto sia nel tempo che nell'ampiezza, è cioè digitale, numerico. Nel prossimo

paragrafo si descrivono nel dettaglio le operazioni di campionamento e quantizzazione.

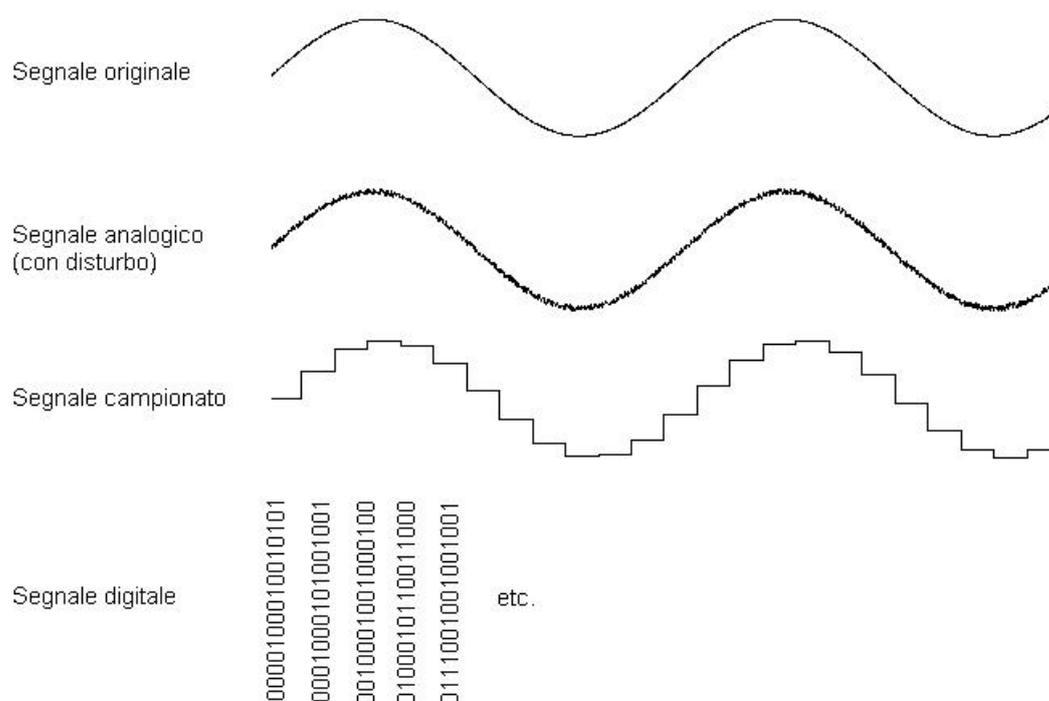


Figura 2.1: Quantizzazione e campionamento

2.2 Il campionamento

Il campionamento è la discretizzazione del segnale analogico nel tempo. Dato che la descrizione del segnale è data dalla variazione dell'ampiezza nel tempo, il campionamento corrisponde all'individuazione dei valori discreti (istanti di tempo) sull'asse delle ascisse. Intuitivamente, perché il segnale rappresenti in modo efficace il segnale analogico di partenza, occorre che i campioni siano prelevati a una velocità abbastanza elevata da riuscire a tener conto di tutte le variazioni del segnale. Cioè se il segnale passa in breve tempo da una fase ascendente a una fase discendente e nuovamente a una ascendente, gli istanti di tempo considerati devono appartenere a tutte e tre le fasi; se una delle fasi viene saltata il segnale digitale non potrà rendere conto di tale variazione. Quindi, la velocità di prelevamento dei campioni, cioè il tasso di campionamento, deve essere una

variabile dipendente dalla velocità di variazione del segnale, che a sua volta dipende dalla parziale con frequenza più alta. Più brevi sono gli intervalli di tempo tra un campione e il successivo, più simile all'originale analogico sarà il segnale a gradini, al limite potranno coincidere.

L'intervallo di tempo tra un campione e il successivo è detto *periodo di campionamento*; l'inverso del periodo, e quindi il numero dei campioni in un secondo è detto *tasso o velocità di campionamento*.

A ogni periodo di campionamento si preleva un campione dell'ampiezza del segnale, cioè il valore quantizzato del segnale analogico in quell'istante; il risultato è una sequenza di valori che corrispondono all'andamento del segnale. Più sono ravvicinati i valori, più efficacemente sarà descritto l'andamento del segnale.

Poichè man mano che si riduce la dimensione dell'intervallo di campionamento aumenta l'accuratezza della descrizione del segnale, è cruciale scoprire qual è la dimensione migliore dell'intervallo. I due parametri per i quali occorre trovare un equo compromesso sono la qualità del suono e il numero di campioni. Una migliore qualità del segnale corrisponde a una maggiore qualità di informazione, cioè a un maggior numero di campioni nell'unità di tempo; un maggior numero di campioni richiede un maggiore spazio in memoria e una superiore velocità di trasferimento dati. Nella digitalizzazione i due parametri sono in contrasto e la scelta del tasso di campionamento dipende esclusivamente dalle esigenze delle applicazioni individuali.

2.3 La quantizzazione

L'operazione di quantizzazione sul singolo campione di ampiezza ha l'obiettivo di assegnare al campione uno dei valori numerici che sono consentiti nella codifica digitale. La codifica digitale divide la gamma della ampiezze possibili in intervalli o regioni: ogni campione ha un'ampiezza che cade in una delle *regioni*. Ad esempio, se l'ampiezza varia tra -5 e +5 volt, la gamma totale delle ampiezze è di 10 volt; se la codifica digitale divide la gamma delle ampiezze in 16 regioni, ogni regione sarà ampia 0.625 volt ($0,625 = 10/16$). Il numero di regioni in cui è suddivisa la gamma delle ampiezze dipende dai bit a disposizione per la codifica. La rappresentazione digitale, infatti, si basa sulle sequenze di *bit*. Il bit è l'unità minima di informazione, e può valere 0 o 1; una sequenza di bit è detta *parola binaria*; i valori tipici per la lunghezza di una parola binaria nell'audio sono

8 o 16 bit. Fissato il numero di regioni possibili mediante i bit a disposizione, occorre stabilire un metodo per associare una configurazione di bit a una regione della gamma delle ampiezze. Un metodo comune è la *quantizzazione lineare*, che divide la gamma delle ampiezze in 2^N regioni uguali, e associa a ciascuna delle regioni uno dei 2^N valori numerici.

Nel processo di quantizzazione si perde informazione, in quanto tutti i valori che rientrano in una regione verranno convertiti nello stesso codice binario, attenuando quindi le differenze tra i valori di tensione, questo è chiamato *errore di quantizzazione*.

L'errore di quantizzazione si può certamente considerare casuale ed è del tutto scorrelato dal segnale vero e proprio: si comporta come un segnale aggiunto, affine al rumore «bianco». Poiché non possiamo andare all'infinito nell'aumentare il numero di bit della codifica digitale, occorre fissare un criterio per determinare se la rappresentazione di un segnale è adeguata. Si può introdurre una misura che confronta l'ampiezza del segnale con l'ampiezza del rumore di quantizzazione. Si può usare l'acronimo SQNR (*Signal-to-Quantization-Noise-Ratio - Rapporto Segnale-Rumore di Quantizzazione*).

2.4 Struttura di un generico audio codificatore

La Figura 2.2 mostra un diagramma della struttura di un generico codificatore audio percettivo. La figura 2.2 (a) mostra la struttura del codificatore, che ha tre fasi principali e una quarta fase di formattazione del bitstream e la figura 2.2 (b) mostra il decoder, che ha invece solo tre fasi.

Il codificatore opera sul segnale audio in ingresso e in uscita sul bitstream codificato, e il decoder opera sul bitstream codificato e ricostruisce il segnale originale.

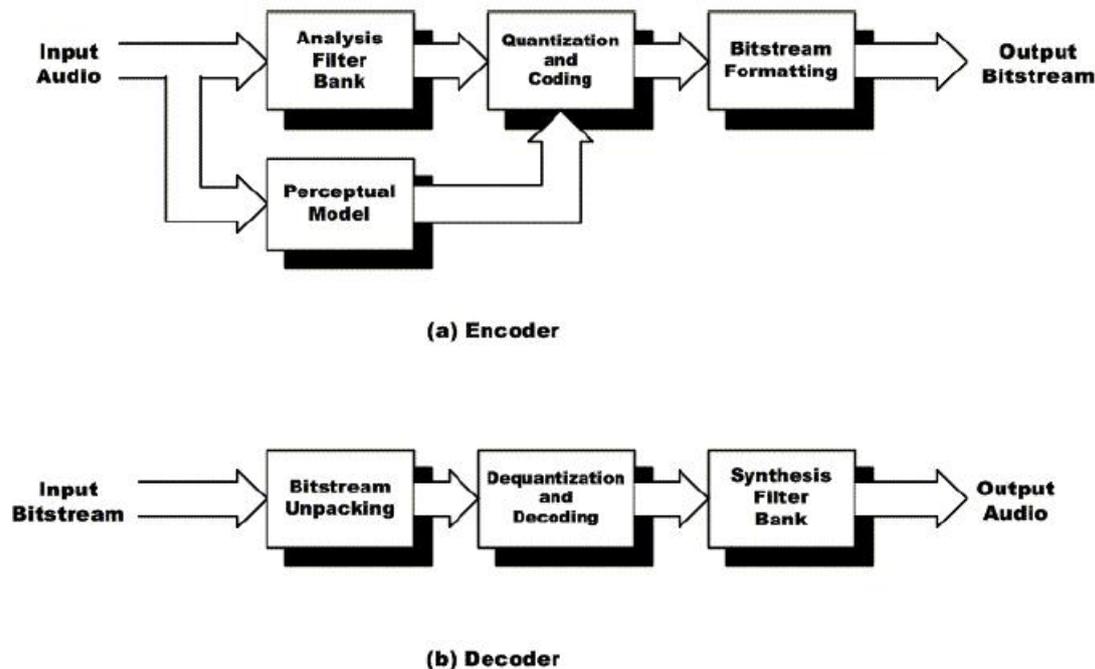


Figura 2.2: Un generico codificatore percettivo: a)encoder b)decoder

Le tre fasi del decoder, come risultato, sono operazioni inverse alle tre fasi dell'encoder.

Vale a dire, le fasi di analisi, quantizzazione e codifica, e formattazione del bitstream dell'encoder corrispondono alla sintesi del segnale, de-quantizzazione, decodifica, ed estrazione del bitstream del decoder, rispettivamente. La fase extra nel codificatore è il modello psicoacustico (cap. 3), che non è necessario nel decodificatore, da cui l'informazione viene implicitamente codificata. Ciò significa che i codificatori percettivi sono asimmetrici, in quanto il codificatore ha una maggiore esigenza di calcolo del decoder, il che può essere molto utile in certe applicazioni in cui un "server" codifica il segnale per molti "client".

Il codificatore funziona come segue. Il segnale di ingresso è di solito suddiviso in blocchi contigui o «frame» in modo che ogni blocco possa essere trattato singolarmente. Questo viene fatto per un serie di motivi. In primo luogo, l'elaborazione di un segnale in termini di segmenti più piccoli riduce la richiesta computazionale e il carico di memoria. In secondo luogo, la segmentazione serve a localizzare il segnale nel tempo in modo che il mascheramento nel dominio della frequenza possa essere applicato ad un segnale

localizzato nel tempo. E in terzo luogo, il flusso di bit codificati può essere inviato come "pacchetto" che può essere trasmesso, decodificato, e rappresentato su una base in tempo reale. All'interno del codificatore, l'input prima entra nel banco filtri e poi nella fase di psicoacustica. La fase banco filtri (vedi paragrafo seguente) trasforma il segnale in una rappresentazione nel dominio della frequenza o in una rappresentazione congiunta tempo-frequenza. Il modello psicoacustico applica prima una risoluzione ad alta frequenza per trasformare e poi applica le regole di psicoacustica per calcolare la frequenza del dominio della soglia di mascheramento. L'output sia del banco filtri sia della psicoacustica poi entra alla fase di quantizzazione e codifica, in cui si verifica la riduzione del bitrate. Nel caso della codifica e nella fase di quantizzazione viene deciso come sono ripartire i bit tra i coefficienti del banco filtri e viene usato un quantizzatore per (ri)quantizzare i coefficienti del filtro. A volte, in questa fase di eliminazione si usa un ulteriore passo di codifica senza perdita di dati (lossless) per ridurre ulteriormente la ridondanza statistica. I coefficienti di quantizzazione, insieme ad alcune informazioni secondarie, sono infine formattati nel flusso di bit in uscita.

Il decoder, sul lato di ricezione, semplicemente esegue le operazioni inverse.

Sebbene la maggior parte dei codificatori percettivi segue questa struttura di base, alcuni sono difficili da descrivere con questo modello semplice e chiaro. Tuttavia, tutti i codificatori percettivi hanno queste quattro tappe fondamentali e questo quadro di base può essere utile per comprendere altri approcci esistenti.

2.5 I banchi filtri

I banchi filtri sono sistemi che scompongono il segnale in varie "bande di frequenza". Essi sono dei circuiti costituiti da filtri passa-basso, passa-banda e passa-alto, combinati secondo opportune architetture, atti a decomporre lo spettro del segnale d'ingresso in un certo numero di bande contigue. La parte che effettua la decomposizione in sotto-bande è denominata banco di analisi, mentre la parte che effettua la ricomposizione in un unico segnale nel dominio temporale è denominata banco di sintesi. La scelta del banco filtri ottimale per lo sviluppo di codificatori percettivi, storicamente, è stata oggetto di molte ricerche e di discussione. Tra il tempo e la risoluzione di frequenza vi è sempre un trade-off inerente. I banchi filtri che hanno risoluzione ad alta frequenza, come accade per esempio nella Trasformata Discreta di Fourier (DFT), hanno bassa

risoluzione temporale, mentre i banchi di filtri ad alta risoluzione temporale, ad esempio, il banco filtri 2-channel QMF, hanno risoluzione a bassa frequenza. Una codifica efficace dipende in larga misura dal modo in cui la scomposizione tempo-frequenza del banco filtri è ottimizzata per le esigenze del segnale in ingresso. Si è constatato che non vi è una singola risoluzione che soddisfa i requisiti di tutti i segnali audio. Questo perché i segnali audio variano notevolmente in termini di caratteristiche tempo-frequenza nel tempo e tra i segnali stessi. La figura 2.3 fornisce esempi di due strumenti musicali le cui soglie di mascheramento sono indicate nel piano tempo-frequenza. I due strumenti, le nacchere (a), che sono uno strumento a percussione atonale e il piccolo (b), che è uno strumento a fiato acuto, sono in un certo senso diametralmente opposti. Si noti che l'energia del piccolo è distribuita con frequenza ad alta risoluzione che rimane sostanzialmente invariata nel tempo, mentre le nacchere sono localizzate nel tempo, ma si espandono nella frequenza. Chiaramente, abbiamo bisogno di una migliore risoluzione di frequenza per il piccolo e una migliore risoluzione temporale per le nacchere. Da questo esempio si deduce quanto detto prima, ovvero che ogni singolo segnale necessita di uno studio a sè stante; nel paragrafo seguente si andranno quindi a esplicitare i parametri di cui tener conto nella scelta della risoluzione ottimale.

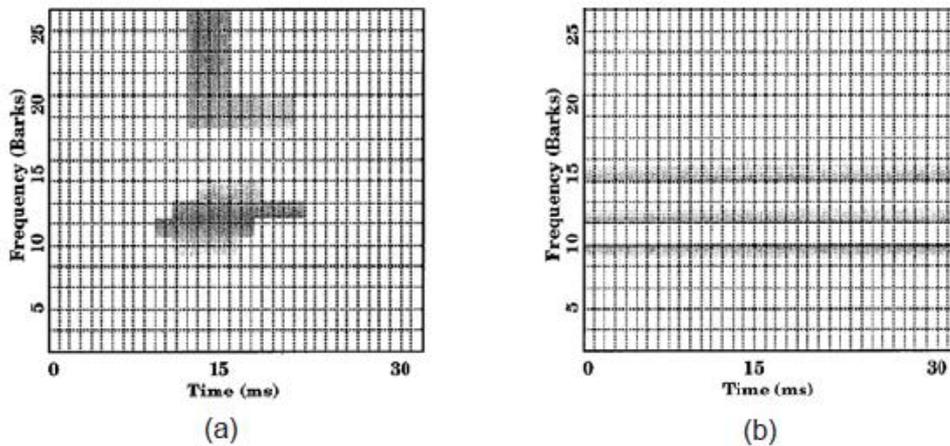


Figura 2.3: Esempio di soglie di mascheramento di un segnale nel piano tempo-frequenza (a) nacchere (b) piccolo

2.6 Esempi di banco filtri

La maggior parte delle trasformate tempo-frequenza e dei banchi filtro utilizzati nella codifica audio oggi, sono inserite nel quadro generale dell'elaborazione multirate del segnale. In questo modo abbiamo un metodo di confronto e di progettazione di vari tipi di banchi filtri in cui ognuno fornisce una diversa serie di vantaggi e svantaggi. In generale, non possiamo progettare un banco filtri che fornisca tutte le proprietà benefiche, ad esempio PR, campionamento critico, buona localizzazione tempo-frequenza, risoluzione tempo-frequenza flessibile, gestione trasparente del confine, e bassa complessità, descritti in precedenza. Esiste sempre però un trade-off tra alcuni parametri, in ogni dato del banco filtri. Alcuni esempi di banchi filtri si trovano comunemente nella letteratura della codifica audio e sono riportati qui di seguito.

- Trasformata di Fourier basata sul banco filtri: alcuni dei primi codificatori audio di alta qualità erano basati sulla Trasformata di Fourier Discreta (DFT) e sulla Transform a discreta del coseno (DCT). Inizialmente furono introdotte per codificare la voce. La trasformata di Fourier cerca di ottenere in maniera relativamente semplice una rappresentazione nel dominio della frequenza e decorrelare il segnale di ingresso. In generale, la trasformata ad alta frequenza fornisce una bassa complessità computazionale, ma manca della risoluzione temporale che a volte è necessaria.
- Filtri quadratura a specchio (QMF) basato su banchi filtri: i banchi di filtri QMF furono introdotti da Croisier, Estaban e Galand nel 1976. Un banco filtri QMF fornisce una divisione a due vie che potrebbero essere utilizzate in cascata insieme per dividere lo spettro delle frequenze in modi diversi. I filtri QMF originali sono soluzioni near-PR con cancellazione dell'alias, come ad esempio i filtri di Johnston QMF, ma sono state successivamente trovate soluzioni PR, per esempio i filtri di quadratura coniugati.
- Modified Discrete Cosine Transform (TCMD): la TCMD è uno dei banchi filtri più diffusi, che combina un elenco di funzionalità che la rendono particolarmente attraente per la codifica dell'audio. Conosciuto anche come "modulated lapped transform" (MLT) e trasformata "time domain aliasing cancellation" (TDAC),

la TCMD è un banco filtri a coseno modulato che fornisce PR ad alta risoluzione di frequenza, alto guadagno di codifica, ed eliminazione di artefatti dei blocchi attraverso la sua struttura LAPPED, il campionamento critico, la procedura semplice di disegno con un solo filtro di prototipo, e l'attuazione efficace attraverso un algoritmo simil-FFT. La TCMD è utilizzata in una serie di codificatori percettivi, compresi i formati MPEG-1 Layer III, MPEG AAC, AC-3, ATRAC, e PAC. 4.

- Banco filtri Wavelet: il banco filtri wavelet è strettamente legato ai due canali dei banchi filtri QMF e può essere utilizzato per garantire una distribuzione flessibile dello spettro di frequenze. Una descrizione approfondita e l'esame del WFB è data nel capitolo 4.

2.7 Sintesi dei problemi di progettazione

Una sintesi dei problemi di progettazione per i banchi filtri può essere la seguente.

- Risoluzione tempo-frequenza: come descritto in precedenza i banchi filtri hanno necessità di considerare la risoluzione tempo-frequenza dell'apparato uditivo, la risoluzione che fornisce la maggior quantità di guadagno di codifica e la risoluzione che fornisce un adeguato controllo degli artefatti temporali. In generale, nessun singolo banco filtri fisso può compiere tutti i requisiti di cui sopra ed è quindi necessario un regime di adattamento.
- Separazione dei canali: la separazione dei canali, o localizzazione di frequenza, si riferisce a quanto bene un canale di banco filtri è separato dagli altri canali. Si verificano sempre delle perdite di inter-banda nella pratica dei banchi filtri a causa della natura non ideale delle operazioni di filtraggio.
- Filtraggio: Il canale di separazione è importante nel contesto della codifica percettiva poiché i canali adiacenti si presume siano indipendenti e che non si sovrappongono quando si applicano i risultati della percezione. Ciò è particolarmente importante per le componenti tonali del segnale, che richiedono la risoluzione ad alta frequenza, ma anche buona frequenza di separazione. E' stato trovato che la

quantità di separazione dei canali, o la loro mancanza, ha un impatto diretto sulle prestazioni di codifica in un codificatore.

- Limite di movimentazione e gli artefatti bloccanti: la segmentazione del segnale in ingresso in blocchi più piccoli dà luogo agli artefatti bloccanti ai confini dei frame. Questo può essere percepito come una distorsione nel segnale ricostruito, in particolare per le porzioni di segnale che sono stazionarie. Per evitare gli artefatti bloccanti, il rumore di codifica deve essere adatto ad essere in qualche modo collocato ai confini dei frame. Per minimizzare le distorsioni di confine viene in genere fatta una sovrapposizione di finestra, oppure si utilizza una trasformazione lambita.
- Ricostruzione perfetta (PR): richiede che il segnale ricostruito sia identico al segnale di ingresso (con un possibile ritardo) in assenza di qualsiasi errore di codifica. Sebbene la condizione di PR non è un requisito rigoroso, in quanto vi sono codificatori che utilizzano un non-PR banco filtri, esso semplifica la progettazione generale di un sistema di codifica.
- Decimazione massima: i banchi filtri massimamente decimati, o criticamente campionati, forniscono lo stesso numero di coefficienti del dominio della trasformata di quanti ve ne sono nel dominio del tempo nel segnale originale, per esempio il numero di campioni in ingresso al secondo è uguale al numero dei campioni al secondo nel dominio della frequenza. Dal momento che l'obiettivo finale è quello di diminuire il tasso dei dati mantenendo un'alta qualità audio, i sistemi di campionamento sono da utilizzare in modo critico.
- Complessità computazionale: anche se è sempre meno un problema, la complessità computazionale e il ritardo di codifica possono essere ancora requisiti importanti per alcune applicazioni. L'analisi e banchi filtri di sintesi dovrebbe fornire implementazioni di algoritmi efficienti e veloci, ad esempio, FFT o DCT.

Capitolo 3

La Psicoacustica

3.1 Definizione

La psicoacustica è lo studio della percezione soggettiva umana dei suoni. Più precisamente è lo studio della psicologia della percezione acustica.

In molte applicazioni dell'acustica e dell'elaborazione del segnale sonoro diventa strettamente necessario conoscere come il suono viene percepito da un essere umano. Il suono, il cui stimolo acustico è composto da onde di pressione che si propagano attraverso l'aria, può essere misurato accuratamente tramite delle apparecchiature sofisticate. Tuttavia capire come queste onde vengano recepite e convertite in pensieri all'interno del nostro cervello non è affatto da sottovalutare: il suono è un segnale analogico continuo che (approssimando a zero il volume delle molecole d'aria) può teoricamente portare un infinito numero di informazioni (essendoci un infinito numero di frequenze portanti, ognuna contenente informazioni relative ad ampiezza e intensità).

Individuare le caratteristiche peculiari della percezione uditiva permette agli scienziati ed agli ingegneri di concentrarsi, per l'analisi e la progettazione di strumenti e apparecchiature acustiche, sulle componenti realmente udibili al fine di valutare quali parti del audio possono essere eventualmente scartate.

La maggior parte dei modelli esistenti della percezione umana sono basate sul lavoro sperimentale di Zwicker e Feldtkeller che è iniziato nel 1950, così come primi lavori da Fletcher e Helmholtz . Il campo della psicoacustica da allora ha compiuto progressi significativi nella caratterizzazione del processo di audizione e come l'orecchio analizza i

segnali audio nel dominio tempo-frequenza. I codificatori audio percettivi possono raggiungere tassi di compressione tanto elevati da identificare informazioni percettivamente irrilevanti nel segnale, applicando i principi della psicoacustica, quali la soglia assoluta di udienza e la diffusione di mascheramento. È importante sottolineare, inoltre, che ciò che "si sente" non è solamente una conseguenza di carattere fisiologico legata alla conformazione del nostro orecchio, ma comporta anche implicazioni psicologiche.

3.2 L'orecchio umano

Il sistema uditivo umano ha una struttura complessa e svolge funzioni notevolmente avanzate. Non solo è in grado di elaborare un ampio insieme di stimoli, ma può identificare precisamente l'altezza o il timbro di un suono, o la direzione da cui esso proviene. Molte funzioni del sistema uditivo vengono svolte dall'organo che chiamiamo orecchio, ma grande enfasi di recente viene attribuita all'elaborazione che ha luogo nel sistema nervoso centrale. Per semplificarne la descrizione, l'orecchio (figura 3.1) è spesso diviso in tre parti principali: orecchio esterno, orecchio medio ed orecchio interno. L'orecchio esterno è formato dalla pinna esterna e dal canale uditivo (meato). L'orecchio medio inizia con la membrana del timpano, alla quale è attaccato il primo dei tre ossicini (chiamati martello, incudine e staffa) che compongono questo stadio. Il compito di questi ultimi è quello di amplificare il moto del timpano (essi formano un sistema di leve) e di trasferirlo ad un'altra membrana, la finestra ovale. Con la finestra ovale inizia l'orecchio interno, formato principalmente dalla coclea. La coclea contiene i meccanismi per trasformare le variazioni di pressione in corrispondenza del timpano in impulsi nervosi che vengono interpretati dal cervello come suono.



Figura 3.1: Morfologia dell'orecchio

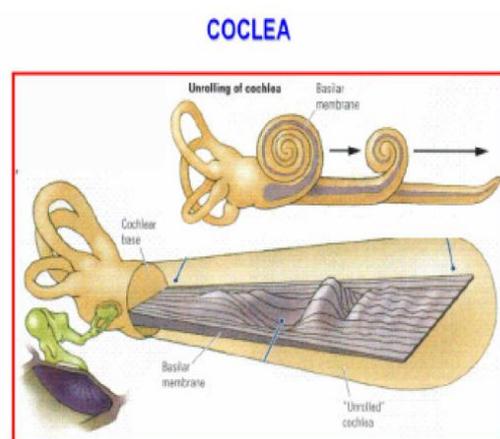


Figura 3.2: Rappresentazione della coclea

SEZIONE DELLA COCLEA

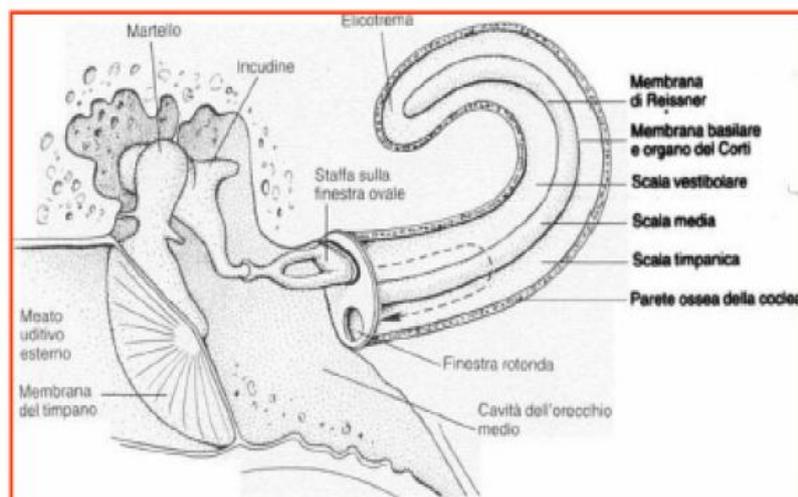


Figura 3.3: Sezione della coclea

Nel dettaglio, vediamo come la coclea (figura 3.2 e figura 3.3) ci aiuta a percepire l'altezza (pitch) di un suono puro. Questo organo è costituito da una cavità divisa in due da una membrana (membrana basilare) e contenente un liquido incompressibile (perilinfo) che passa da una parte all'altra della membrana attraverso un'apertura (elicotrema).

Le vibrazioni trasmesse dall'orecchio medio alla finestra ovale si trasmettono al fluido del dotto cocleare, che a sua volta provoca un moto della membrana basilare simile a quello di una bandiera. Lungo la membrana basilare, circa 30000 recettori nervosi (cellule ciliate) convertono il moto della membrana in segnali che sono a loro volta trasmessi ai neuroni del nervo acustico. E' importante sottolineare che, a fronte di un suono puro di una data frequenza, il massimo della ampiezza di oscillazione della membrana basilare è localizzato in una regione ben delimitata della membrana. La posizione di questa regione dipende dalla frequenza del suono. Per ogni frequenza c'è dunque una regione di massima sensibilità della membrana (regione di risonanza). Più bassa è la frequenza e più la regione di risonanza è prossima all'Apex (elicotrema). L'estensione delle frequenze udibili va da 16 Hz a 20 kHz. (figura 3.4).

FREQUENZE PERCEPIBILI

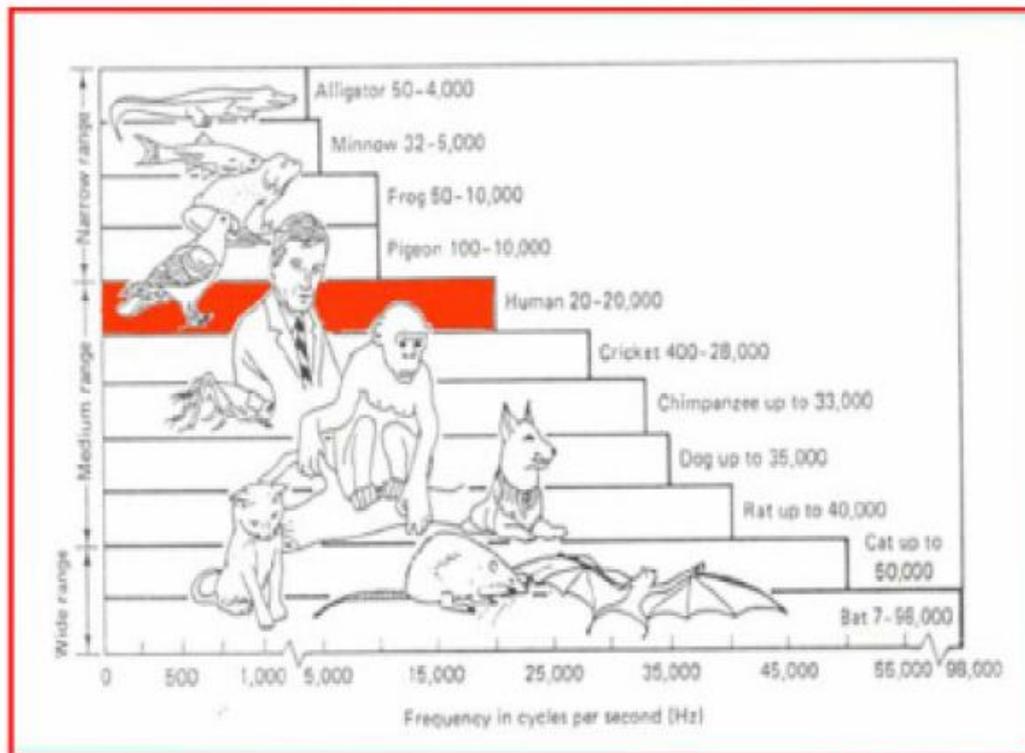


Figura 3.4: Le frequenze sonore percepite dagli esseri viventi

3.3 La psicologia dell'udito e caratteristiche fisiche del suono

La natura percettiva del suono è stata spiegata solo in parte. Alla fine dell'elaborazione effettuata dal nostro apparato uditivo è possibile percepire:

- Musica
- Linguaggio
- Rumore

Sentiamo un suono quando il timpano dell'orecchio viene eccitato da un'onda di pressione avente caratteristiche fisiche ben definite (come intensità o periodicità). La conseguenza per noi è una percezione del fenomeno costituita da diverse sensazioni che ci permettono di distinguere quel suono rispetto ad altri. Le tre sensazioni primarie che accompagnano l'ascolto di un suono sono: altezza, intensità e timbro (in inglese, rispettivamente pitch, loudness e timbre). Queste sensazioni sono il risultato della elaborazione dell'orecchio e del cervello e non sono grandezze misurabili direttamente (sono, invece, misurabili le quantità fisiche del suono che sono principali cause di altezza, intensità e timbro). Per suoni periodici (o quasi periodici), il pitch è determinato principalmente dalla frequenza fondamentale (si ricorda che la frequenza fondamentale di un suono è il numero di ripetizioni in un secondo del pattern di vibrazione). Tra i suoni periodici si distinguono i suoni puri, formati cioè da una sola componente sinusoidale. Visto nel dominio della frequenza, un suono puro è rappresentato con una riga in corrispondenza della frequenza della sinusoide. Per un suono composto da più armoniche (sinusoidi), la frequenza fondamentale è il massimo comun divisore della serie di frequenze che costituiscono lo spettro. L'intensità percepita dipende dal flusso di energia che accompagna la vibrazione. Essa è tuttavia dipendente anche da altri fattori quali pitch, durata e presenza di altri suoni.

3.4 I parametri della percezione

In generale, le relazioni esistenti tra le grandezze fisiche e le grandezze percettive sono molto complesse; tipicamente più di una grandezza fisica corrisponde a una singola grandezza percettiva con un contributo non sempre facilmente determinabile. Più che teorie scientifiche complete, che riescano a mettere assieme sia la parte fisiologica che quella cognitiva, si dispone attualmente di teorie che descrivono in modo sistematico l'intero corpus di studi empirici, stabilendo delle correlazioni tra grandezze fisiche e grandezze percettive.

Individuiamo innanzi tutto quali sono le grandezze percettive e i loro correlati fisici più immediati; vedremo quindi come la nostra percezione di una grandezza percettiva risulti dalla combinazione di più grandezze fisiche, con contributi differenti.

Le grandezze percettive con cui di solito si descrive un suono sono *volume*, *altezza*

e *timbro*. Il volume del suono è il correlato percettivo diretto dell'intensità sonora. La frequenza è il principale responsabile della sensazione di altezza di un suono (o *pitch*); infine, il timbro ha come correlato fisico diretto lo spettro o la forma dell'onda del suono.

3.4.1 Altezza

L'altezza è la qualità che fa distinguere un suono acuto da uno grave. Dipende in massima parte dalla frequenza ma anche dalla intensità. L'orecchio umano percepisce solo i suoni che vanno da 20 a 20.000 oscillazioni al secondo. Al di sotto abbiamo gli infrasuoni, al di sopra gli ultrasuoni. Il sonar, ma anche i delfini ed i pipistrelli percepiscono gli ultrasuoni mentre gli elefanti percepiscono gli infrasuoni. La pratica musicale copre una gamma di suoni, le cui fondamentali vanno dal do grave che ha circa 65 oscillazioni semplici al secondo al do acuto che ha 8276 oscillazioni semplici. La voce umana ha un registro ancora più limitato. Per calcolare l'altezza dei suoni, è stato scelto come punto di riferimento il La4 (= ottava centrale del pianoforte) che chiamiamo diapason o corista. La frequenza del diapason, che fino al XIX secolo variava di Paese in Paese e anche a seconda del tipo di musica da eseguire (sacra, da camera etc.) è stata determinata da diversi congressi: nel 1885, al Congresso di Vienna, si stabilì che il La4 corrispondesse a 870 oscillazioni semplici che, a loro volta, corrispondevano a 435 oscillazioni doppie. Ora invece il valore di riferimento, stabilito dalla Conferenza di Londra del 1939, è 440 vibrazioni doppie, quindi 880 semplici.

3.4.2 Timbro

Il timbro, è la qualità che, a parità di frequenza, distingue un suono da un altro. Il timbro dipende dalla forma dell'onda sonora, determinata dalla sovrapposizione delle onde sinusoidali caratterizzate dai suoni fondamentali e dai loro armonici. Dal punto di vista della produzione del suono, il timbro è determinato dalla natura (forma e composizione) della sorgente del suono e dalla maniera in cui questa viene posta in oscillazione. La scomposizione di un suono nelle proprie componenti sinusoidali fondamentali è detta analisi in frequenza. Le frequenze vengono misurate in Hz, ovvero oscillazioni al secondo. Le armoniche sono suoni con frequenze che sono multipli interi del suono principale. Nella musica, tanto più un suono è composto da diverse componenti, tanto più esso risulta complesso: si va dal suono di un flauto dolce, composto dalla fondamentale e da pochissime armoniche, al suono degli strumenti ad arco, composto da moltissime fre-

quenze armoniche secondarie. Tanto più le frequenze secondarie che si sovrappongono alla principale non sono armoniche (ovvero hanno frequenze che non sono multipli interi della fondamentale), tanto più ci si avvicina al rumore.

3.4.3 Volume

Il volume che viene spesso anche chiamato - colloquialmente ed erroneamente - pressione, è la qualità sonora associata alla percezione della forza di un suono, ed è determinato dalla pressione che l'onda sonora esercita sul timpano: quest'ultima è a sua volta determinata dall'ampiezza della vibrazione e dalla distanza del punto di percezione da quello di emissione del suono.

Per misurare il volume percepito di un suono, si fa spesso riferimento al livello sonoro, che viene calcolato, in decibel, come segue:

$$L_p = 10 \log_{10} \left(\frac{p^2}{p_0^2} \right) = 20 \log_{10} \left(\frac{p}{p_0} \right) \text{ dB}$$

nella formula, p indica la pressione quadratica media, mentre p_0 è una pressione sonora di riferimento (soglia del silenzio o udibile). È la più piccola variazione di pressione che l'orecchio umano è in grado di percepire) comunemente posta $20 \mu Pa$ ($2 \cdot 10^{-5} \text{ N/m}^2$) in aria e a $1 \mu Pa$ in acqua. La misura in decibel risulta più appropriata per indicare il livello sonoro percepito, perché la risposta dell'orecchio umano è all'incirca logaritmica.

3.5 Mascheramento uditivo

Il nostro apparato uditivo si comporta come un analizzatore di Fourier, e percepisce le componenti individuali di un suono distribuendole lungo la membrana basilare della coclea. Tuttavia il picco di vibrazione interessa una regione che, sebbene limitata, non è un punto infinitesimo sulla membrana; poiché ha invece una dimensione, più frequenze ravvicinate ricadono in una stessa regione, causando imprecisioni nella percezione delle singole componenti. Considerando la membrana basilare come un banco di filtri accordato su frequenze diverse, le capacità discriminatorie del nostro apparato uditivo non sono assolute, ma dipendono dalla acutezza dei filtri. Il fenomeno psicoacustico che ci consente di rilevare l'ampiezza di banda dei filtri uditivi lungo la membrana basilare è detto *mascheramento uditivo*.

Il mascheramento è quel fenomeno che si verifica nel dominio della frequenza, dove un segnale di piccola ampiezza può essere reso inudibile (mascherato) da un segnale

contemporaneo più intenso (mascherante) se i due hanno frequenze prossime tra loro. A causa della dipendenza in frequenza della capacità di risoluzione dell'udito umano, la soglia di mascheramento del rumore a qualsiasi frequenza dipende esclusivamente dalla potenza del segnale entro un limitato intorno di quella frequenza. Si può perciò definire una soglia di mascheramento e i segnali sotto questa soglia non risulteranno udibili. L'idea generale è che i suoni possono essere discriminati perfettamente dal nostro apparato uditivo solo quando ricadono in differenti bande critiche; quando ricadono nella stessa banda critica, la discriminazione diventa più difficile, ed è possibile solo in determinate condizioni.

3.6 Rumore (noise) ed effetti del rumore

I cosiddetti noise (rumori) possono essere trovati dappertutto in natura. Nel mondo della tecnologia e della scienza spesso è un fenomeno non desiderato. Il noise può essere comune o "ammorbido", qualche volta è descritto con un colore (es. rumore rosa, rumore bianco, ecc.). La caratteristica più importante del rumore è l'imprevedibilità. Il vero rumore naturale non può essere previsto completamente ma può essere simulato con varie formule matematiche. Qui di seguito sono riportati alcuni esempi di rumore classificato in base al colore, ma ne esistono molti altri:

- Rumore rosso. Un genere di rumore molto "basso" ottenuto con un pesante filtraggio Low-pass. Questo rumore suona come un rimbombo molto basso - un treno che passa nel sottopassaggio pedonale o un sistema di aria condizionata chiassoso. La definizione di rumore rosso non è precisa come quella del rumore bianco e rosa. Il termine è usato soprattutto per rumori a basso-pitch nella musica elettronica.
- Rumore bianco. Si intende con questa dicitura un rumore di ampiezza costante su tutto lo spettro di frequenza. In sostanza si tratta di un rumore termico solo che in questo caso si intende un rumore appositamente generato con finalità di test. Per vedere infatti il comportamento di un componente audio, per esempio di un canale di un mixer, si invia in ingresso un rumore bianco e si esamina il segnale di uscita. Generalmente in questo caso l'obiettivo sarà quello di ottenere un segnale in uscita mediamente costante a tutte le frequenze, questo significherà che il componente è

affidabile a tutte le frequenze. In generale il rumore bianco viene usato per i test sui componenti elettronici.

- Rumore rosa. Questo rumore ha una frequenza spettrale di $1/f$ con una caduta di 3dB per ottava e quindi uguale energia per ottava. Di solito il rumore rosa è prodotto con un filtraggio Low-pass del rumore bianco. Il rumore rosa avrà la stessa ampiezza nella banda da 110 a 220 Hz come da 220 a 440 Hz o da 7.040 a 14.080 Hz. Se si ingrandisce o si restringe l'onda con un oscilloscopio la forma d'onda ha sempre la stessa "trama". Il pink noise non è nè casuale nè prevedibile ed ha una struttura frattale. Il rumore rosa suona più naturale degli altri rumori ed è molto rilassante. Lo troviamo spesso in natura e suona come la pioggia, le cascate d'acqua, l'acqua che scorre nei vimini o dell'oceano praticando il surf e altri suoni naturali. Spesso si usa per "ambienti" musicali elettronici e come segnale test per "accordare" il rinvigorimento dei sistemi sonori (molti equalizzatori e analizzatori di spettro audio hanno generatori di rumore rosa).

Gli effetti fisiologici prodotti dal rumore sull'individuo possono variare in funzione, sia delle caratteristiche fisiche del rumore, quali l'intensità, la composizione spettrale, il tempo di esposizione, sia dalla risposta dei soggetti esposti. Gli effetti sono suddivisibili in tre classi:

- effetto di danno
- effetto di disturbo
- effetto di fastidio (annoyance)

Per danno si intende ogni alterazione anche parzialmente non reversibile dell'apparato uditivo. Il disturbo è invece un'alterazione reversibile delle condizioni psicofisiche dei soggetti esposti al rumore. Il fastidio è invece un effetto che il rumore provoca sugli individui non dovuto esclusivamente al rumore ma anche alla combinazione di fattori di natura psicologica e sociologica. La figura seguente (figura 3.5) ci mostra una tabella con alcuni danni uditi associati al livello di pressione acustica che esercitano.

Gamma di rumore	Livello di pressione acustica [dBA]	Caratteristica del danno uditivo
α	0 - 35	Rumore che non arreca né fastidio né danno
β	36 - 65	Rumore fastidioso e molesto, che può disturbare il sonno e il riposo
γ	66 - 85	Rumore che disturba e affatica, capace di provocare danno psichico e neurovegetativo e in alcuni casi danno uditivo
δ	86 - 115	Rumore che produce danno psichico e neurovegetativo, che determina effetti specifici a livello auricolare e che può indurre malattia psicosomatica
ϵ	116 - 130	Rumore pericoloso: prevalgono gli effetti specifici su quelli psichici e neurovegetativi
ζ	131 - 150 e oltre	Rumore molto pericoloso: impossibile da sopportare senza adeguata protezione; insorgenza immediata o comunque molto rapida del danno

Figura 3.5: Tabella danni uditivi associati al livello di pressione esercitata

Capitolo 4

Le funzioni Wavelet

4.1 Introduzione storica

Lo strumento principe nell'analisi dei segnali è stato, sino alla prima metà del secolo scorso, la trasformata di Fourier. E' curioso notare come lo sviluppo della teoria del grande matematico dell'800 prenda spunto da un campo abbastanza distante quale la termodinamica. L'introduzione delle Wavelet si può far risalire all'inizio del XX secolo ad Alfred Haar. L'approccio usato dalle due teorie non è molto diverso, ed è sempre basato sul trovare le funzioni più adeguate per rappresentare i differenti segnali offerti dai fenomeni naturali. A sfavore della teoria di Fourier c'è lo sfruttamento di seni e coseni come funzioni elementari, che sono sì maneggevoli e regolari, ma poco si adattano a seguire andamenti con cambi repentini, infatti, il fenomeno di Gibbs evidenzia molto bene le difficoltà che sorgono quando il segnale in esame è, per esempio, un'onda quadra (Figura 4.1). Nel caso dell'uso di wavelet, invece, la situazione cambia completamente: l'insieme delle funzioni elementari da cui attingere è ben più vasto e all'occorrenza si possono creare ad hoc.

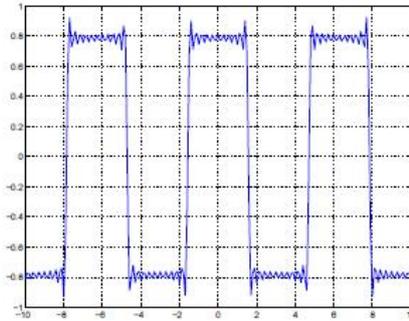


Figura 4.1: Rappresentazione di Fourier di un'onda quadra

Il caso di Haar citato prima è forse troppo semplice (la Wavelet in questione è un'onda quadra) ma si possono trovare anche esempi come questo (Figura 4.2):



Figura 4.2: Esempio di una wavelet

Il caso di Haar citato prima è forse troppo semplice (la Wavelet in questione è un'onda quadra) ma si possono trovare anche esempi come questo (Figura 4.2): anche agli occhi del profano una forma d'onda così fatta appare subito più adatta a seguire tutte le particolarità che si possono presentare. Non è questo l'unico punto di forza dell'analisi fatta con wavelet: se nel caso di Fourier il passaggio dal dominio del tempo a quello della frequenza comporta la perdita totale d'informazione spaziale ora, sostituendo il concetto di frequenza con quello di scala, invece di spostarsi da una frequenza ad un'altra, ci si sposta da una risoluzione più fine a una più grossolana. Si colgono così in un colpo solo sia i dettagli sia le informazioni che potrebbero essere definite di massima.

Il fatto che nel grafico di figura 4.3 le aree tracciate abbiano la stessa estensione è l'esemplificazione del tentativo della trasformata Wavelet di rendere meno rigida l'imposizione del principio di Heisenberg:

«Non è possibile determinare esattamente e contemporaneamente posizione e quantità di moto di un corpuscolo.»

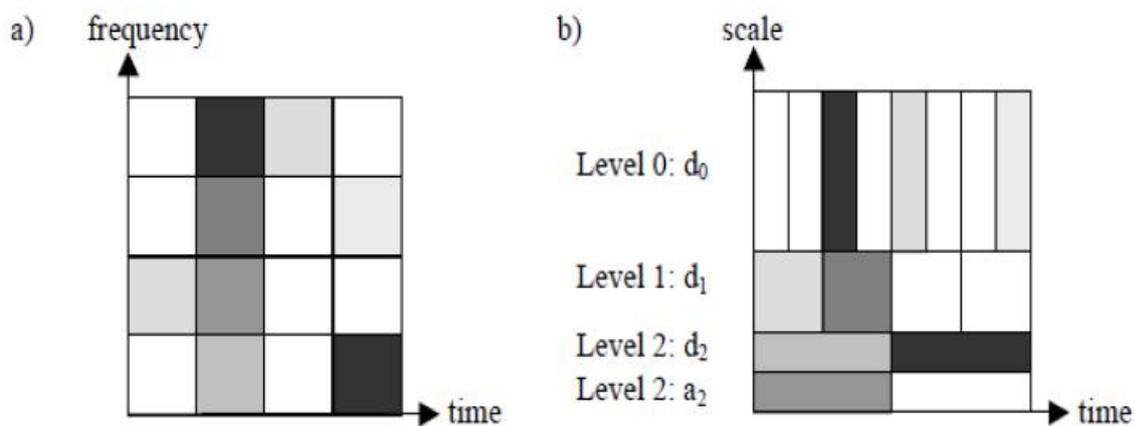


Figura 4.3: Rappresentazione tempo- frequenza e tempo-scala

Riportato in questo campo significa che non è possibile avere pari accuratezza dal punto di vista del tempo e della frequenza nell'analisi di un segnale. Per dare un'idea di come lavora la trasformata si prenda il segnale $x(t) = \sin(t)$ e in un punto a caso si inserisca una discontinuità impercettibile (per esempio a $t = \pi$). La trasformata di Fourier di questo segnale mostra solo un picco sulla frequenza interessata e anch'essa, come l'occhio umano, non riesce a evidenziare la piccola discontinuità (Figura 4.4).

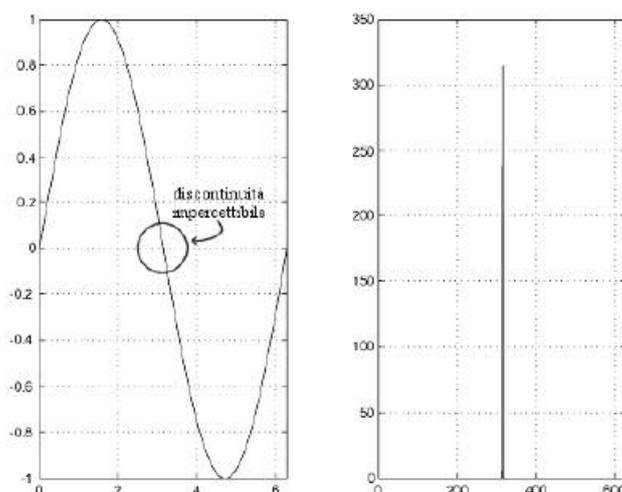


Figura 4.4: Una sinusoide imperfetta e la sua trasformata di Fourier

Usando invece la trasformata wavelet è possibile trattare separatamente i dettagli e le approssimazioni del segnale arrivando ad un'immagine di questo tipo (Figura 4.5). L'immagine delle approssimazioni (sulla sinistra) corrisponde alla visione definita «di massima» e infatti anche qui non vi è traccia dell'imperfezione (questa è una caratteristica utile quando per esempio si debbano analizzare dei trend), ma osservando i coefficienti dei dettagli (sulla destra) si nota che i valori in modulo maggiore sono posti esattamente nella posizione dove la discontinuità era stata in origine inserita, mantenendo quindi anche l'informazione temporale.

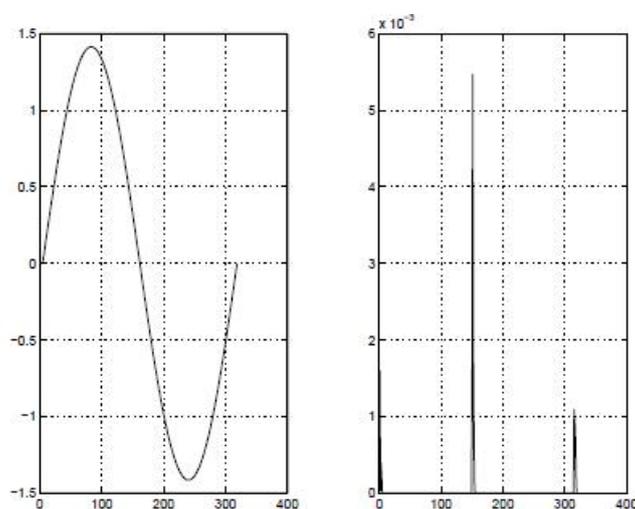


Figura 4.5: Approssimazioni e dettagli della stessa sinusoide

Non occorre molto per accorgersi della potenza di questo strumento: la lista che segue abbraccia solo una parte delle possibili applicazioni:

- Nel campo monodimensionale (segnali da \mathbb{R} a \mathbb{R})
 - eliminazione del fruscio dalle note musicali
 - studio di onde sismiche
 - studio dei trend finanziari

- Nel campo bidimensionale (segnali da \mathbb{R}^2 a \mathbb{R})
 - svariate applicazioni in campo medico
 - tutto ciò che concerne l'elaborazione d'immagine quindi rilevamento contorni e caratteristiche, compressione. . .

4.2 Le funzioni wavelet

Adesso spiegherò meglio cosa è davvero una wavelet e le teorie dietro di essa. In particolare spiegherò come lo schema del banco filtro è sufficiente per calcolare la trasformata wavelet.

Le wavelets sono funzioni continue nel tempo; esse possiedono speciali proprietà. In genere si utilizza il simbolo ψ per identificare la wavelet. Le funzioni possono assumere valori che vanno da $-\infty$ a $+\infty$, e l'integrale è sempre zero. Inoltre, una wavelet può avere al massimo un cambio di segno che rende la sua forma molto simile a una piccola onda (è proprio da questa proprietà che nasce il nome «wavelet», in italiano odina).

Il risultato che si ottiene dalla trasformata wavelet è una matrice contenente valori che indicano l'intensità di una certa frequenza in un certo tempo. Per poter ottenere questa matrice i principali algoritmi eseguono il prodotto tra la funzione in ingresso e l'ondina al tempo 0 (Figura 4.6), l'ondina viene così traslata di un tempo b e viene eseguito nuovamente il prodotto tra essa e la funzione in ingresso (Figura 4.7).

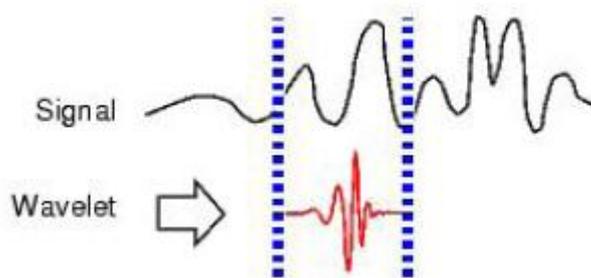


Figura 4.6: Traslo l'ondina nel tempo

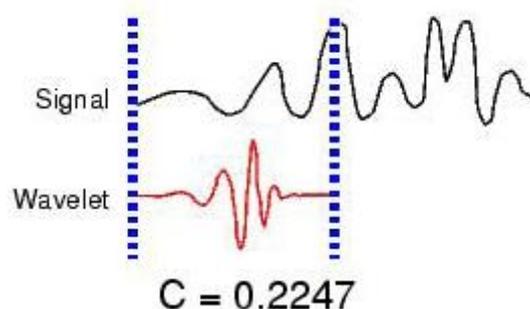


Figura 4.7: Ripeto il procedimento con una scala maggiore

L'ondina ha la scala minima. Le operazioni si concludono quando si arriva alla fine della funzione di ingresso. Così facendo riempio la prima riga della matrice contenente la soluzione. Dopo tale operazione l'ondina viene riscalata con una scala maggiore s e viene ripetuto il procedimento precedente. In tal modo si riempie la seconda riga della matrice. Il numero di riga indica la scala, mentre il numero di colonna il tempo. Il procedimento si ripete iterativamente fino a che si arriva alla scala massima (Figura 4.8).

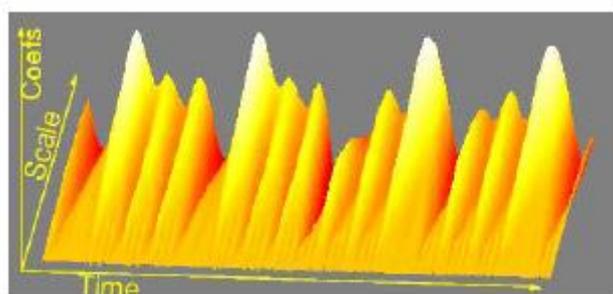


Figura 4.8: Grafico rappresentante la matrice della soluzione della trasformata wavelet

Questo procedimento richiede un'ingente quantità di operazioni. A volte anche le macchine più veloci risolvono i calcoli con uno sforzo non indifferente. La ricerca di nuovi strumenti matematici si è spostata verso una trasformata che svolgesse i calcoli in una maniera più veloce: la trasformata wavelet discreta.

I paragrafi seguenti illustrano le caratteristiche delle wavelet continue e discrete da un punto di vista matematico.

4.3 Funzione wavelet continua

Un approccio valido non solo nel campo delle wavelet è quello di modellare il segnale come una combinazione lineare di funzioni: per rappresentare il vettore v si consideri una base b per lo spazio vettoriale \mathbf{V} , quindi

$$\forall v \in V, v = \sum_{k=1}^n \mu_k b_k \quad (1.1)$$

dove k è il k -esimo coefficiente e b_k è il k -esimo vettore della base.

Passando alle funzioni si può usare lo stesso approccio:

$$f(t) = \sum_{k=1}^n \mu_k \phi_k(t) \quad (1.2)$$

dove in questo caso $\phi_k(t)$ è la k -esima funzione di base; se poi f e g soddisfano la condizione

$$\int_{-\infty}^{+\infty} |f|^2 \leq +\infty \quad (1.3)$$

ovvero $f \in \mathbf{L}^2(\mathbb{R})$, si può definire il prodotto scalare tra due funzioni (* denota il complesso coniugato):

$$\langle f, g \rangle = \int_a^b f(t)g^*(t)dt \quad (1.4).$$

A questo punto la definizione di CWT si vede come prodotto scalare tra la Mother Wavelet e il segnale $x(t)$ usando la (1.4):

$$CWT_x^\psi = \Psi_x^\psi(\tau, s) = \int_a^b x(t)\psi_{\tau,s}^*(t)dt \quad (1.5)$$

dove

$$\psi_{\tau,s} = \frac{1}{\sqrt{s}}\psi\left(\frac{t-\tau}{s}\right) \quad (1.6)$$

La 1.6 deve godere di una proprietà importante, chiamata ipotesi di regolarità:

$$\int_a^b \psi dt = 0 \quad (1.7)$$

cioè la wavelet generatrice deve avere valor medio nullo nel dominio del tempo, quindi un andamento oscillatorio (dall'inglese wave, onda).

4.4 Funzione wavelet discreta

La funzione wavelet discreta (DWT) è molto utile quando ci troviamo davanti a un segnale quantizzato, cioè conosciuto esclusivamente in precisi istanti di tempo; se si considera inoltre il desiderio di una valutazione numerica della trasformata, allora essa dovrà essere valutata solo con un numero finito di valori delle variabili a e b . La scelta non accurata dei parametri ovviamente può determinare da una parte una rappresentazione ridondante, ossia il numero di campioni della trasformata risulta molto più grande di quello del segnale, dall'altra parte si ha una rappresentazione non completa, cioè il numero di campioni non è sufficiente per la ricostruzione. La situazione ideale è la cosiddetta rappresentazione ortogonale, ovvero il segnale e la sua trasformata hanno lo stesso numero di campioni.

Si considerino le seguenti osservazioni:

- al variare della scala a le wavelet hanno banda relativa costante, e costante è l'indeterminazione

$$\frac{\Delta * f}{f}$$

sulla frequenza. E' quindi ragionevole effettuare una discretizzazione logaritmica di questo parametro.

- quando la scala è piccola le ondine hanno supporto temporale breve, quindi è logico che il passo di traslazione Δb sia piccolo per garantire una sufficiente copertura dell'asse-tempo da parte della serie di wavelet.

- viceversa, quando la scala è grande le ondine sono lunghe, per cui il passo di traslazione può essere maggiore.

4.5 Filter Bank Wavelet Transform

In questo paragrafo, è spiegata la rappresentazione del banco filtri della trasformata wavelet, leggermente accennato nel paragrafo 2.6. Questo tipo di approccio può apparire quasi come partire dalla fine, dal momento che storicamente le rappresentazioni di banche filtri sono nate dopo la trasformazione continua, il collegamento è stato fatto alla fine degli anni ottanta da S. Mallat. Tuttavia, il lettore ha già familiarizzato con filtri e segnali discreti, quindi questo approccio si integra meglio nella logica della tesi.

4.5.1 Wavelet filter banks

I banchi filtri Wavelet sono perfette ricostruzioni di banchi filtri.

Essi appaiono in alberi di filtri: nel primo livello, lo spettro delle frequenze è diviso in due parti, una metà inferiore e una superiore. Dopo il downsampling (o decimazione), questi possono essere divisi di nuovo, fino ad un determinato livello. La *wavelet packet tree* segue il modello in figura 4.9: ogni filtro decimato è diviso di nuovo in 2 segnali.

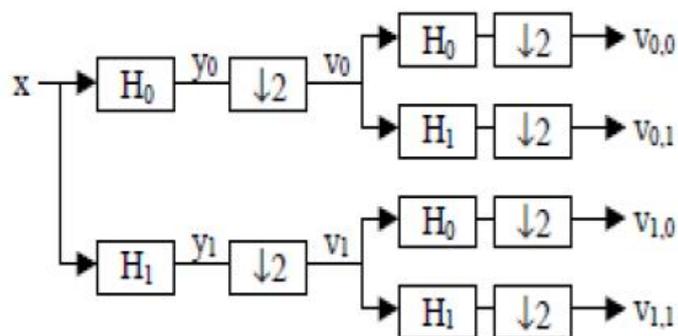


Figura 4.9: Primo livello di decimazione del banco filtro wavelet

La Fig. 4.10 mostra il 3-livello di decomposizione.

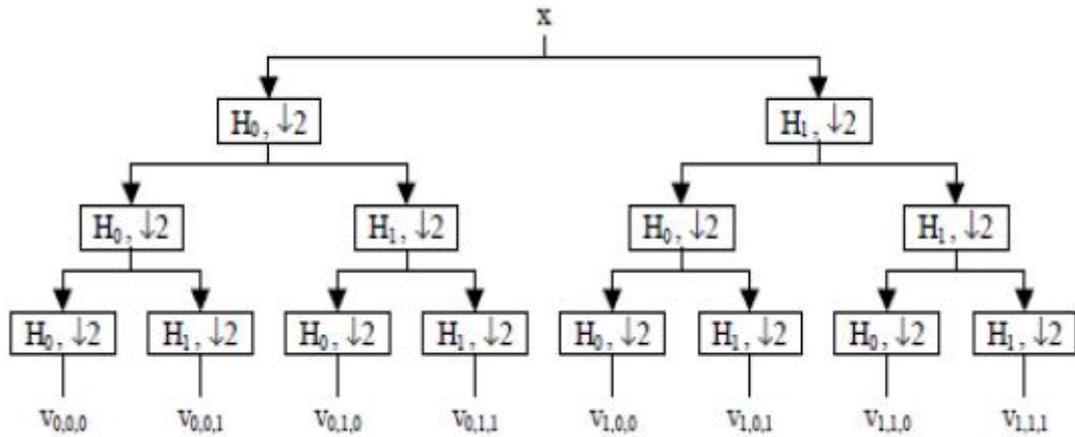


Figura 4.10: Terzo livello di decomposizione del banco filtro wavelet

Il secondo metodo è invece *l'albero wavelet*. Qui l'output pass-alto non è separato ulteriormente. Questo metodo si chiama algoritmo piramidale o algoritmo di Mallat (figura 4.11).

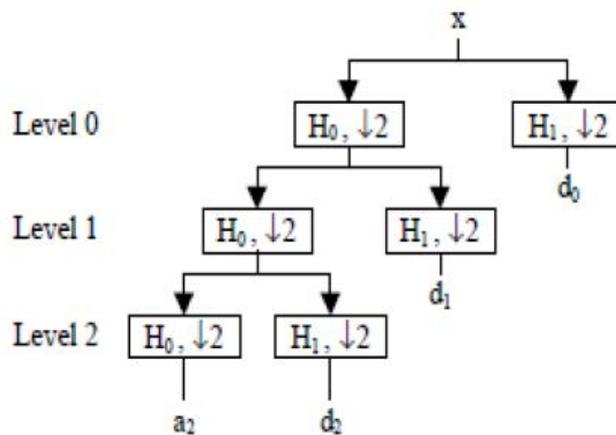


Figura 4.11: Esempio di albero wavelet

L'albero wavelet ha alcune importanti proprietà. I segnali di output che essa produce sono chiamati «dettagli D» e «approssimazioni a», e sono indicati come coefficienti wavelet. Le approssimazioni sono l'output decimato del filtro passa-basso e al contrario i dettagli provengono dal filtro passa-alto. In ciascun livello, le approssimazioni sono ulteriormente separate e sono mantenute solo le approssimazioni dell'ultimo livello. I numeri

dei livelli determina il numero dei coefficienti del dettaglio. d_0 , i dettagli del livello 0, ha la metà del numero dei coefficienti come il numero di campioni del segnale originale. Di conseguenza, d_1 ha un quarto del numero dei coefficienti, e così via. Occorre sottolineare che a questa ripartizione dei coefficienti corrisponde nello spazio delle frequenze (o spazio di Fourier) una ben precisa suddivisione della frequenze che compongono il segnale. Infatti a ciascun passo di decomposizione wavelet, che sottolineo ancora, si effettua sempre nello spazio reale, corrisponde una suddivisione in basse e alte frequenze nello spazio di Fourier, per cui d_0 contiene la metà superiore delle frequenze del segnale originale, d_1 contiene la gamma di frequenze da un quarto alla metà delle frequenze, ecc. Poiché si tratta di questa struttura, l'albero di wavelet divide in modo efficace il segnale in ottave. L'approssimazione dell'ultimo livello contiene le frequenze più basse del segnale.

In letteratura, spesso la dimensione dei livelli non è detta "frequenza", si chiama scala. L'uscita della WT è dunque un dominio tempo-scala. Come uno spettrogramma, i coefficienti della wavelet possono essere rappresentati in una griglia scala-tempo. La figura 4.12 mostra un confronto tra la rappresentazione della STFT e la WT.

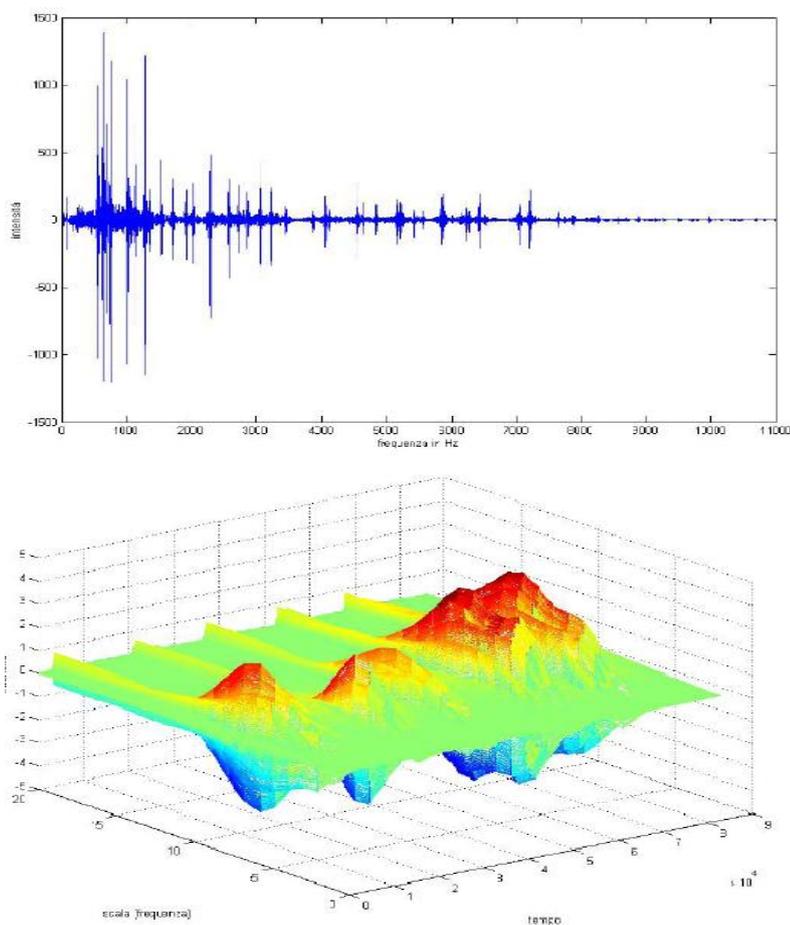


Figura 4.12: Grafico analisi di frequenza tramite FT e tramite CWT

4.5.2 Collegamento tra wavelet e banche filtro

Nella maggior parte dei casi la trasformata wavelet discreta viene realizzata attraverso dei banche di filtri. La funzione passa per un filtro passa alto ed uno passa basso (in questo caso sono filtri numerici) del corrispondente banco di filtri di ricostruzione. Per questo, è introdotta una funzione ausiliaria, la funzione di ridimensionamento f . Può essere calcolata utilizzando l'equazione di dilatazione:

$$\Phi(t) = 2 \sum_{k=0}^{+N} h_0(k) \phi(2t - k)$$

Poiché l'equazione dilatazione è ricorsiva a se stessa, non sempre vi è una soluzione per f . La funzione di ridimensionamento è una funzione a tempo continuo, ma è probabile che non sia continua, anzi potrebbe non essere regolare e contenere salti.

Infine, la funzione wavelet Ψ può essere calcolata a partire dalla funzione di ridimensionamento con l'equazione di wavelet

$$\Psi(t) = 2 \sum_{k=0}^{+N} h_1(k) \phi(2t - k)$$

Queste equazioni sono chiamate «two-scaling function» e sono direttamente individuate dei filtri h_0 e h_1 .

Il funzionamento dei filtri è esemplificato nella figura 4.13 e 4.14. Come si vede chiaramente se noi volessimo analizzare il frame d_1 dovremmo semplicemente seguire il percorso dei filtri (H, H, H, G) e troveremo il nostro livello di decomposizione. Allo stesso modo con il processo inverso si può ricostruire il segnale decomposto.

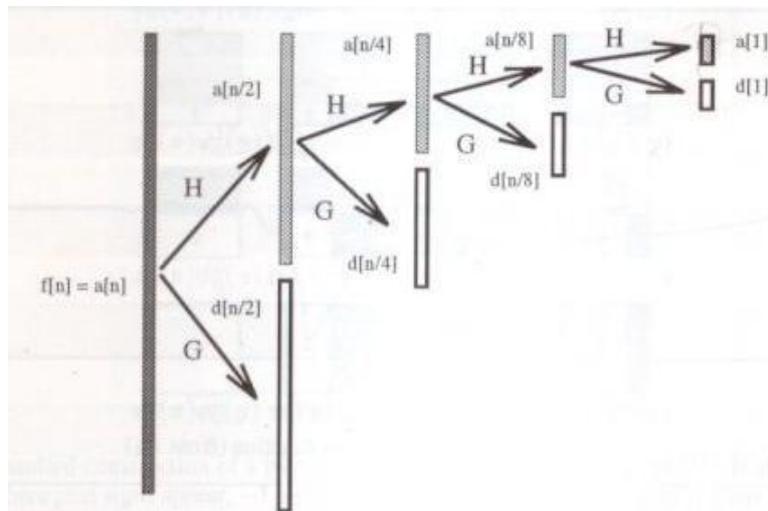


Figura 4.13: Scomposizione tramite filtri

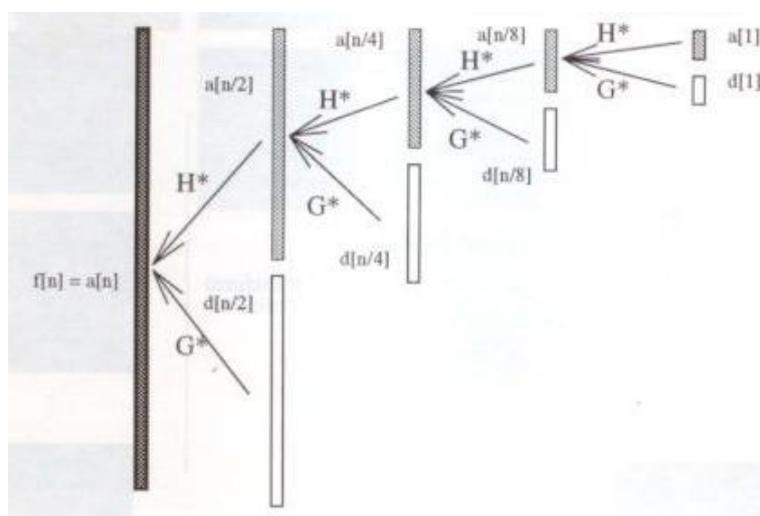


Figura 4.14: Ricomposizione tramite filtri

Una volta che la funzione di ridimensionamento è nota, la wavelet madre può essere calcolata direttamente senza ricorsione. La figura 4.9 mostra:

- a) la funzione di ridimensionamento
- b) la funzione wavelet
- c), d), e) ed f) i coefficienti del filtro
- g) le risposte in frequenza dei filtri di analisi per la Daubechies 2 wavelet.

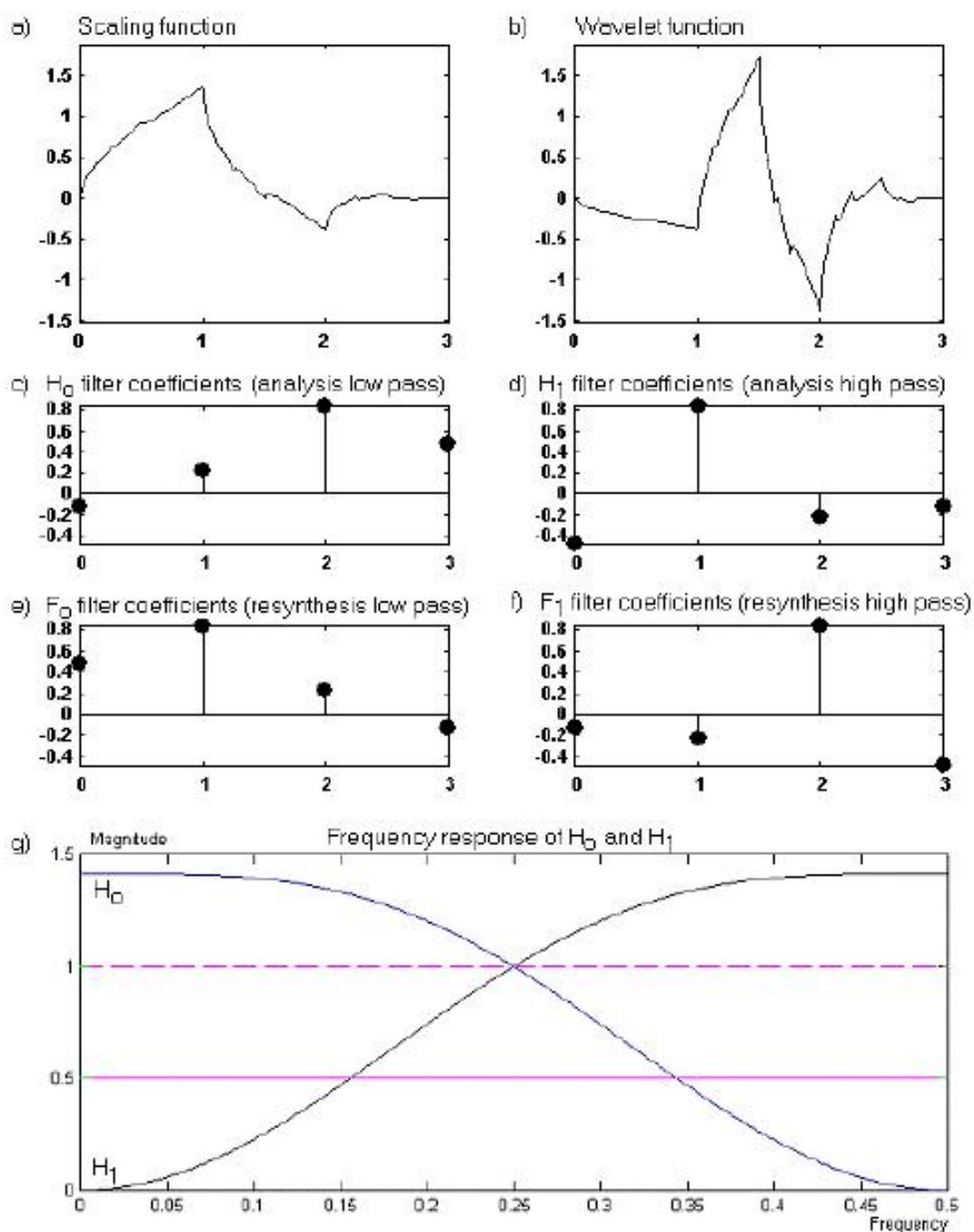


Figura 4.15: Immagine riassuntiva: a) la funzione di ridimensionamento, b) la funzione wavelet, c), d), e) ed f) i coefficienti del filtro e g) le risposte in frequenza dei filtri di analisi per la Daubechies 2 wavelet

La famiglia di wavelet Daubechies ha un'importanza enorme ed è stata sviluppata da Ingrid Daubechies. E' di ordine 2, pari ai coefficienti di 4 filtri.

4.6 Proprietà della trasformata wavelet

La WT è lineare. Ciò significa che la trasformazione della somma di due segnali è pari alla somma della loro trasforma. Inoltre, la trasformazione conserva l'energia, vale a dire che l'energia del segnale è uguale alla energia dei coefficienti. La proprietà locale dei chicchi consente la localizzazione degli eventi del segnale originale. Una wavelets approssima il segnale. Infatti la forma della wavelet ne determina la precisione. Un aspetto importante per la precisione è il numero di «vanishing moments» della funzione: una wavelet con p vanishing moments significa che è in grado di approssimare un polinomio di ordine $p - 1$. Così, più più è alto il valore di p , più i coefficienti di dettagli risultano chiari e più accuratamente la wavelet descrive il segnale. Inoltre le funzioni wavelet possono essere classificate in base ad alcune loro caratteristiche:

- esistenza dei filtri (e quindi lo schema di calcolo del banco filtro).
- ortogonalità: se il banco filtro è ortogonale o biortogonale.
- simmetria: la wavelet simmetrica e le funzioni di scala riportano ai rispettivi filtri simmetrici.
- supporto compatto: la funzione di scala a supporto compatto e la funzione wavelet riportano ai filtri finiti (FIR). Una wavelet a supporto non compatto dovrebbe avere un decadimento rapido in modo che i filtri FIR possano essere approssimati abbastanza bene.
- levigatezza: in molte applicazioni la scorrevolezza è importante, modificando i coefficienti wavelet si avrà come risultato un segnale in uscita liscio. La scorrevolezza è collegata alla regolarità della funzione wavelet.

4.7 Applicazione delle wavelet

Le wavelet sono utilizzate con successo per l'elaborazione di segnali. Come già più volte detto, il campo più evidente è la compressione dei segnali digitali. Abbastanza affermati sono gli algoritmi wavelet per l'elaborazione di immagini digitali. La capacità delle WT di estrarre le caratteristiche principali (più importanti per l'occhio e l'orecchio) si traduce in una compressione ad alta qualità senza grossa perdita. Sino al 2000 lo standard JPEG si basava sulla trasformata coseno, ora il nuovo standard impiega le trasformate wavelet. L'FBI ha adottato la compressione tramite wavelet per il suo personale archivio di immagini di impronte digitali. Le wavelets sono utilizzate con successo anche per la compressione video.

La riduzione del rumore funziona bene per ragioni analoghe: i coefficienti a bassa frequenza potrebbero contenere rumore ad ampio spettro incongruente. Impostando i coefficienti al di sotto di una certa soglia a 0, l'immagine può essere «denoised». Questo metodo funziona bene per i segnali audio e verrà spiegato più dettagliatamente in seguito. Altri campi di elaborazione del segnale, dove la WT è efficiente, comprendono l'individuazione delle singolarità o delle rotture, determinando l'evoluzione del segnale a lungo termine e il riconoscimento dello schema.

4.8 La WT per elaborare segnali musicali in tempo reale

E' stato anticipato che la WT offre alcune funzionalità particolarmente utili per l'elaborazione di segnali musicali.

La sua decomposizione multirisoluzionale offre un alto livello di localizzazione temporale per le alte frequenze mentre offre risoluzione ad alta frequenza per le basse frequenze. Un evento ad alta frequenza (ad esempio la rottura di un cembalo) sarà analizzato da alcune wavelet "veloci", brevi, e ad alta frequenza. Le note basse saranno analizzate dalle wavelets "lente", lunghe, e a bassa frequenza. In genere, queste operazioni non sono possibili con l'analisi di Fourier. La decomposizione logaritmica delle bande di frequenza assomiglia alla percezione umana di frequenze. La WT offre bande di frequenza pari logaritmicamente (ottave), mentre la FT ha una risoluzione logaritmica bassa per le basse frequenze.

La WT si presta molto bene all'analisi in tempo reale di un segnale perchè permette la localizzazione delle discontinuità nello spazio temporale. La WT necessita di un alto tasso di parametrizzazione, più della FT. Può essere calcolata in modo efficiente con l'algoritmo a piramide del banco filtro; il calcolo è in generale più lungo di quanto non lo sia il calcolo della FFT in quanto occorre scegliere in modo accurato il tipo di wavlet sulla base dell'indagine che si intende condurre. Tuttavia, è abbastanza veloce sia per l'analisi in tempo reale e sia per la risintesi di dati audio. Nel confronto con la FT, si può dire che la WT fornisce le proprietà, che si adattano meglio all'analisi e all'elaborazione in tempo reale di dati audio.

Capitolo 5

Le wavelet nella codifica audio percettiva

Il grande successo ottenuto dall'applicazione di wavelet nella codifica delle immagini, ha portato alcuni ad indagare anche la sua potenziale utilità nella codifica audio. Sono stati proposti molti tipi di schemi per le wavelet nella codifica audio, e come risultato hanno indicato che la WFB fornisce un modo interessante e potenzialmente utile di rappresentare e codificare dei segnali audio.

Nella codifica audio generalmente, un regione audio è composta da regioni o parti stazionarie o quasi stazionarie e da alcune parti non stazionarie e si verificano cambiamenti significativi generalmente a tutte le risoluzioni, cioè in tutta la gamma spettrale, e per tutto la risoluzione, cioè per tutta la gamma temporale. Si è cercato pertanto di adeguare la struttura della decomposizione wavelet in modo da ottimizzare la concentrazione dell'energia su base statistica. L'utilizzo di una strategia simile, tuttavia, potrebbe comportare l'adeguamento della struttura ad albero di decomposizione e del filtro di base in modo che la concentrazione di energia nel dominio wavelet possa essere ottimizzato su base statistica.

I codificatori percettivi d'altra parte, si basano su un criterio percettivo che viene calcolato da un modello psicoacustico in modo che la distorsione di codifica possa essere plasmato in maniera da ridurre al minimo la distorsione percepita. Applicare un criterio percettivo a una WFB dunque significa scegliere una struttura ad albero e una base del filtro, trovare modo in cui sono codificati i coefficienti di wavelet di ogni banda, infine

esaminare in primo luogo i requisiti percettivi del segnale audio e le caratteristiche del sistema uditivo umano. Inoltre, a differenza della codifica delle immagini in cui pensiamo alla decomposizione wavelet come un'analisi multirisoluzione, l'audio codifica percettiva tratta la WFB come un banco di filtri che fornisce una rappresentazione nel dominio del tempo-frequenza del segnale di ingresso

5.1 La simulazione in MATLAB

Ho analizzato un codice proposto da Matias Zanartu, uno studente della Purdue University (Indiana, U.S.A), che quanto si prestava molto bene alle mie necessità e ai miei scopi dimostrativi. L'applicazione è stata implementata usando MATLAB 7.2 provvisto di «Wavelet Toolbox». Questa simulazione non intende replicare un particolare algoritmo ma vuole solo mostrare alcune idee esplicitate in questa tesi nella pratica.

5.2 Principali caratteristiche dell'implementazione

Il software utilizzato può essere così schematizzato:

1. suddivisione del segnale in frames di 2048 elementi
2. la decomposizione wavelet di ogni frame
3. compressione nel dominio della wavelet
4. applicazione del modello psicoacustico
5. quantizzazione non lineare dei coefficienti wavelet mediante modello psicoacustico
6. ricostruzione del frame nello spazio temporale
7. ricomposizione del segnale audio.

Il codice qui di seguito mostra la compressione wavelet di un file .wav. Lo schema è una versione semplificata del documento pubblicato da Deepen Sinha e Ahmed Tewfik "Low Bit Rate Transparent Audio Compression using Adapted Wavelet" su IEEE Trans. ASSP, Vol.41, Numero 12, Dicembre 1993. Il file audio deve essere allocato nella stessa

cartella in cui è situato il programma. Se si vuole provare il programma con un altro file audio, occorre solo cambiare il nome della variabile "file". Lo scopo finale del programma è quello di comprimere i file audio, tramite il modello psicoacustico, in modo tale da ridurre la grandezza e avere file meno ridondanti. La compressione è utile ad esempio se si pensa allo sviluppo del formato MP3 a partire dagli anni '90. L' algoritmo è progettato per ridurre notevolmente la quantità di dati necessari per rappresentare la registrazione audio ed essere riprodotta ancora come una fedele riproduzione del audio non compresso originale per la maggior parte degli ascoltatori. Una volta stabilito su quale file si voglia lavorare, scegliamo quale wavelet utilizzare (nel nostro caso ho scelto di usare la 'dB10'(wavelet proposta da Ingrid Daubechies con 10 «vanishing moments», individuata da 20 coefficienti scaling e 20 coefficienti wavelet). Imponiamo il livello di decomposizione (nel nostro caso ho scelto 5), la grandezza dei frame e settiamo il modello psicoacustico ad 'on' (nel qual caso fosse settato ad off, esso userà di default 8). Inizia a questo punto la fase chiamata "ENCODER". Con la funzione wavread viene letto il nostro file .wav e ci vengono restituite le alte frequenze di campionamento (Fs) in Hertz, e il numero di bit utilizzato per codificare i dati nel file (bits).La function 'wavdec' si occuperà di scomporre il frame usando la 'dB10'. Da Qui inizia la fase vera e propria di compressione attraverso 'dencmp' e 'wdencmp'. La prima funzione restituisce i valori di compressione, la seconda compie la compressione effettiva. A questo punto entra in gioco il modello psicoacustico, ampiamente trattato in questo progetto. Esso analizza e ispeziona lo spettro e trova i mascheratori di toni. Nella fase finale il 'DECODER' ricostruisce il segnale usando gli stessi frame contenuti in 'step' (ad eccezione dell'ultimo) e quindi crea un nuovo file audio compresso ascoltabile. Infine viene anche disegnata la compressione in una semplice immagine.

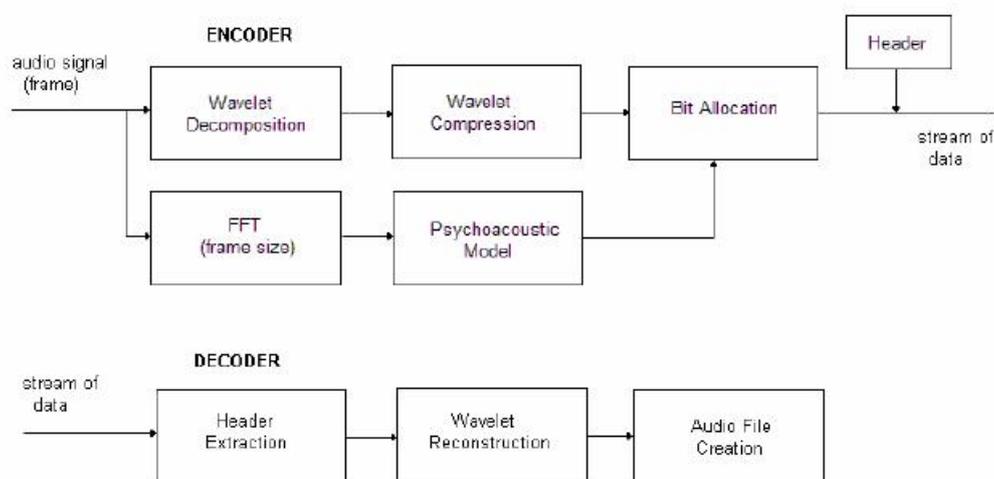


Figura 5.1: Diagramma a blocchi dell'implementazione in MATLAB

5.3 Considerazioni

Poichè sarebbe risultato molto complicato e talvolta ridondante ai fini della mia sperimentazione, sono state inserite alcune modifiche e considerazioni nella progettazione di questa simulazione MATLAB:

- Non viene eseguita la ricerca di una base ottimale: anche se questo è uno dei punti chiave, per la sua attuazione è necessario un modello di programmazione di grandi dimensioni. Per compensare a tale mancanza, è stata utilizzata un'altra tecnica di compressione basata sulla decomposizione della wavelet discreta nota che utilizza una soglia ottimale a livello mondiale. Questa tecnica è stata usata con successo per la compressione audio. I risultati migliori sono stati osservati quando si è utilizzata una wavelet Daubechies con 10 vanishing moments (DB10 in MATLAB), e 5 livelli di decomposizione.
- Non sono compresi riquadri sovrapposti: in questa implementazione non si sovrappongono i frame per evitare la complessità computazionale.

- E' stato semplificato il modello psicoacustico: data la complessità associata alla costruzione di un modello psicoacustico, è stata considerata una versione semplificata. Questo modello può solo rilevare i toni mascherati nel segnale e dà un limite generale per tutte le frequenze. Un esempio di questo modello psicoacustico semplificato è mostrato nella figura seguente, in cui vengono rilevati le principali componenti tonali. La potenza media di queste componenti viene utilizzata come soglia di mascheramento per ogni frequenza.

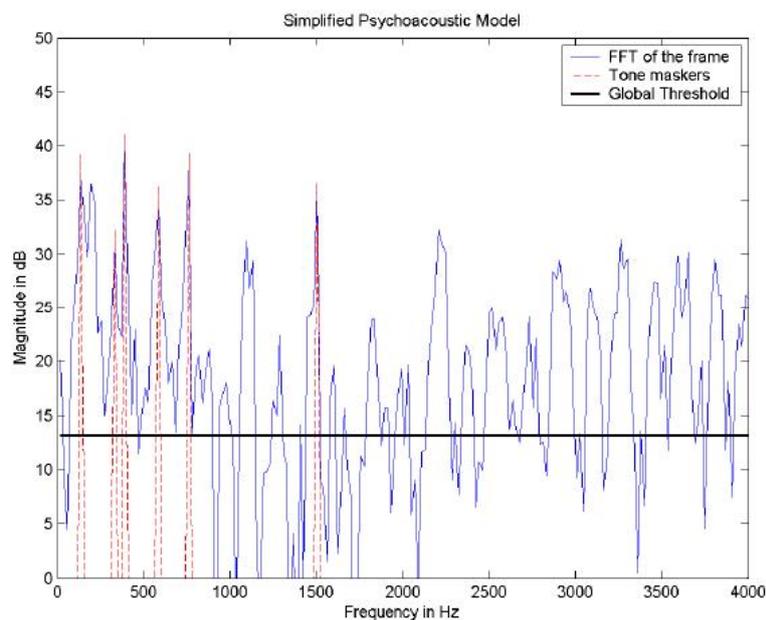


Figura 5.2: Rilevamento dei toni mascherati in ogni frame

-
- Non viene creato nessun nuovo formato audio: anche se questa implementazione matlab semplificata effettua una compressione sul segnale audio, questo non si riflette nella dimensione dei nuovi file audio. Ciò è dovuto al fatto che è stato utilizzato semplicemente il comando *wavwrite* per creare i file audio (.Wav).
- La compressione senza perdita è stata testata e abolita: rendeva il programma troppo lento . Tuttavia le sue prestazioni sono considerate nei risultati.

5.4 Risultati

Il mio studio fornisce alcuni risultati importanti: l'ottimizzazione della base wavelet per adattarla a risultati chiari sui dati audio porta ad una significativa riduzione del fabbisogno di bit rate. In generale, è vantaggioso utilizzare più wavelets, poiché queste forniscono una maggiore compressione. Tuttavia, non è necessario eseguire un'ottimizzazione in piena regola per individuare la migliore wavelet. Il risultato ottimale è stato l'utilizzare delle wavelet con un grande numero «vanishing moments» fornisce risultati ottimali. Sono attuabili diversi miglioramenti nel metodo proposto sia in termini di riduzione della complessità computazionale e sulle esigenze bit rate.

Come spiegato in precedenza, non sono stati utilizzati parametri oggettivi (come l'errore quadratico medio) per valutare la qualità dell'audio compresso. Pertanto i risultati di questa implementazione sono stati valutati utilizzando test soggettivi. Questi risultati devono essere considerati solo come un riferimento, perché non forniscono alcun supporto statistico. Il materiale utilizzato è composto da file audio monofonici. Facendo una valutazione generale, la qualità raggiunta da questa implementazione non è trasparente, con una distorsione sonora che ricorda il rumore tipico del sibilo di un disco in vinile. Tuttavia, la qualità raggiunta è migliore di quella fornita da una linea telefonica, ma peggiore rispetto a quella fornita da una radio FM, che in ogni caso è peggio della tipica qualità dell'mp3. Questo risultato è dovuto alle considerazioni e le semplificazioni che sono state prese in considerazione. Tuttavia, se confrontiamo i risultati di questo schema di compressione con quello ottenuto solo riducendo il numero di bit direttamente sul segnale audio (vale a dire "compressione al buio"), la qualità è notevolmente migliore. Per identificare il comportamento del sistema di compressione sono stati testati diversi file audio . I risultati della valutazione sono riassunti nella tabella seguente.

Stumenti - Generi musicali	Considerazioni
Sassofono	Nessuna rilevazione particolare
Orchestra classica	Nessuna rilevazione particolare
Flauto	Il suono del flauto si sente meglio
Piano	Si sentono meglio le medie e le alte frequenze
Chitarra elettrica	Si sentono meglio le medie e le alte frequenze
Pop - Vocal	Si sentono meglio le consonanti occlusive e fricative

Un altro metodo di valutazione della nostra implementazione è stato misurare di quanto è stato compresso il segnale e quali sono i tassi di bit risultante. Per fare questo abbiamo misurato la media bit per campione su ogni segnale di prova.

Infine, si può constatare che anche se la qualità di questa implementazione non è comparabile con gli attuali formati di compressione standardizzati, l'obiettivo è quello di comprimere il segnale in modo efficiente con un bit rate comparabile a quello del documento originale. E' inoltre necessario ricordare che questa implementazione Matlab può essere ulteriormente ottimizzata.

I grafici seguenti (figure 5.3, 5.4, 5.5) mostrano la compressione vera e propria. Inizialmente si farà fatica a notare delle differenze tra il grafico del segnale non compresso e il grafico ottenuto dalla compressione ma se si confrontano i valori del bitstream associandoli al grafico sarà più facile notare i cambiamenti.

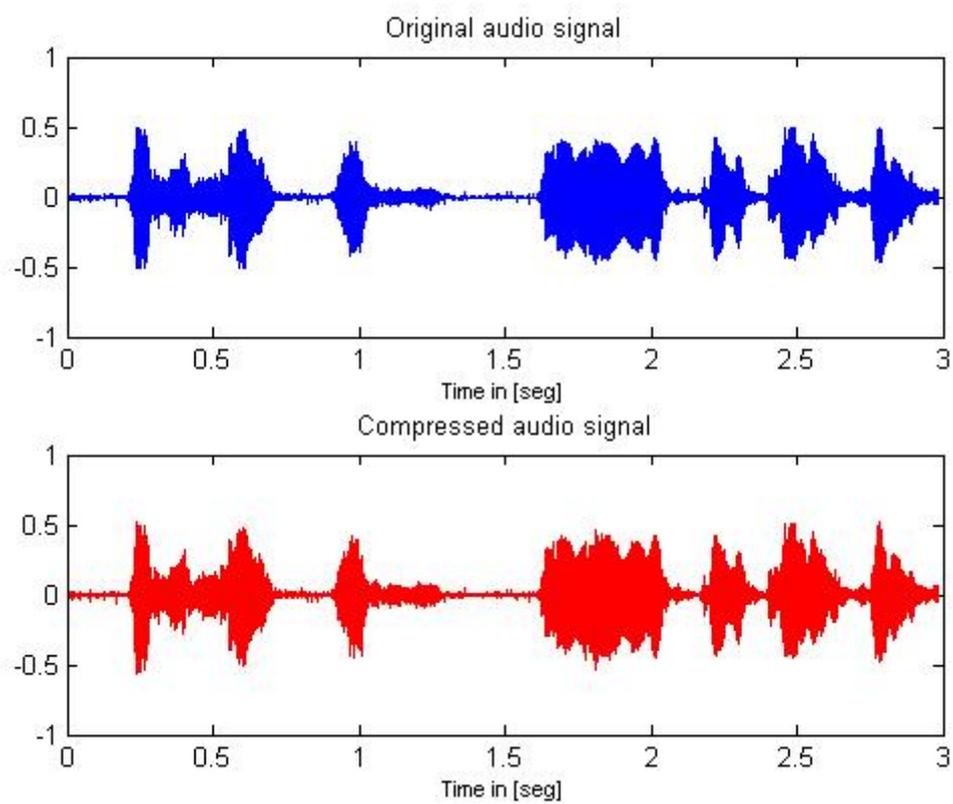


Figura 5.3: Grafico ottenuto dalla compressione di un file vocale

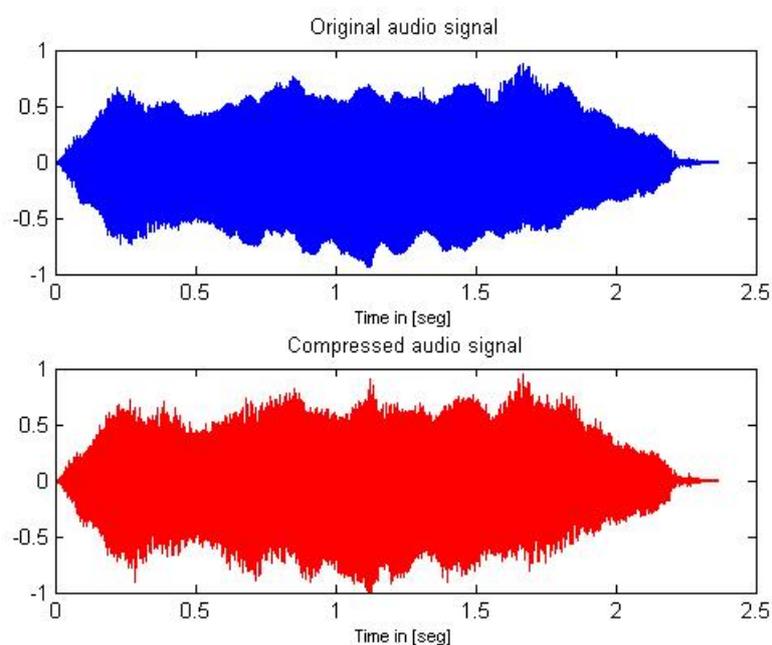


Figura 5.4: Compressione di un file con note di chitarra elettrica

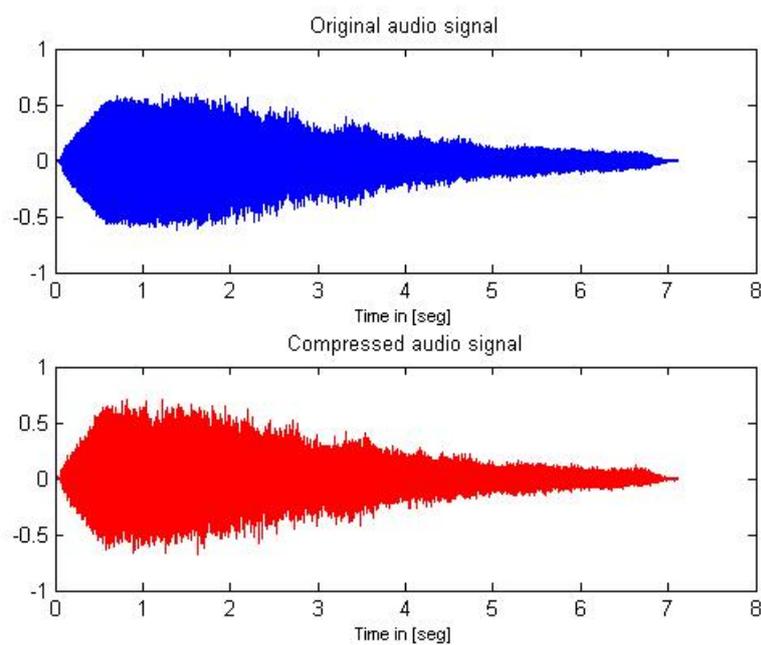


Figura 5.5: Grafico ottenuto dalla compressione di un file con note di pianoforte

5.5 Codice MatLab

```

%Final Project: MATLAB Code
%ECE 648 – Spring 2005
%Wavelet, Time–Frequency, and Multirate Signal Processing
%Professor: Ilya Pollak
%Project Title: Audio Compression using Wavelet Techniques
%Matias Zanartu – mzanartu@ecn.purdue.edu
%Purdue University
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%This one–file code performs wavelet
%compression over a .wav file.
% The scheme is a simplified version of
%the one described on the paper
%"Low Bit Rate Transparent Audio Compression
%using Adapted Wavelets"
%by Deepen Sinha and Ahmed H. Tewfik published in IEEE Trans.
%ASSP, Vol.41, No. 12, December 1993.
%
%NOTES: The file must be in the same folder where
%this file is located.
%If you want to try this scheme
%with other audio file , please change
%the name of the %variable "file".
%Avoid using long audio files or long
%silences at the beginning of the file
%for computational constrains.
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
clear;
clc;
file='beat_indian.wav';
wavelet='dB10';
level=5;
frame_size=2048;
psychoacoustic='on ';
%if it is off it uses 8 bits/frame as
%default wavelet_compression = 'on ';
heavy_compression='off ';
compander='on ';
quantization = 'on ';

```

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% ENCODER %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
[x,Fs,bits] = wavread(file);
xlen=length(x);
t=0:1/Fs:(length(x)-1)/Fs;
%decomposition using N equal frames
step=frame_size;
N=ceil(xlen/step);
%computational variables
Cchunks=0;
Lchunks=0;
Csize=0;
PERF0mean=0;
PERFL2mean=0;
n_avg=0;
n_max=0;
n_0=0;
n_vector=[];
for i=1:1:N
    if (i==N)
        frame=x([(step*(i-1)+1):length(x)]);
    else
        frame=x([(step*(i-1)+1):step*i]);
    end
    %wavelet decomposition of the frame
    [C,L] = wavedec(frame,level,wavelet);
    %wavelet compression scheme
    if wavelet_compression=='on'
        [thr,sorh,keepapp] = ddencmp('cmp','wv',frame);
        if heavy_compression == 'on'
            thr=thr*10^6;
        end
    [XC,CXC,LXC,PERF0,PERFL2] = wdencmp('gbl',C,L,wavelet,...
level,thr,sorh,keepapp);
        C=CXC;
        L=LXC;
        PERF0mean=PERF0mean + PERF0;
        PERFL2mean=PERFL2mean+PERFL2;
    end
    %Psychoacoustic model

```

```

if psychoacoustic == 'on '
    P=10.*log10((abs(fft(frame,length(frame))))).^2);
    Ptm=zeros(1,length(P));
    %Inspect spectrum and find tones maskers
for k=1:1:length(P)
    if ((k<=1) | (k>=250))    bool = 0;
    elseif ((P(k)<P(k-1)) | (P(k)<P(k+1))), bool = 0;
    elseif ((k>2) & (k<63)), bool = ((P(k)>(P(k-2)+7)) &...
(P(k)>(P(k+2)+7)));
    elseif ((k>=63) & (k<127)),...
bool = ((P(k)>(P(k-2)+7)) & ...
(P(k)>(P(k+2)+7))& (P(k)>(P(k-3)+7))& (P(k)>(P(k+3)+7)));
    elseif ((k>=127) & (k<=256)), ...
bool = ((P(k)>(P(k-2)+7)) &(P(k)>(P(k+2)+7))...
& (P(k)>(P(k-3)+7)) & (P(k)>(P(k+3)+7)) & ...
(P(k)>(P(k-4)+7)) & (P(k)>(P(k+4)+7)) & ...
(P(k)>(P(k-5)+7)) & (P(k)>(P(k+5)+7)) ...
& (P(k)>(P(k-6)+7)) & (P(k)>(P(k+6)+7)));
    else
        bool = 0;
    end
    if bool==1
        Ptm(k)=10*log10(10.^(0.1.*(P(k-1)))+10.^(0.1.*(P(k))))...
+10.^(0.1.*P(k+1)));
    end
    end
    sum_energy=0;
%sum energy of the tone maskers
    for k=1:1:length(Ptm)
        sum_energy=10.^(0.1.*(Ptm(k)))+sum_energy;
    end
    E=10*log10(sum_energy/(length(Ptm)));
    SNR=max(P)-E;
    n=ceil(SNR/6.02);
%number of bits required for quantization
    if n<=3
%to avoid distortion by error of my psychoacoustic model.
        n=4;
        n_0=n_0+1;
    end
    if n>n_max

```

```

        n_max=n;
    end
    n_avg=n+n_avg;
    n_vector=[n_vector n];
end
%Compander( compressor )
    if compander=='on '
        Mu=255;
        C = compand(C,Mu,max(C) , 'mu/compressor ');
    end
%Quantization
    if quantization=='on '
        if psychoacoustic=='off '
            n=8;
%default number of bits for each frame
%sounds better but uses more bits
        end
        partition = [min(C):((max(C)-min(C))/2^n):max(C)];
        codebook = [min(C):((max(C)-min(C))/2^n):max(C)];
        [index,quant,distor] = quantiz(C,partition,codebook);
        %find and correct offset
        offset=0;
        for j=1:1:N
            if C(j)==0
                offset=-quant(j);
                break;
            end
        end
        quant=quant+offset;
        C=quant;
    end
%Put together all the chunks
    Cchunks=[Cchunks C];
%NOTE: if an error appears in this line just
% modify the transpose of C
    Lchunks=[Lchunks L'];
    Csize=[Csize length(C)];
    Encoder = round((i/N)*100)
%indicator of progress
end
Cchunks=Cchunks(2:length(Cchunks));

```

```

Csize=[Csize(2) Csize(N+1)];
Lsize=length(L);
Lchunks=[Lchunks(2:Lsize+1) Lchunks((N-1)*Lsize+1:...
length(Lchunks))];
PERF0mean=PERF0mean/N %indicator
PERFL2mean=PERFL2mean/N %indicator
n_avg=n_avg/N%indicator
n_max %indicator
end_of_encoder='done'%indicator of progress
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%In this part the signal is stored with the new format
%or transmitted by frames
%This new format uses this parameters:
%header: N, Lsize, Csize.
%body: Lchunks (small),
%Cchunks(smaller signal because now it is
%quantized with less bit and coded)
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
% DECODER %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%reconstruction using N equal frames of length
step (except the last % one)
xdchunks=0;
for i=1:1:N
    if i==N
        Cframe=Cchunks([(Csize(1)*(i-1))+1]:Csize(2)...
+(Csize(1)*(i-1))]);
        %Compander (expander)
        if compander=='on '
            if max(Cframe)==0
                else
            Cframe = compand(Cframe,Mu,max(Cframe),'mu/expander');
            end
        end
        xd = waverec(Cframe,...
Lchunks(Lsize+2:length(Lchunks)),wavelet);
    else
        Cframe=Cchunks([(Csize(1)*(i-1))+1]:Csize(1)*i]);
        %Compander (expander)

```

```

        if compander=='on '
            if max(Cframe)==0
                else
                    Cframe = compand(Cframe,Mu,max(Cframe),'mu/expander');
                end
            end
        end
        xd = waverec(Cframe,Lchunks(1:Lsize),wavelet);
    end
    xdchunks=[xdchunks xd];
    Decoder = round((i/N)*100)
%indicator of progress
end
xdchunks=xdchunks(2:length(xdchunks));
distorsion = sum((xdchunks-x').^2)/length(x)
end_of_decoder='done'
%creating audio files with compressed schemes
wavwrite(xdchunks,Fs,bits,'output1.wav')
%this does not represent the
% real compression achieved. It is only to hear the results
end_of_writing_file='done'
%indicator of progress
% figure(1);
clf;
subplot(2,1,1) plot(t,x);
ylim([-1 1]); title('Original audio signal');
xlabel('Time in [seg]','FontSize',8);
subplot(2,1,2) plot(t,xdchunks,'r');
ylim([-1 1]);
title('Compressed audio signal');
xlabel('Time in [seg]','FontSize',8);

```

Capitolo 6

Conclusioni

L'audio codifica richiede la rimozione due due elementi principali: le irrilevanze percettive e la ridondanza statistica al fine di fornire il livello di compressione richiesta da molte applicazioni a banda limitata. E' stato riscontrato che uno schema di codifica percettiva fornisce il migliore contesto in cui sia l'irrilevanza percettiva sia la ridondanza statistica possono essere rimossi. Un codificatore percettivo è tipicamente costituito da tre fasi principali, vale a dire, un banco di filtri, un modello psicoacustico, e una fase di codifica e quantizzazione.

La fase del banco filtro viene utilizzata per trasformare il segnale di ingresso in un dominio più appropriato all'applicazione di criteri di percezione, e di un dominio che fornisce alcune decorrelazioni del segnale in ingresso.

La fase psicoacustica è utilizzata per calcolare i criteri di percezione, cioè la soglia di mascheramento, utilizzando elementi di psicoacustica e psicofisica. L'effettiva riduzione del bitrate da quantizzare viene effettuata dalle fasi di codifica e quantizzazione, riquantizzando i coefficienti del dominio della trasformata in base ai risultati del mascheramento.

Il calcolo della soglia di mascheramento è di importanza centrale in un codificatore percettivo dato che indica come un ascoltatore percepisce il suono e come il rumore di codifica può essere introdotto e modellato senza introdurre distorsione della percezione. Come risultato, il modello psicoacustico detta direttamente o indirettamente la progettazione della fase banco di filtri e, successivamente, l'azione della fase di codifica e quantizzazione.

I comuni modelli psicoacustici che si trovano nei codificatori oggi sono basati sul

modello di mascheramento. I modelli di mascheramento sono particolarmente apprezzati grazie alle loro implementazioni relativamente semplici e ai risultati ragionevolmente accurati, anche se sono state rilevate alcune carenze.

La trasformata wavelet, o più in generale il banco filtro wavelet (WFB), ha le sue radici in varie branche della matematica, fisica e ingegneria e rappresenta uno strumento interessante e potenzialmente molto utile in molte applicazioni, compreso l'audio codifica percettiva.

Il WFB può essere definito nel dominio continuo e anche nel dominio discreto, ma può essere stabilito un collegamento tra i due. Inoltre, il WFB offre una grande flessibilità nel design attraverso la scelta della struttura ad albero e la base del filtro. La struttura ad albero controlla essenzialmente la risoluzione tempo-frequenza, ad esempio le «piastrelle» nel piano tempo-frequenza, e la base del filtro controlla la localizzazione di ogni banda, cioè la localizzazione delle singole tessere.

Per valutare l'attuale state-of-the-art delle wavelet dell'ambito della compressione audio, è conveniente dividere i segnali tra audio/musica generici e segnali vocali. Cominciamo con audio e musica. Storicamente, negli anni 1970 e 1980, l'idea di sottobanda di codifica audio era molto popolare e ampiamente studiata. Nel corso degli anni questa tecnica è stata affinata e ora si è trasformata in quella che è chiamata *codifica audio percettiva* che è la base per gli standard di codifica audio MPEG. Utilizzando questa idea, la compressione con perdita di dati ha raggiunto qualità trasparente con un tasso di bit basso. Tuttavia, i file compressi con perdita non sono adatti per le applicazioni professionali di ingegneria dell'audio, quali studi in camera acustica, editing audio o la registrazione multitraccia.

Come osservato, il banco filtro con struttura ad albero è la migliore descrizione del modello psicoacustico. Ma questa struttura non è la migliore rappresentazione del modello del canale uditivo, soprattutto perché tale banco filtro dà vita a due scomposizioni e non approssima bene la scala di Bark.

Le bande critiche associate all'udito umano sono più o meno uniformi, ad un primo ordine. Le bande sono più piccole alle basse frequenze e diventano più grandi mano a mano che ci muoviamo in frequenze più alte. Inoltre, le strutture ad albero risultano equivalenti a filtri molto lunghi con eccessivo ritardo. Spesso quando si scrive il codice delle sottobande, i filtri lunghi possono provocare distorsioni che sono udibili a volte.

Un altro tipo di discorso è invece da fare per un segnale, che è più strutturato. Questa

struttura può essere sfruttata attraverso modelli come LPC, vocoder, e HMM. Ancora una volta, la struttura ad albero non aiuta molto nel nuovo metodo.

A causa di ciò, le tecniche di wavelet non sono incluse in nessuno standard corrente di codifica audio. La tecnologia attuale è basata sul MDCT (Modified Discrete Cosine Transform) che meglio descrive il modello psicoacustico e che ha meno problemi di attuazione. Tuttavia esistono alcune applicazioni delle wavelets con l'audio e i segnali vocali. Alcuni autori hanno applicato con successo le wavelet ad esempio per il «watermarking» di segnali audio e vocali.

Bibliografia e Sitografia

Libri e articoli

- «Tecniche di registrazione - Nuova edizione» di Bruce Bartlett, Jenny Bartlett, Apogeo 2001
- «Matlab. Concetti e progetti 2a edizione» di Naldi, Pareschi, Apogeo 2007
- «Audio digitale» di Uncini Aurelio McGraw Hill 2006
- «Low Bit Rate Transparent Audio Compression using Adapted Wavelet» di Deepen Sinha e Ahmed Tewfik su IEEE Trans. ASSP, Vol.41, Numero 12, Dicembre 1993
- «Audio e multimedia» terza edizione di Vincenzo Lombardo e Andrea Valle, Apogeo 2008
- «Informatica e Musica» di Lonello Tarabella, Jackson libri, Milano 1992
- «La mente musicale, Psicologia cognitivista della musica» di John A. Slobosa, Bologna Il Mulino 1988

Siti Web

- <http://www.worldlingo.com/ma/enwiki/it/Wavelet>
- <http://www.audiomaster.it/tutorial/introduzionepsicoacustica.htm>
- <http://www.nemesi.net/audio3.htm>
- <http://www.neuronica.polito.it/pubb/tesi.pdf>

- http://www.ing.unisi.it/biblio/tesi/daprileroberto_0607.pdf
- <http://www.ismprofessional.net/pascucci/documenti/suono-tempo/x357.html>
- http://www.inquinamentoacustico.it/_download/acustica_psicoacustica.pdf
- http://www.dmi.unict.it/~battiato/mm0910/Parte2_Audio%20Digitale.pdf
- <http://www.oocities.com/tuttowindows98/soundc2.htm> <http://www.nemesi.net/audio4.htm>
- http://www.clurican.com/nozioni_elettroacustiche/sensazione_sonora/sensazione_sonora.asp
- http://www.maurograziani.org/text_pages/acoustic/acustica/MG_Acustica05.html
- <http://ispac.ing.uniroma1.it/scarpiniti/files/LabEMM/Less4.pdf>
- <http://www.taolab.it/algorithms/wavelet/wavelet.htm>
- <http://www.ge.infn.it/~pesce/Wavelets%20e%20denoising.pdf>
- <http://www.na.iac.cnr.it/~bdv/wt.pdf>
- <http://www.maurouberti.it/psicoacustica/psicoacustica.html>
- <http://www.nemesi.net/audio3.htm>
- <http://www.audiomaster.it/tutorial/introduzionepsicoacustica.htm>
- <http://www.wikipedia.it>
- <http://www.surveillance-video.com/wavelet-feb-2010.html>
- <http://www.scribd.com/doc/14333043/Matlab-Wavelet-Toolbox-Users-Guide-Version-2>
- <http://www.mathworks.com/>
- <http://www.wavlist.com/>

Ringraziamenti

I miei più sentiti ringraziamenti vanno alla mia relatrice, la professoressa Carla Guerini, con cui è più di un anno che collaboro sia per la realizzazione di questa tesi, sia per lo svolgimento del tirocinio. Sempre paziente e disponibile, è la mamma-professoressa che tutti vorrebbero.

Un grazie speciale va anche a Matias Zanartu che si è reso sempre disponibilissimo nell'inviarmi il materiale da consultare e utilizzare.

Subito dopo vorrei ringraziare i miei genitori che con il loro incrollabile sostegno morale ed economico, mi hanno permesso di raggiungere questo traguardo.

Un grazie infinito va a mia sorella Giulia che con tanta pazienza mi sta vicina notte e giorno, sopportando ogni mio capriccio o desiderio. Allo stesso modo voglio ringraziare il mio gruppo di gallinelle composto da Martina, Virginia, Bruna e Valeria, fedeli compagne di «curtigghio» e divertimento.

Un sentito grazie anche ai miei colleghi della triennale e della specialistica, in particolare a Bruno, Fabrizio, Fabietto, Vincenzino, Donz, Dora Veronica. Con voi ho trascorso (e ancora trascorro) degli anni meravigliosi, siete stati tutti importanti per il mio cammino, chi per un motivo e chi per un altro. Adesso so che qualcuno di voi si lamenterà perchè non faccio un elenco dettagliato di quello che ho trascorso con ognuno di voi, ma per quello vi rimando ai ringraziamenti della tesi della specialistica, tanto siamo ancora tutti in gioco!

Vorrei dire grazie anche a una persona che adesso non vedo più ma che in questi anni è stata, nel bene e nel male forse, la più importante in assoluto, e questa persona è Elio. Grazie perchè con te sono cresciuta e maturata, sei sempre stato capace di aiutarmi anche quando eri straimpegnato con le tue cose, sei stato capace di darmi tante attenzioni e di sostenermi e per questo meriti uno spazio in questi ringraziamenti.

Grazie anche agli amici di sempre, quelle persone che dopo la maturità ho dovuto abbandonare, ma che in realtà sono sempre rimaste con me: Miki, Lambo e Valle. Voi siete il mio trio delle meraviglie.

Ultimi ma non per questo meno importanti, ci sono «gli sciclitani». Tommaso, Peppe, Morana, Enrico e Mauro siete stati i miei compagni di avventura sin dai primi anni qui a Bologna e siete stati il mio punto di riferimento. C'è anche chi tra voi ha dovuto convivere con me e quindi doppio grazie perchè non deve essere stato facile.

Grazie infinite a tutti.