

# **CORREZIONE DEGLI ARTEFATTI NELLA RIPRODUZIONE AUDIO DIGITALE**

**Dott. Ing. Tommaso Giunti**  
Estratto da Tesi di Laurea

Gennaio 2009

---

*Adattato e Ridotto da:*



**UNIVERSITÀ DI PISA**

**FACOLTÀ DI INGEGNERIA**

Corso di Laurea Specialistica in Ingegneria delle Telecomunicazioni  
Tesi di Laurea

**PROGETTO E REALIZZAZIONE DI UN SISTEMA SOFTWARE PER  
LA CORREZIONE DEGLI ARTEFATTI NELLA RIPRODUZIONE AUDIO  
DIGITALE**

Relatori:

Prof. Filippo Giannetti

Prof. Ruggero Reggiannini

Dott. Graziano Bertini

Dott. Massimo Magrini

Candidato:

Tommaso Giunti

**ANNO ACCADEMICO 2007-2008**

---

## Indice generale

Introduzione.....	4
Capitolo 1. Distorsione acustica ambientale.....	5
1.1. Il suono in uno spazio chiuso.....	5
1.2. Teoria modale.....	8
1.3. Acustica geometrica.....	10
Capitolo 2. Soluzioni per un ambiente Hi-Fi.....	15
2.1. Interventi strutturali.....	15
2.2. Correzione acustica digitale.....	20
2.3. Progetto di nuovo sistema di correzione.....	22
Capitolo 3. Analisi della funzione di trasferimento del locale d'ascolto.....	27
3.1. Algoritmo di Goertzel.....	29
3.2. Fast Fourier Transform.....	33
3.3. Constant-Q Transform.....	42
Capitolo 4. Sviluppo Software per Macintosh.....	53
4.1. Sottosistema audio di Mac OS X.....	54
4.2. Strumenti di sviluppo per l'ambiente Mac.....	58
4.3. Descrizione dell'interfaccia utente.....	60
4.4. Una architettura Hi-End.....	64
Conclusioni.....	66
Ringraziamenti.....	67
Bibliografia.....	68

## Introduzione

Il presente lavoro di tesi si inserisce al termine di un percorso formativo che ho compiuto presso il laboratorio Segnali e Immagini dell'istituto ISTI-CNR di Pisa. Il **Dott. Graziano Bertini** e il **Dott. Massimo Magrini** mi hanno proposto di usare le moderne tecniche di elaborazione numerica dei segnali nella ricerca dell'alta fedeltà (Hi-Fi) del suono nelle riproduzioni musicali. In particolare mi è stato proposto di progettare e sviluppare un software di elaborazione numerica dei brani musicali memorizzati e riprodotti dal computer. Questa elaborazione ha una importante finalità: compensare le distorsioni lineari provocate dall'acustica del locale d'ascolto e dalla catena di riproduzione, che compromettono le caratteristiche timbriche del suono. La correzione delle distorsioni lineari del locale d'ascolto avviene pre-elaborando il segnale con una combinazione di filtri parametrici. Per permettere all'utente la regolazione di questi parametri il software (attualmente in testing) integra un analizzatore di spettro che rappresenta la risposta in frequenza dell'ambiente al rumore rosa. Questa tematica è stata proposta da **TangerineTech Engineering**.

In particolare il mio lavoro ha riguardato:

- lo studio dei fenomeni fisici che provocano le distorsioni acustiche negli ambienti chiusi e il progetto di un sistema di correzione
- la sperimentazione di tre tecniche d'analisi nel dominio della frequenza dei segnali in banda audio
- lo sviluppo di un analizzatore di spettro per l'analisi real-time della funzione di trasferimento del locale d'ascolto.

Le elaborazioni sono state implementate in un software per la piattaforma Apple Macintosh.

## **Capitolo 1. Distorsione acustica ambientale**

Sebbene al giorno d'oggi la musica venga fruita sempre di più in mobilità grazie ai dispositivi di riproduzione portatili ed ascoltata attraverso auricolari o cuffie, vi sono molte persone che ascoltano ancora la musica anche in ambiente domestico, dove viene diffusa dagli speaker del proprio impianto di riproduzione. Molto spesso è sempre lo stesso locale dell'abitazione ad essere destinato a “teatro” dell'esperienza musicale.

Un'idea sbagliata che si è andata diffondendo è quella che la qualità dell'ascolto possa aumentare con l'impiego di altri diffusori o che dipenda esclusivamente dal livello tecnico del proprio impianto. Al contrario anche un buon sistema Hi-Fi offre un ascolto non ottimale della musica riprodotta se viene posizionato in un pessimo ambiente.

Infatti la stanza si comporta come uno strumento indesiderato che suona oltre ai musicisti: il locale d'ascolto introduce una propria connotazione timbrica al segnale mentre sarebbe auspicabile un comportamento acustico quasi neutro. Le informazioni spettrali del suono sono già presenti sul supporto stesso su cui la musica è registrata e il locale d'ascolto in cui viene riprodotta non deve aggiungere informazioni che non gli appartengono.

In questo capitolo vengono brevemente descritti quei fenomeni che generano la distorsione acustica ambientale e motivano la necessità di minimizzare l'accoppiamento elettroacustico tra sorgente, ambiente ed ascoltatore.

### ***1.1. Il suono in uno spazio chiuso***

Le proprietà acustiche di un locale domestico sono determinate da molti fattori come i materiali impiegati nella costruzione del pavimento e delle pareti,

## 1.1. Il suono in uno spazio chiuso

---

l'arredamento e la presenza di aperture, e in particolare dalla forma geometrica della stanza e dalle sue misure. Per i fenomeni acustici il parametro da utilizzare nel giudicare grande o piccolo un ambiente è il rapporto tra le dimensioni lineari del locale e la lunghezza d'onda  $\lambda$  :

$$\lambda = c / f \quad (1.1)$$

dove  $c \approx 343 \text{ m/s}$  è la velocità del suono ed  $f$  la frequenza del suono preso in considerazione. Se si considerano le frequenze di interesse per l'orecchio umano, la lunghezza d'onda in aria per la frequenza di 20 Hz è pari a circa 17 m, mentre quella per la frequenza di 20 kHz è di circa 1,7 cm. Pertanto la stessa camera d'abitazione in cui è collocato un impianto Hi-Fi sarà piccola per i suoni emessi dall'altoparlante dei bassi, mentre sarà grande per quelli emessi dall'altoparlante degli acuti. Per un ambiente chiuso è perciò necessario distinguere il problema della descrizione a bassa frequenza da quello ad alta frequenza, ovvero, è possibile definire l'ambiente di dimensioni piccole o grandi secondo la frequenza di interesse. Schroeder ha definito l'omonima frequenza limite  $f_{Sch}$ , diversa per ogni singolo ambiente, che possa distinguere la regione di comportamento a bassa frequenza da quella ad alta frequenza [1]:

$$f_{Sch} = K \cdot \sqrt{\left(\frac{RT_{-60dB}}{V}\right)} \quad (1.2)$$

dove  $K=2000$  è una costante,  $V$  è il volume dell'ambiente e  $RT_{-60dB}$  è il tempo di riverbero: definito come il tempo necessario perché l'intensità sonora decada di 60 decibel rispetto al massimo valore raggiunto prima del transitorio di estinzione, può essere misurato empiricamente o calcolato analiticamente con la (1.5).

La formula (1.2) dimostra che ambienti più grandi e più fonoassorbenti, ossia con un piccolo tempo di riverberazione, presentano questa frequenza a valori più bassi, mentre ambienti piccoli e poco assorbenti mostrano questa frequenza a valori più elevati. Ad esempio una stanza di  $54 \text{ m}^3$  con tempo di riverberazione di un secondo presenta  $f_{Sch} = 272 \text{ Hz}$ , mentre una stanza di  $400 \text{ m}^3$  con tempo di riverberazione di 2 secondi presenta  $f_{Sch} = 141 \text{ Hz}$ .

Pertanto il comportamento del suono può essere studiato distinguendo idealmente tre regioni spettrali: alle frequenze basse, ossia al disotto della frequenza di Schroeder, l'acustica è dominata dalla presenza delle onde stazionarie, per cui il livello della pressione sonora rilevato in regime permanente sinusoidale fluttua ampiamente, spostandosi da punto a punto nell'ambiente, così pure la risposta in frequenza dell'ambiente è fortemente non lineare, inquinando drasticamente l'ascolto. La seconda regione spettrale, al disopra della frequenza di Schroeder, è una banda di transizione che copre circa due ottave fino alla terza regione, a medio-alta frequenza, dove il comportamento acustico dell'ambiente è di tipo statistico: le onde di pressione possono essere approssimate come raggi sonori, simili a quelli della luce secondo la teoria dell'acustica geometrica. Infatti ciò che ascoltiamo in uno spazio chiuso è la complessa combinazione tra il suono diretto che proviene dalla sorgente sonora e il suono indiretto, costituito dalle molteplici riflessioni che ci giungono all'orecchio dalle varie superfici della stanza.

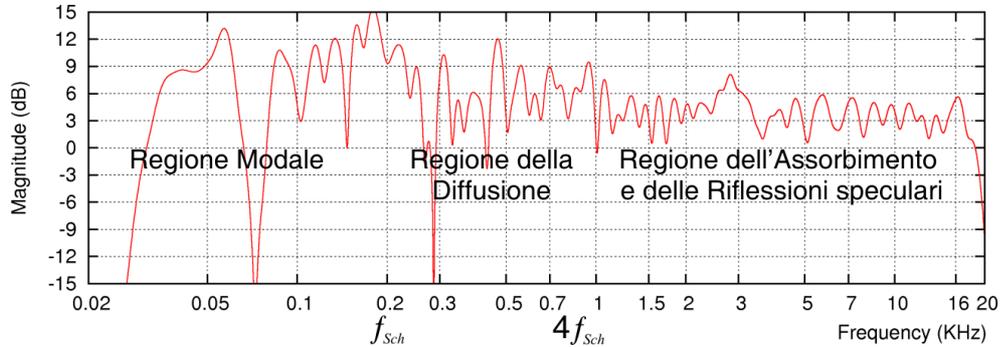


Figura 1.1.: Esempio della risposta in frequenza  $H(f)$  di una stanza.

## 1.2. Teoria modale

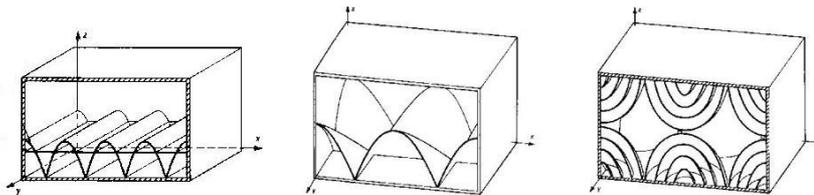
Concentriamoci prima sulla regione delle basse frequenze dove il campo acustico nello spazio è dominato dalle risonanze dell'ambiente stesso: le frequenze di risonanza di una stanza si possono calcolare secondo l'analisi modale. Consideriamo una guida d'onda a sezione rettangolare  $(w \cdot \hat{z} \times h \cdot \hat{y})$  in cui si propaghi il campo sonoro in direzione  $\hat{x}$ , la soluzione generale per la pressione in un punto è:

$$P(x, y, z) = \cos\left(\frac{n\pi z}{w}\right) \cos\left(\frac{m\pi y}{h}\right) e^{\pm j\beta x} \quad \beta = \sqrt{k^2 - \left(\frac{n\pi}{w}\right)^2 - \left(\frac{m\pi}{h}\right)^2} \quad (1.3)$$

dove  $k$  è la costante di propagazione in spazio libero ed  $n$ - $m$  sono numeri interi, zero compreso, che definisco il *modo*: configurazione di campo che si propaga nella guida mantenendosi immutata.

Se chiudiamo la guida d'onda su due pareti in  $x=0$  e  $x=l$  otteniamo una cavità risonante che è un buon modello per una stanza rettangolare. L'onda di pressione acustica viene riflessa dalle due pareti e la distribuzione del campo sonoro è la complessa combinazione lineare di modi di propagazione assiali tangenziali ed

obliqui: i modi assiali considerano una sola dimensione della stanza, quelli tangenziali coinvolgono due superfici e quelli obliqui tutte e tre. La distribuzione della pressione sonora è fortemente simmetrica e presenta zone a maggiore e minore pressione.



*Figura 1.2.: Modi di propagazione del campo sonoro in una stanza rettangolare.*

Quando all'onda di pressione sonora proveniente dalla sorgente si sovrappone un'onda che è stata riflessa dalla parete si origina un'interferenza: se l'interferenza è costruttiva il risultato è l'instaurarsi di onde stazionarie all'interno del locale. Di conseguenza la pressione sonora di un tono a quella frequenza di risonanza verrà rinforzato dalla vibrazione dell'aria contenuta all'interno della stanza; gli altri toni, al di fuori delle frequenze di risonanza, vengono attenuati senza ricevere alcun sostegno energetico da parte del comportamento acustico del locale. Le frequenze di risonanza si calcolano come:

$$f_{res} = \frac{c}{2} \cdot \sqrt{\left(\frac{q}{l}\right)^2 + \left(\frac{n}{w}\right)^2 + \left(\frac{m}{h}\right)^2} \quad (1.4)$$

dove q-n-m sono numeri interi, zero compreso, che definiscono il modo. Ciascun multiplo intero di  $f_{res}$  è anch'esso una frequenza di risonanza. È evidente che la combinazione di questi fenomeni determini la forma della risposta in frequenza  $H(f)$  della stanza: ne comprometta la linearità e ne conferisca una qualità timbrica.

<b>q</b>	<b>m</b>	<b>n</b>	<b>frequenza (Hz)</b>	<b>q</b>	<b>m</b>	<b>n</b>	<b>frequenza (Hz)</b>
0	1	0	34.30	2	0	1	103.06
1	0	0	42.88	2	1	1	108.62
1	1	0	52.91	2	2	0	109.81
0	0	1	57.17	2	2	1	123.80
1	0	1	71.46	3	0	0	128.63
1	1	1	79.26	3	1	0	133.12
2	0	0	85.75	3	0	1	140.76
2	1	0	92.36				

*Tabella 1.1.: Frequenza di risonanza per una stanza ( $w=5, l=4, h=3$  [m])*

### **1.3. Acustica geometrica**

Al di sopra della frequenza di Schroeder le risonanze sono talmente sovrapposte tra loro da non poter più distinguere i singoli contributi di queste al suono complessivo, perciò è necessario sostituire la descrizione del comportamento del suono secondo la teoria modale con un nuovo modello valido anche alle frequenze la cui lunghezza d'onda è trascurabile rispetto alle dimensioni dell'ambiente e degli oggetti presenti in esso. La letteratura suggerisce un approccio di natura geometrica basato sul concetto di raggio sonoro, anziché su quello di onda, che si propaga nello spazio lungo traiettorie rettilinee e che viene riflesso in modo speculare quando incontra un ostacolo.

La riflessione di un raggio sonoro su una superficie rigida e di dimensioni infinite deve soddisfare la condizione al contorno per cui la velocità molecolare della componente normale all'interfaccia sia nulla: pertanto il raggio incidente e quello riflesso si trovano sullo stesso piano e formano angoli uguali con la perpendicolare alla superficie d'incidenza così che la pressione delle due onde

sonore si sommi coerentemente e cancelli la componente normale. È lecito fare riferimento al teorema delle immagini e pensare al raggio sonoro riflesso come se fosse emesso da un diffusore ipotetico dall'altra parte della superficie: è una sorgente dello stesso segnale in posizione speculare rispetto all'altoparlante reale.

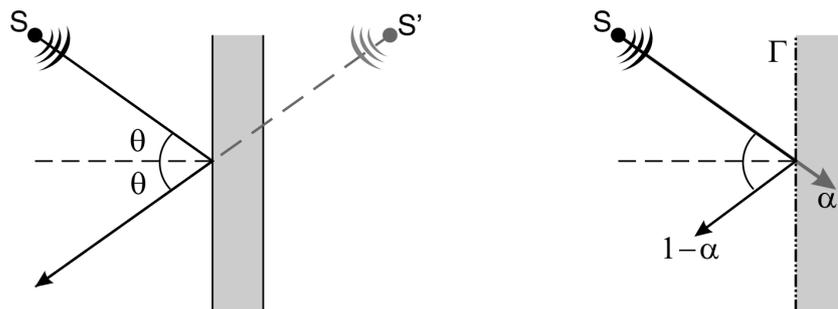


Figura 1.3.: Riflessione di un raggio sonoro su una superficie.

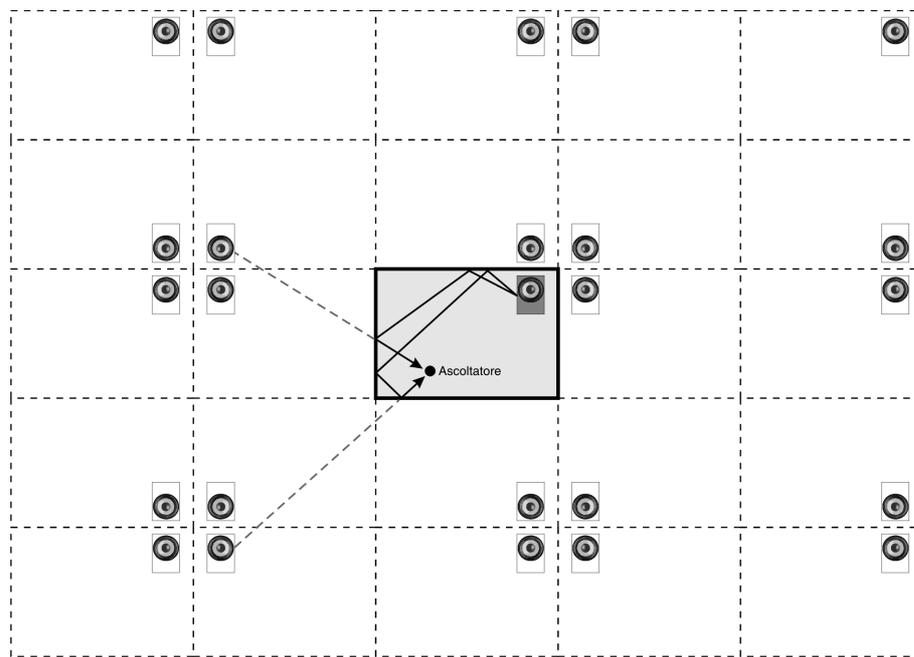
Se l'interfaccia non è perfettamente rigida la componente normale della velocità molecolare può essere non nulla: per rispettare la nuova condizione al contorno il raggio riflesso ha un'ampiezza ridotta dal coefficiente di riflessione

$\Gamma = (1 - \alpha)$  della superficie, il quale vale 1 se una superficie è totalmente riflettente e 0 per una superficie totalmente assorbente. Il coefficiente di assorbimento acustico  $\alpha$  dipende dall'angolo d'incidenza e dalla frequenza del suono, variabili che modificano lo spettro di potenza del suono riflesso e che ne riducono l'intensità. Tuttavia spesso si utilizza un valore relativo a tutte le superfici dell'ambiente: il coefficiente di assorbimento acustico medio  $\bar{\alpha}$  viene calcolato invertendo la relazione che definisce il tempo di riverbero, che viene misurato empiricamente [2]:

$$RT_{-60dB} = \frac{K}{\ln(1 - \bar{\alpha})} \cdot \frac{V}{S \cdot \bar{\alpha}} \quad (1.5)$$

dove  $K=0.161$  è una costante,  $V$  è il volume ed  $S$  è la superficie totale della stanza.

Questi principi dell'acustica geometrica permettono di studiare il comportamento di un raggio sonoro anche all'interno di uno spazio chiuso come ad esempio la stanza di un appartamento: si può considerare ciascuna riflessione come se avvenisse su un'interfaccia di dimensioni infinite a patto che la superficie sia uniforme, ovvero non presenti irregolarità di dimensioni paragonabili a quelle della lunghezza d'onda del suono. Si possono descrivere riflessioni multiple creando una sorgente immagine del second'ordine  $S''$ , simmetrica a quella di prim'ordine  $S'$  che a sua volta è posta dietro la superficie d'incidenza simmetricamente alla sorgente reale  $S$ , e così via fino a che il raggio sonoro non incontra una parete totalmente assorbente o la sua energia si esaurisce.



*Figura 1.4.: Teorema delle immagini: dislocazione bidimensionale dei diffusori fino al secondo ordine.*

La figura 1.4 illustra, in due dimensioni, la dislocazione di alcuni diffusori immagine rispetto ad una stanza rettangolare al centro: evidenzia come un raggio sonoro riflesso coincida con quello diretto diffuso da un ipotetico altoparlante posteriore. Il suono totale percepito nel punto d'ascolto è la somma coerente del contributo diretto e di ciascun diffusore immagine; è lecito formulare una trattazione analitica secondo il modello di canale *multi-path*:

$$r(t) = \sum_{n=1}^{N_p} c_n \cdot s(t - \tau_n) \quad (1.6)$$

dove  $\tau_n$  è il ritardo del cammino n-esimo;  $c_n = 1/\sqrt{d_n} \cdot (\bar{\alpha})^{N_r}$  l'attenuazione proporzionale alla radice della distanza dalla sorgente e al numero  $N_r$  delle riflessioni.  $N_p$  è il numero teoricamente infinito delle sorgenti immagine, che per una stanza rettangolare si estendono in tutte e tre le dimensioni, ma la potenza del suono si riduce a ciascuna riflessione per assorbimento dunque basta computare empiricamente al più 10 riflessioni, che tuttavia corrispondono in tre dimensioni ancora a  $N_p = 1500$  immagini.

Conoscendo tutti i fattori della (1.6) potremmo calcolare analiticamente la risposta in frequenza della stanza mediante trasformata di Fourier:

$$H(f) = \frac{R(f)}{S(f)} = \sum_{n=1}^{N_p} c_n e^{-j2\pi f \tau_n} \quad (1.7)$$

La risposta in ampiezza di un canale *multi-path* è una funzione fluttuante in frequenza che effettivamente assomiglia alla risposta di una stanza come quella illustrata nella figura 1.1.

La descrizione del comportamento acustico dell'ambiente d'ascolto, secondo i criteri dell'acustica geometrica, si complica notevolmente nel caso di stanze la cui geometria non sia rettangolare; a maggior ragione se all'interno della stanza vi

sono comuni oggetti ed arredamento che rendono impraticabile l'uso di questa teoria se non attraverso approssimazioni. Tuttavia questo approccio geometrico aiuta a comprendere la natura della distorsione acustica ed è molto utile nello studio della risposta di un ambiente ai fenomeni transitori, sia che si tratti di suoni impulsivi isolati, sia che si tratti delle fasi di transizione che precedono e seguono lo stato stazionario di un suono.

## **Capitolo 2. Soluzioni per un ambiente Hi-Fi**

In questo capitolo vedremo quali sono gli interventi possibili che mirano ad eliminare le varie sorgenti di distorsione del segnale audio, affinché quest'ultimo arrivi all'orecchio dell'ascoltatore senza aver subito grosse modificazioni rispetto al segnale originale.

Quando si parla di trattamento acustico ci si riferisce a tutta una serie di modifiche dell'ambiente finalizzate alla creazione di condizioni d'ascolto ottimali. Senza scendere nel dettaglio, per cui si rimanda alla letteratura specifica, saranno brevemente descritte le soluzioni strutturali per la correzione acustica di un ambiente domestico: tuttavia emerge che spesso non si prestano ad essere adottate nei casi pratici perché questi interventi si scontrano con limiti dovuti alle dimensioni fissate di uno spazio preesistente e di norma, con lo scopo d'uso cui è destinato, che molto raramente è dedicato esclusivamente alla musica.

Al contrario, le moderne tecnologie di elaborazione digitale del suono hanno assunto fondamentale importanza per la loro capacità di realizzare un sistema correttivo in maniera non invasiva. È all'interno di questo settore che si inserisce la progettazione del nuovo sistema, oggetto di questa tesi, e che verrà descritto nei successivi capitoli.

### ***2.1. Interventi strutturali***

In fase di costruzione dell'abitazione, la scelta di misure adeguate della stanza dedicata all'ascolto della musica può ridurre gli effetti udibili delle risonanze modali alle basse frequenze ed evitare gli effetti indesiderati di colorazione del suono: infatti la distribuzione in frequenza dei modi di un ambiente è dovuta principalmente al rapporto delle dimensioni lineari del locale.

Gli studi [3] eseguiti al riguardo hanno prodotto un elenco di *room ratios* che caratterizzano gli ambienti con un comportamento acustico ottimale: cioè in grado di aumentare la densità modale e diffondere la loro distribuzione nella maniera uniforme.

<b>Dimensione</b>	<b>Design #1</b>	<b>Design #2</b>	<b>Design #3</b>
<b>Larghezza</b>	$1.14 \cdot h$	$1.28 \cdot h$	$1.60 \cdot h$
<b>Lunghezza</b>	$1.39 \cdot h$	$1.54 \cdot h$	$2.33 \cdot h$

*Tabella 2.1.: Tre Optimum-Room-Ratio in funzione del parametro  $h$  che rappresenta l'altezza e spesso è imposto da direttive legislative sull'abitabilità.*

Secondo la teoria modale il peggior comportamento acustico è quello delle stanze a forma di cubo; anche le stanze le cui misure sono un multiplo intero dell'altezza hanno un comportamento molto distorto. Queste considerazioni possono essere utili in fase di progettazione di un locale d'ascolto ancora da costruire mentre nel caso dei locali preesistenti sono un metro di valutazione per determinare la necessità di un eventuale trattamento acustico.

Un primo intervento consiste nel modificare le posizioni reciproche di diffusore-ambiente-ascoltatore per ridurre le distorsioni: la dislocazione degli altoparlanti in un ambiente chiuso influisce sensibilmente sul contenuto spettrale dei suoni che vengono uditi, così come la scelta della posizione d'ascolto ha effetti sulla loro percezione. La teoria delle immagini suggerisce di avvicinare il punto d'ascolto ai diffusori in modo che il suono diretto sia dominante, tuttavia più ci si avvicina alla sorgente e più è probabile percepire la direzione di provenienza del suono perdendo la sensazione di stereofonia. Assumendo un rapporto tra suono diretto e suono riflesso pari ad 8 la tabella 2.2 suggerisce diverse distanze dalla

sorgente in funzione del tempo di riverbero.

$RT_{-60dB}$ [s]	Sorgente - Ascoltatore [m]	Rapporto riflesso/diretto
0.4	$d=3.16$	8
0.5	$d=2.75$	8
0.6	$d=2.30$	8
1	$d=1.83$	8.38
1.2	$d=1.60$	7.9

Tabella 2.2.: Valori d'esempio per una stanza di  $60 \text{ m}^3$  (5x4x3)

I diffusori vanno poi orientati verso il punto d'ascolto per formare un triangolo isoscele il cui vertice cada sulla testa dell'ascoltatore: se il triangolo si chiude davanti l'ascoltatore si ha la sensazione che il “palcoscenico virtuale” si allontani, se si chiude alle spalle dell'ascoltatore si rischia di alterare la percezione del “canale centrale” tra i due diffusori. La teoria modale invece sconsiglia il collocamento dei subwoofer vicino agli angoli della stanza e suggerisce di allontanarli il più possibile dalle pareti per non compromettere l'equilibrio della risposta nella parte bassa dello spettro.

Un fenomeno di distorsione acustica può essere affrontato cercando di preservare l'energia sonora all'interno della stanza oppure disperdendo parte del suono, la scelta dipende dal tempo di riverbero. Per migliorare la resa acustica dell'ambiente è possibile installare materiali e strutture assorbenti con coefficiente  $\alpha$  maggiore di quello delle pareti che permettono di attenuare il suono riflesso rispetto a quello diretto. Questi materiali devono rivestire le superfici in quei punti dove avvengono le prime riflessioni che hanno maggior contributo energetico. A questo scopo si è soliti impiegare tappeti sul pavimento, tendaggi pesanti in

prossimità dei diffusori o pannelli artificiali come quelli illustrati in figura 2.1. Fortunatamente l'assorbimento acustico è bidirezionale ed i pannelli fonoassorbenti sono efficaci anche per isolare l'ambiente da sorgenti di rumore esterne.



*Figura 2.1.: Pannelli fonoassorbenti e Tube-trap.*

Tali accorgimenti sono in grado di controllare le varie distorsioni del suono quasi a tutte le frequenze, tuttavia per creare un assorbimento omogeneo che copra tutte le frequenze possibili, può essere necessario abbondare nei materiali e nelle strutture assorbenti. In un locale di piccole dimensioni ciò si ripercuote in un'acustica non ottimale e troppo smorzata: vanno persi anche gli effetti benefici delle riflessioni, come la spazialità e la sonorità.

Questa limitazione viene corretta con l'uso simultaneo di elementi diffusivi che garantiscono un sostegno energetico da parte delle riflessioni: ciò restituisce al suono complessivo una serie di qualità importanti, come la brillantezza, la definizione, la pienezza e la sensazione di un suono acusticamente vivo. I due tipi di trattamento sono complementari tra loro e, se utilizzati in modo appropriato, possono essere entrambi di aiuto per la costruzione di un locale d'ascolto di qualità.

I diffusori acustici devono offrire un trattamento acustico che assicuri un campo acustico un po' più uniforme e contemporaneamente permetta di scampare al rischio di riflessioni speculari che comportano distorsioni sonore. Questa diffusione del suono può essere ottenuta accentuando l'irregolarità della forma geometrica della stanza che contribuisce a diminuire la simmetria spaziale del campo sonoro. Anche le asperità delle pareti possono diffondere il suono a quelle frequenze la cui lunghezza d'onda è paragonabile alle dimensioni delle sporgenze, mentre le superfici convesse hanno la capacità di disperdere il suono in direzioni diverse da quella dell'onda incidente.

La presenza di oggetti o strutture all'interno dell'ambiente contribuisce alla diffusione del suono a frequenze diverse: un comune diffusore è una semplice libreria riempita alternando volumi di dimensione differenti. Oppure si possono installare diffusori acustici veri e propri come quelli illustrati a titolo d'esempio in figura 2.2: sono dispositivi progettati appositamente per la diffusione del suono e, in quanto tali, hanno caratteristiche migliori rispetto alle superfici non specifiche.



*Figura 2.2.: Elementi diffusori acustici.*

Tuttavia nei casi più comuni è impossibile intervenire sulla struttura muraria e sulle dimensioni dell'ambiente, spesso la dislocazione degli altoparlanti e del punto d'ascolto sono vincolate e a meno che non si voglia tappezzare il soggiorno con pannelli di dubbio gusto estetico, è possibile ricorrere alle nuove tecnologie.

## **2.2. Correzione acustica digitale**

Per descrivere il comportamento del locale d'ascolto, di cui si voglia linearizzare la risposta in frequenza  $H(f)$ , si assuma ancora il modello di canale trasmissivo distorto e tempo invariante. Nei sistemi di telecomunicazioni l'equalizzazione del canale avviene al ricevitore che può stimarne il comportamento grazie ad una fase di training in cui viene trasmesso un segnale noto. Poiché nel contesto della riproduzione musicale è impossibile far affidamento sull'ascoltatore-ricevitore è necessario che sia il sistema di riproduzione-diffusione ad operare l'equalizzazione pre-distorcendo il suono.

L'evolversi delle tecnologie DSP (Digital Signal Processing) hanno permesso lo sviluppo di metodi di controllo attivo applicate al processo di linearizzazione della risposta in ambiente: meglio note come Digital Room Correction (DRC). Questa definizione non identifica un sistema o un algoritmo specifico ma un processo di correzione che coinvolge software e dispositivi diversi.

Il criterio si basa sulla stima della funzione di trasferimento dell'ambiente a partire dalla sua risposta impulsiva  $h(t)$ . È necessaria dunque una fase di acquisizione che consiste nel sollecitare il locale d'ascolto con un impulso sonoro di test emesso dal proprio impianto di riproduzione; è possibile sintetizzare un segnale audio impulsivo di test con specifici software (Cool Edit, Soundforge, etc). La pressione sonora nel punto d'ascolto è la risultante delle distorsioni introdotte dall'ambiente: può essere catturata con un microfono di misura professionale (con risposta lineare in banda audio) che traduce la pressione sonora in un segnale elettrico. Dopo una conversione A/D in formato PCM si può memorizzare in un file.

I motori di calcolo (quali: DRC, Acourate, Room EQ Wizard, etc) ottengono una stima della risposta in frequenza della stanza mediante FFT (Fast Fourier

Transform) di questo segnale numerico. Dopodiché determinano la funzione di trasferimento del filtro inverso e anti-trasformando stimano la sua risposta impulsiva. I valori della sua ampiezza, valutata ad intervallo di campionamento, rappresentano i coefficienti FIR per l'implementazione digitale del filtro inverso di correzione. Questi coefficienti vengono memorizzati in un altro file.

Il segnale musicale in ingresso viene pre-distorto inserendo il filtro FIR nella catena di riproduzione. Ciò è possibile con l'aiuto di un *convolver* (BruteFir, Convolver, etc): un altro plug-in software che esegue la convoluzione tra il segnale in ingresso e il file che contiene i coefficienti.

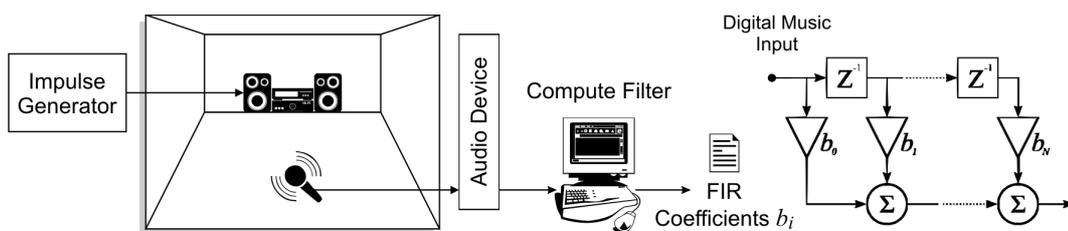


Figura 2.3.: Digital Room Correction. Schema a blocchi.

Un inconveniente di queste tecniche è l'errore intrinseco nella stima di  $H(f)$  che risiede nel segnale di test impiegato. Secondo la teoria, la funzione di trasferimento di un sistema è univocamente determinata dalla risposta all'eccitazione impulsiva  $\delta(t)$ . Tuttavia la delta di Dirac è un concetto ideale e nella fisica del suono è possibile trovar solo delle approssimazioni: un suono, seppur breve, non può avere né durata infinitesima né ampiezza infinita. Nella pratica si ricorre allo sparo di una pistola caricata a salve però il suo diagramma di radiazione è omnidirezionale, diverso da quello degli altoparlanti e dunque non va bene per la misura cui siamo interessati.

Inoltre i fenomeni fisici di risonanza modale si instaurano dopo un transitorio che non riesce ad esaurirsi nel breve tempo di acquisizione legato alla risposta

impulsiva. Nei locali in cui il punto d'ascolto è distante dagli altoparlanti si richiede un segnale impulsivo di potenza maggiore che però rischia di raggiungere la soglia di saturazione dell'amplificatore o di distorsione dei diffusori. Quando si provocano distorsioni non lineari, nascono componenti armoniche che inficiano la stima di  $H(f)$ . Pertanto il filtro FIR va a correggere le distorsioni della funzione di trasferimento dell'ambiente ma anche quelle del sistema di riproduzione che tuttavia non sono presenti a regime d'ascolto.

L'efficacia della correzione ambientale dipende dall'ordine del filtro FIR. Il numero di coefficienti è pari al numero di campioni della risposta impulsiva e una stanza molto riverberante può essere corretta solo aumentando l'ordine del FIR. Il segnale d'uscita dal *convolver* è una replica di quello d'ingresso, ritardata di tanti intervalli di campionamento quant'è l'ordine del filtro. Se questo è trascurabile nella riproduzione musicale, può rappresentare un inconveniente nelle applicazioni di Home Theater, dove il sincronismo tra audio e video è una prerogativa indispensabile.

Affinché possa avvenire la convoluzione, è necessario che la risposta impulsiva del filtro correttivo sia calcolata alla stessa frequenza di campionamento del segnale audio in ingresso. Se il sistema di riproduzione può leggere formati audio con *Sample Rate* diverse è necessario ripetere l'acquisizione della risposta impulsiva per ciascun caso: 44,1 kHz per i CD; 48 kHz per i DAT, Dolby digital o MPEG audio.

### **2.3. Progetto di nuovo sistema di correzione**

Parte dell'attività di progetto svolta presso il laboratorio ISTI-CNR di Pisa è stata promossa dal gruppo TangerineTech Engineering con l'obiettivo di progettare un sistema che superasse gli inconvenienti descritti precedentemente. Il lavoro di

tesi che mi è stato proposto è stato quello di far coincidere in un unico sistema sia la capacità di stimare la  $H(f)$  del locale d'ascolto che di compensarne le distorsioni in maniera efficiente e flessibile.

Il sistema che viene presentato impiega un equalizzatore digitale parametrico per la correzione ambientale e integra un sintetizzatore di rumore bianco o rosa. Questo rumore di test viene prima filtrato dall'equalizzatore e poi inviato al sistema di riproduzione: o in formato digitale, o come segnale analogico.

La risposta del locale d'ascolto alla sollecitazione rumorosa viene acquisita dallo stesso sistema con un microfono di misura che viene connesso al suo ingresso. Lo spettro di questo segnale rappresenta una misura della  $H(f)$  che viene calcolata, aggiornata in tempo reale e illustrata su un'interfaccia grafica. Il processo d'analisi sarà descritto dettagliatamente nel capitolo 3.

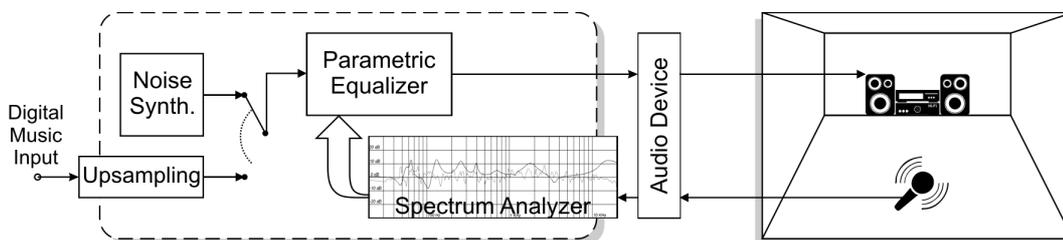


Figura 2.4.: Schema a blocchi del processo di correzione.

All'interno di questa architettura viene dunque a crearsi un anello di retroazione in grado di dare un *feedback* sull'effetto che la modifica di ciascun parametro dei filtri abbia sul suono percepito dal microfono nel punto d'ascolto.

Infatti ciò che il grafico illustra è la risposta in frequenza risultante da:

$$\hat{H}(f) = EQ(f) \cdot H(f) \quad (2.1)$$

dove  $EQ(f)$  è la curva di equalizzazione data dalla combinazione di tutti i

filtri. Questa curva appare illustrata sullo stesso grafico dell'interfaccia utente dove è rappresentata la misura della  $\hat{H}(f)$  in modo da facilitare il settaggio dei parametri dei filtri. Determinare  $EQ(f) \approx 1/H(f)$ , tale che  $\hat{H}(f) \approx K$  con  $K$  una costante indipendente dalla frequenza, si semplifica con un'astrazione grafica dell'intervento: diventa un'operazione intuitiva che può essere svolta da qualunque utente anche senza una preparazione tecnica specifica, eventualmente attraverso una procedura guidata.

Il rumore è un utile segnale di test perché ha un contenuto spettrale che copre l'intera banda audio. Il rumore bianco è un segnale audio ottenuto a partire da un processo aleatorio a media nulla i cui campioni sono variabili aleatorie equidistribuite e indipendenti con varianza  $\sigma^2$ . Sia  $w(t)$  il processo a tempo continuo, la sua funzione di autocorrelazione è:

$$R_{ww}(t_1, t_2) = E[w(t_1)w(t_2)] = \sigma^2 \cdot \delta(t_1 - t_2) \rightarrow S_w(f) = \sigma^2 \quad (2.2)$$

che implica una densità spettrale di potenza uniforme.

Il rumore può essere facilmente sintetizzato *real-time* con un generatore di numeri casuali e rispetto al suono impulsivo usato dai sistemi DRC non ha una connotazione ideale. Il rumore rosa  $p(t)$  invece ha una densità spettrale che diminuisce di 3 dB per ottava, ciò conferisce al rumore la stessa energia per ottava:

$$S_p(f) = \frac{\sigma^2}{f} \quad (2.3)$$

Un altro vantaggio del rumore come segnale di test è che la sua ampiezza si mantiene ad un livello costante per tutta la durata dell'acquisizione senza picchi che rischiano di spingere alla saturazione l'impianto di riproduzione. Inoltre, la

fase di acquisizione non ha più una durata infinitesima e questo garantisce l'esaurirsi dei transistori nei componenti elettronici e l'innescarsi dei fenomeni fisici di risonanza modale.

La correzione effettuata da un equalizzatore parametrico è del tutto paragonabile a quella realizzata con un filtro FIR, eccetto per la fase di calibrazione che richiede l'impegno dell'utente (è stata programmata un'attività di ricerca successiva con lo scopo di progettare un algoritmo che esegua il settaggio automatico dei filtri); tuttavia un equalizzatore digitale parametrico è un sistema scalabile e in linea teorica con un numero infinito di filtri è possibile approssimare qualunque risposta in frequenza. Pertanto l'accuratezza della correzione ambientale migliora al crescere del numero di filtri impiegati che, a differenza di quanto succede per un filtro FIR causale, non implica una latenza tra Input/Output ma un aumento dell'onere computazionale, che per i moderni processori non rappresenta più un vincolo.

Il lavoro si è articolato a partire dal prototipo di un equalizzatore parametrico progettato dal gruppo TangerineTech Engineering all'interno di una partnership che lo lega all'istituto del CNR. Il sistema che viene presentato in questo lavoro di tesi permette all'utente di regolare il guadagno, la frequenza di centro-banda e il fattore di merito  $Q$  di quindici filtri parametrici *peak/notch* per canale. Un filtro *peak/notch* è un filtro elimina-banda se è stato impostato un guadagno minore di uno, quando il guadagno è invece maggiore di uno è un filtro passa-banda alla cui uscita viene nuovamente sommato l'ingresso. L'implementazione consiste nel calcolare i coefficienti dei filtri a partire dai valori di frequenza, guadagno e  $Q$ . Per l'implementazione di questo software si è partiti dal prototipo di un filtro analogico determinando poi l'equivalente a tempo discreto usando la trasformazione bilineare [4].

Una volta settati i parametri dell'equalizzatore, il sintetizzatore di rumore

### 2.3. Progetto di nuovo sistema di correzione

---

viene disattivato e all'ingresso del sistema può essere connessa qualunque sorgente audio digitale. L'algoritmo è flessibile e ricalcola i coefficienti opportuni all'implementazione dei filtri per lavorare con *Sample Rate* diversi. Come sarà descritto nel capitolo 4, a cui si rimanda per le specifiche che riguardano lo sviluppo software, il sistema sovra-campiona il segnale d'ingresso perché la funzione di trasferimento dei filtri numerici si avvicina a quella dell'equivalente analogico all'aumentare della frequenza di campionamento.

Nel capitolo 3 verrà descritto il processo d'analisi della funzione di trasferimento del locale d'ascolto.

## Capitolo 3. Analisi della funzione di trasferimento del locale d'ascolto.

La seconda parte dello studio è stata dedicata all'analisi della risposta in frequenza del locale d'ascolto e alla sua rappresentazione in forma grafica. La prima specifica che ci si è trovati a definire riguarda la risoluzione che doveva avere questa rappresentazione spettrale. Lo spettro del segnale acquisito dal microfono di misura è una funzione su un piano cartesiano in cui sull'asse delle ascisse vi sono le frequenze mentre l'asse delle ordinate rappresenta l'intensità espressa in decibel; si trascuri invece l'informazione sulla fase perché l'orecchio umano difficilmente è in grado di apprezzarne le variazioni. Quando si tratta di segnali in banda audio si è soliti usare l'asse delle frequenze in scala logaritmica.

Il grafico è rappresentato sul monitor di un computer da un'immagine che ha una risoluzione di  $700 \times 200$  pixel ed è aggiornato con un *refresh* di 300 ms. Se si campiona in frequenza la curva dello spettro in scala logaritmica, si trovano i coefficienti discreti di Fourier per quelle frequenze che sono equi-spaziate nel

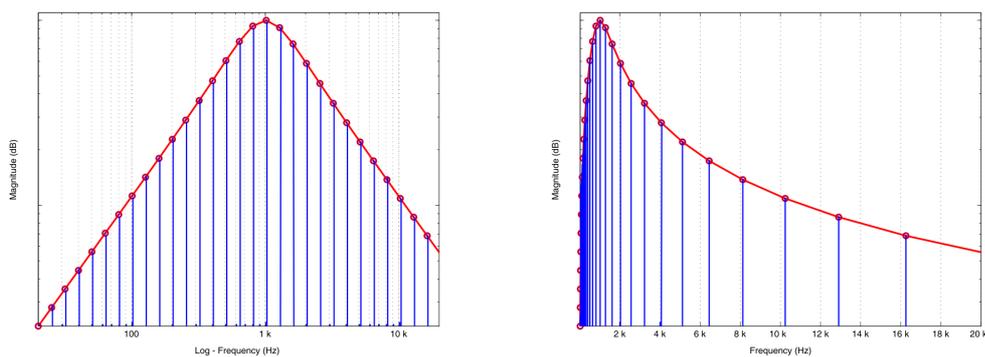


Figura 3.1.: Rappresentazione nel dominio logaritmico (a) e lineare (b) della frequenza.

dominio logaritmico delle frequenze. La curva continua dello spettro si ottiene con una funzione, scritta in linguaggio C, che unisce i punti con una spezzata.

Questa risoluzione presenta una somiglianza con l'apparato uditivo umano e trova riscontro sia in psicoacustica che nella teoria musicale. Dunque per fissare la risoluzione dell'analisi ci siamo ricondotti ai paradigmi usati dagli algoritmi per il riconoscimento delle note musicali: la funzione di trasferimento del locale d'ascolto verrà campionata in corrispondenza delle componenti spettrali che coincidono con i semitoni della scala temperata nella musica occidentale. Infatti l'obiettivo dell'equalizzazione è quello di garantire un comportamento “trasparente” della sala d'ascolto rispetto ad ogni nota musicale che sia stata incisa nel brano che viene riprodotto: così che due note suonate nella stessa maniera non siano mascherate dall'ambiente ma vengano percepite con la medesima intensità.

Il temperamento equabile della scala si fonda sulla suddivisione di ciascuna ottava in dodici intervalli spettrali tali che il rapporto  $x$  tra le due frequenze di ciascun intervallo rimanga costante. Questo rapporto è detto *semitono temperato* e ciascun tono è formato da due semitoni. Considerando due frequenze distanti un'ottava:

$$f_1 = 2 \cdot f_0 = x^{12} \cdot f_0 \quad \rightarrow \quad x = \sqrt[12]{2} = 2^{\frac{1}{12}} \quad (3.1)$$

il numero  $x \approx 1,06$  rappresenta la base per la costruzione della scala secondo una progressione geometrica tale che:

$$f_k = x^k \cdot f_{min} = (2^{1/12})^k \cdot f_{min} \quad k = 0, 1, \dots, k_{max} \quad (3.2)$$

in cui  $f_{min}$  è il limite inferiore di banda dello spettro che siamo interessati ad analizzare, mentre  $f_k$  varia con  $k$  tra  $f_{min}$  e la frequenza di Nyquist. Se l'ottava

invece è divisa in un numero  $n$  di intervalli, solitamente un multiplo di 3, si parla di  $n$ -esimi d'ottava e la ragione della progressione diventa  $(2^{1/n})^k$ . Per poter apprezzare contemporaneamente come ciascuna componente armonica venga attenuata o rinforzata dalle risonanze nel locale d'ascolto, un'analisi accurata è quella a 24-esimi d'ottava perché fornisce due righe spettrali per ciascuna nota musicale (quarti di tono). Sia  $f_{min}=20$  Hz e  $f_{max}\leq f_{SR}/2=20$  kHz in accordo con l'intero *range* dello spettro audio:

$$\begin{aligned} f_k &= (2^{1/24})^k \cdot f_{min} \\ f_{max} &= (2^{1/24})^{k_{max}} \cdot f_{min} \rightarrow k_{max} \approx 240 \end{aligned} \tag{3.3}$$

Lo spettro verrà campionato nel dominio delle frequenze logaritmiche per ottenere  $k_{max} \approx 240$  coefficienti discreti di Fourier.

Tuttavia non tutte le tecniche d'analisi che sono state sperimentate garantiscono una risoluzione spettrale congruente con questa rappresentazione né danno informazioni esattamente sulle componenti spettrali a 24-esimi d'ottava: come verrà descritto in seguito, spesso è necessario interpolare i risultati ottenuti.

### **3.1. Algoritmo di Goertzel**

Un prototipo del sistema di correzione ambientale è stato testato usando come dispositivo d'analisi il software SpectraLAB di Sound Technology con una risoluzione d'ottava così da poter intervenire semplicemente anche con un equalizzatore non parametrico da dieci filtri d'ottava: un plug-in integrato in molti player (WinAmp, iTunes, Windows Media Player).

Come approccio iniziale si è cercato di ricreare questo tipo d'analisi sebbene ritenuta da subito insufficiente: infatti una risoluzione inadeguata sarebbe in contraddizione con lo sviluppo di un preciso equalizzatore parametrico

(sintonizzabile, guadagni e bande controllabili) e non ne sfrutterebbe al meglio le potenzialità.

Un'analisi d'ottava può essere realizzata con un banco di filtri: l'energia del segnale d'uscita da ciascun filtro rappresenta il contributo spettrale di quella sotto-banda e può essere rappresentato con un grafico a barre. Tuttavia si è pensato di calcolare l'ampiezza dello spettro soltanto alle frequenze di centro banda assumendo, nel caso di rumore a banda larga, che l'ampiezza dello spettro si mantenga pressoché costante all'interno di ciascuna sotto-banda.

Allo scopo è efficiente l'algoritmo di Goertzel che permette di calcolare l'ampiezza solo di predeterminate componenti spettrali, a differenza della FFT che considera l'intera banda del segnale [5]. Questo algoritmo è utilizzato per riconoscere i toni prodotti nel sistema DTMF (Dual Tone Multi Frequency).

Sia  $f_{0,i}$  la  $i$ -esima frequenza di interesse, definiamo  $\omega = f_{0,i}/f_c$  la frequenza normalizzata rispetto a quella di campionamento. L'algoritmo di Goertzel consiste nella cascata di un filtro IIR la cui trasformata  $z$  è:

$$\frac{S(z)}{X(z)} = \frac{1}{1 - 2\cos(2\pi\omega)z^{-1} + z^{-2}} = \frac{1}{(1 - e^{+2\pi j\omega} z^{-1})(1 - e^{-2\pi j\omega} z^{-1})} \quad (3.4)$$

dove  $x(n)$  è la sequenza d'ingresso e  $s(n)$  una variabile di stato:

$$s(n) = x(n) + 2\cos(2\pi\omega)s(n-1) - s(n-2) \quad (3.5)$$

seguito da un altro filtro FIR:

$$\frac{Y(z)}{S(z)} = 1 - e^{-2\pi j\omega} z^{-1} \quad (3.6)$$

la cui trasformata  $z$  complessiva è:

$$H_{IIR+FIR}(z) = \frac{S(z) Y(z)}{X(z) S(z)} = \frac{1}{(1 - e^{+2\pi j\omega} z^{-1})} \quad (3.7)$$

che dà una sequenza d'uscita:

$$\begin{aligned} y(n) &= x(n) + e^{+2\pi j\omega} y(n-1) = \sum_{k=-\infty}^n x(k) e^{+2\pi j\omega(n-k)} \\ y(n) &= e^{+2\pi j\omega n} \sum_{k=-\infty}^n x(k) e^{-2\pi j\omega k} = e^{+2\pi j\omega n} \sum_{k=0}^n x(k) e^{-2\pi j\omega k} \end{aligned} \quad (3.8)$$

assumendo  $x(n)=0 \forall n < 0$ . A meno di un fattore esponenziale, la (3.8) rappresenta la DFT di  $x(n)$  valutata proprio per la frequenza  $\omega$  e calcolata su  $N=(n+1)$  campioni:

$$X(\omega) = y(N-1) e^{-2\pi j\omega(N-1)} \rightarrow |X(\omega)|^2 = y(N-1) y'(N-1) \quad (3.9)$$

che, in funzione della variabile di stato  $s$ , vale:

$$|X(\omega)|^2 = s(N-2)^2 + s(N-1)^2 - 2\cos(2\pi\omega) s(N-2) s(N-1) \quad (3.10)$$

ripetendo in parallelo questo algoritmo per ciascuna frequenza centrale delle dieci ottave che coprono la banda audio, otteniamo un grafico a gradini. La figura 3.2 documenta come questo tipo di analisi rilevi lo spettro del rumore bianco e del rumore rosa.

### 3.1. Algoritmo di Goertzel

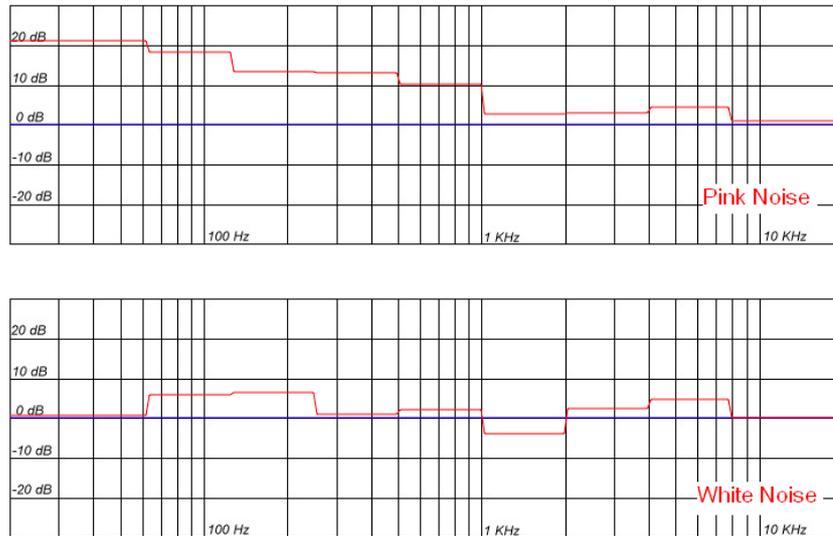


Figura 3.2.: Algoritmo di Goertzel per toni ad intervalli d'ottava.

Appare evidente come sia povera di risoluzione spettrale. Tuttavia si dimostra che l'algoritmo di Goertzel è più efficiente della FFT solo quando il numero  $M$  di toni da analizzare è:

$$M < \log_2 N \quad (3.11)$$

Per l'analisi ad ottave ( $M=10$ ) è sufficiente una risoluzione della DFT pari alla banda della prima ottava (32 Hz) che significa  $N = 44100/32 \approx 1378 \approx 2^{10.4}$ , che verifica la (3.11). Desiderando invece una risoluzione a terzi d'ottava ( $M=30$ ) servirebbe una DFT con  $N \approx 5512 \approx 2^{12.5}$  che non verifica la (3.11): pertanto la FFT diventa computazionalmente preferibile all'algoritmo di Goertzel.

### **3.2. Fast Fourier Transform**

È stato sperimentato l'efficiente algoritmo computazionale di FFT (Fast Fourier Transform) per l'analisi della funzione di trasferimento del locale d'ascolto; non verrà descritto l'algoritmo perché ampiamente documentato in letteratura [6],[7]. Lo studio è stato affrontato prima in ambiente MatLab e successivamente è stata sviluppata la funzione C che implementa l'analizzatore di spettro nel software di correzione ambientale.

Sebbene lo scopo dell'analisi non sia il riconoscimento dei toni puri, è stato sperimentato il comportamento della FFT nei confronti di un segnale sinusoidale con frequenza in banda audio. Infatti si pensi al rumore bianco come ad un accordo ottenuto dalla somma di infiniti toni puri della stessa ampiezza le cui frequenze comprano l'intera banda audio. Pertanto una sola componente armonica di ampiezza diversa dalle altre rappresenta il caso limite nella perturbazione della regolarità dello spettro. La capacità della FFT di risolvere una riga spettrale è la stessa che evidenziare gli eventuali picchi di risonanza, o i *notch*, nella risposta in frequenza della stanza al rumore bianco.

Inoltre un segnale sinusoidale è molto più facile da generare, sia come funzione matematica in MatLab che come segnale audio, rispetto ad un altro che presenti un spettro costante e dei "buchi" strettissimi. Sono stati appositamente creati con il software Soundforge 6.0 oppure sintetizzati real-time da un generatore di toni puri implementato con l'ambiente di programmazione Pure Data.

Il tradizionale approccio matematico all'analisi di Fourier è basato sull'approssimazione di forme d'onda continue. Viceversa le elaborazioni al calcolatore si basano su una forma d'onda campionata e generano uno spettro con solo  $N/2$  armoniche della fondamentale  $\delta f = f_{SR}/N$ : dove  $N$  è il numero di dati

collezionati in un buffer in un periodo di  $T=1/\delta f$  secondi e con  $f_{SR}$  la frequenza di campionamento. L'algoritmo FFT può processare solo buffer la cui lunghezza  $N$  sia una potenza di 2. Pertanto, fissata secondo il teorema di Nyquist una  $f_{SR}$  che scongiuri il fenomeno di *aliasing* (solitamente 44,1 kHz o 48 kHz per segnali audio), la scelta di  $N$  impatta sulla risoluzione in frequenza.

È stata sperimentata una risoluzione che sia pari ad un terzo della prima ottava: se la frequenza di campionamento è di 48 kHz è sufficiente collezionare 8192 campioni per ottenere una risoluzione spettrale di circa  $\delta f=5,85$  Hz. La figura 3.3 illustra lo spettro ottenuto con FFT di una sinusoide di ampiezza unitaria e frequenza 585 Hz:

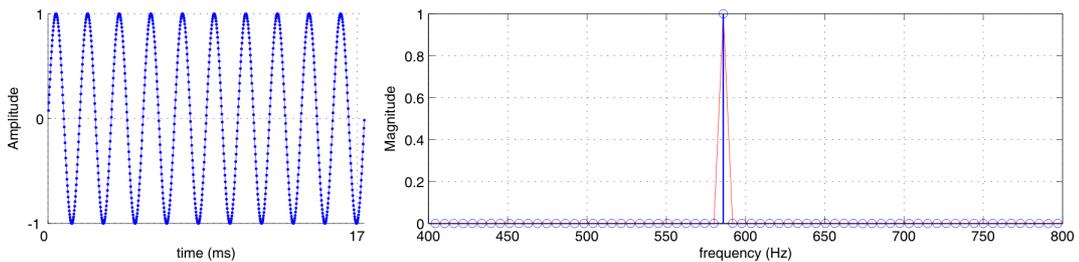


Figura 3.3.: Analisi di una sinusoide a 585 Hz.

In accordo con la teoria dei segnali, lo spettro è composto da una sola riga spettrale nel semi-asse positivo delle frequenze di ampiezza pari a quella del segnale. Si osservi che l'ampiezza è stata corretta perché la FFT genera una serie di Fourier esponenziale anziché trigonometrica, le due coincidono a meno di un fattore 2 di scala. Lo spettro appare chiaro e ideale perché l'intervallo di osservazione  $T$  è un multiplo intero del periodo della sinusoide esaminata: il tono ha una frequenza che è un multiplo intero di  $\delta f$ . Purtroppo questo non avviene per ogni frequenza, la figura 3.4 illustra come viene rilevato lo spettro di una sinusoide con frequenza di 600 Hz:

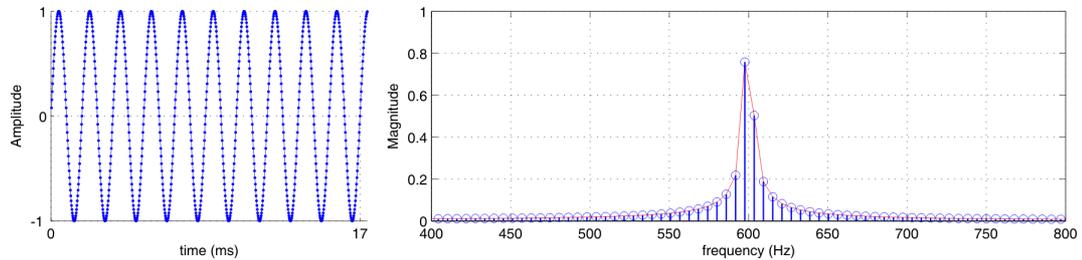


Figura 3.4.: Analisi di una sinusoide a 600 Hz.

Lo spettro è diverso ed è in contraddizione con la teoria: appaiono delle componenti armoniche che in realtà non sono presenti nel segnale ma soprattutto l'ampiezza della riga centrale si è ridotta. Questo effetto è chiamato *Leakage* o è noto come fenomeno di Gibbs [8]. Tuttavia l'energia del segnale è ancora rappresentata ma è stata distribuita sulle componenti armoniche adiacenti.

Il *leakage* è la conseguenza del troncamento del segnale nel dominio del tempo. Memorizzare un buffer di  $N$  campioni equivale a moltiplicare nel dominio del tempo il segnale per una finestra che vale 1.0 nell'intervallo  $[t, t+T]$  e zero al di fuori. L'algoritmo di DFT (Discrete Fourier Transform), di cui la FFT è l'implementazione veloce, elabora la sequenza di campioni nel buffer come se fosse in un loop continuo: simula così una sequenza periodica che può essere espressa come sommatoria di componenti finite dello spettro. Se il primo campione e l'ultimo hanno ampiezza diversa, il segnale periodicizzato presenta una discontinuità a gradino che è la causa nel dominio della frequenza delle componenti spurie che formano il *leakage*.

Per migliorare l'accuratezza dello spettro ci sono due approcci. Potremmo aumentare il numero di campioni del buffer in accordo con le componenti armoniche del segnale in esame, ma nel nostro caso è impossibile perché il rumore bianco contiene tutte le armoniche e dovremmo processare un numero

infinito di campioni dopo un tempo di acquisizione infinito. Un'altra strategia è l'operazione di finestrazione con finestre non rettangolari. Dopo che i campioni sono stati acquisiti in un buffer, comunque di lunghezza finita, vengono pesati con una funzione opportuna; i valori così scalati vengono processati con FFT. Sono state proposte diverse funzioni di finestrazione, verrà fatto riferimento alle finestre di Hamming e Blackman-Harris, definite come:

$$w_{Hamn}(n) = 0.54 - 0.46 \cdot \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.12)$$

$$w_{Bl-Hr}(n) = 0.36 - 0.48 \cos\left(\frac{2\pi n}{N-1}\right) + 0.13 \cos\left(\frac{4\pi n}{N-1}\right) - 0.01 \cos\left(\frac{6\pi n}{N-1}\right)$$

la caratteristica comune è che sono progettate per avere un massimo al centro del buffer e scendere a zero alle due estremità: così che la loro ripetizione non presentino discontinuità nel dominio del tempo. La figura 3.5 confronta lo spettro di un tono a 600 Hz con quello dello stesso segnale pesato dalle finestre:

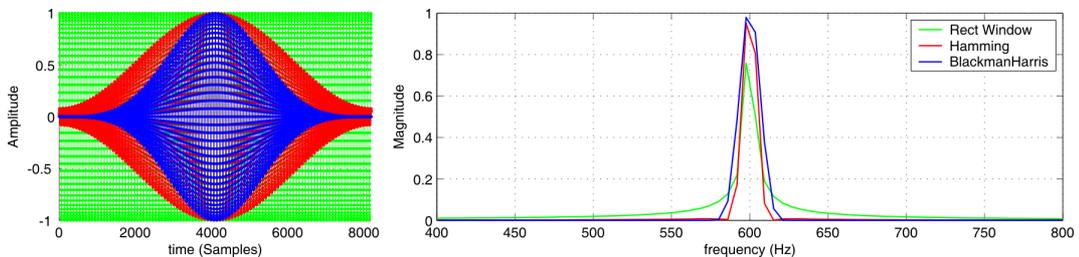


Figura 3.5.: Analisi con finestrazione dei campioni di una sinusoide a 600 Hz.

Si osserva come il fenomeno del *leakage* non sia scomparso ma notevolmente ridotto e l'ampiezza sia tuttavia più accurata.

Alla luce di queste considerazioni fatte con l'ambiente MatLab, si è intrapreso lo sviluppo in linguaggio C di una classe che implementa l'analizzatore di spettro con FFT. Questo metodo è uno strumento dell'applicazione software per la

correzione ambientale che sfrutta la sinergia con l'equalizzatore parametrico.

Come già detto, il processo di equalizzazione parametrico è in grado di elaborare segnali con una frequenza di campionamento  $f_{SR}$  diverse: 44,1 kHz, 48 kHz, 96 kHz e 192 kHz. Tuttavia appare inutile computare con FFT le componenti spettrali al di sopra della banda audio. Pertanto il segnale proveniente dal microfono viene sotto-campionato di un fattore 4 se codificato a 192 kHz, o di un fattore 2 se codificato a 96 kHz. In questo modo l'analisi viene fatta solo su segnali campionati a 48 o 44,1 kHz: il cui spettro è contenuto nella banda audio in accordo al teorema di Nyquist.

Dopo il sotto-campionamento i dati vengono memorizzati in un *buffer* circolare di  $N=8192$  allocazioni in attesa di essere processati ogni 300 ms. Il tempo di *refresh* è stato scelto con un compromesso tra la fluidità dell'aggiornamento del grafico ed un onere computazionale che non gravi sulle prestazioni della CPU.

La FFT è una procedura ormai affermata e ampiamente disponibile sulla rete Internet: ci si è avvalsi di una routine che implementa la stessa funzione di MatLab tradotta in C. La funzione ritorna un vettore di  $N/2=4096$  valori che rappresentano lo spettro in banda audio del segnale d'ingresso campionato nel dominio della frequenza con un passo  $\delta f$ . Tuttavia questa informazione non può essere rappresentata con la risoluzione del monitor di un computer ed è necessaria una decimazione. Come già detto all'inizio del capitolo, si è soliti rappresentare lo spettro di un segnale di natura musicale su un grafico con l'asse delle frequenze in scala logaritmica: si vuole campionare l'ampiezza dello spettro in corrispondenza di quelle frequenze che corrispondono ai quarti di tono della scala temperata. Si è dovuto cercare dunque un criterio che mappasse i risultati della FFT in queste frequenze.

Sia  $x(n)$  la sequenza del segnale d'ingresso e  $X_m$  i coefficienti discreti di Fourier della FFT definiti per le frequenze armoniche  $f_m = m \cdot \delta f$ . Verrà usato d'ora in poi il pedice “ $m$ ” per riferirsi alle frequenze e ai coefficienti calcolati mediante FFT, mentre verrà usato il pedice “ $k$ ” per riferirsi alle frequenze  $f_k$ , definite dalla progressione geometrica (3.3) ed ai loro coefficienti discreti di Fourier. Gli  $X_k$  sono stati calcolati con un'interpolazione lineare:

$$a = \text{floor} \left\{ \frac{f_k}{\delta f} \right\} \quad b = \text{ceil} \left\{ \frac{f_k}{\delta f} \right\} \quad (3.13)$$

$$X_k = \frac{f_k - f_{m=b}}{\delta f} \cdot X_{m=a} + \frac{f_k - f_{m=a}}{\delta f} \cdot X_{m=b}$$

La figura 3.6 illustra lo spettro di un processo di rumore bianco rilevato con quest'operazione:

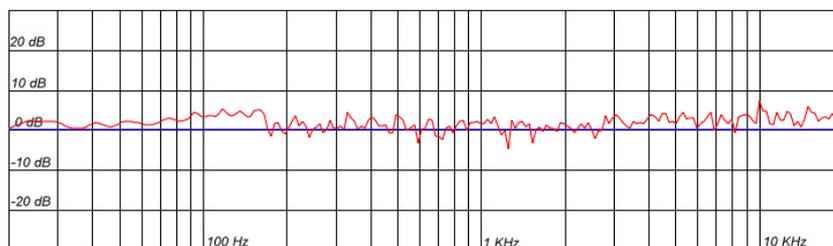


Figura 3.6.: Analisi del rumore bianco.

La spettro di questo segnale a banda larga è quello che teoricamente ci si attende. Invece quando si analizza un segnale composto da cinque componenti armoniche (500 Hz, 1 kHz, 2 kHz, 4 kHz, 8 kHz) tutte della stessa ampiezza si ottiene un risultato non attendibile :

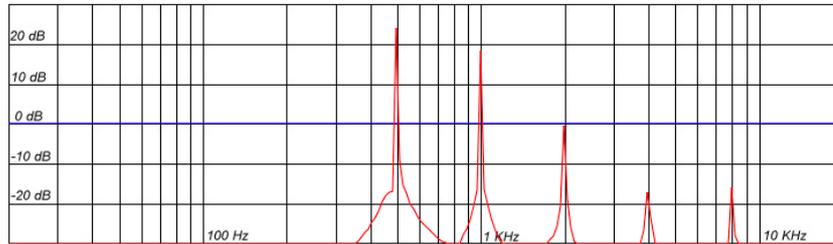


Figura 3.7.: Analisi di un segnale con cinque armoniche.

Lo spettro indica che il segnale è composto da cinque componenti fondamentali ma l'ampiezza relativa appare distorta al crescere della frequenza. : Si trascuri pure l'ampiezza assoluta, tuttavia questo fenomeno renderebbe impossibile equalizzare la funzione di trasferimento del sistema di riproduzione perché, assumendo come riferimento l'ampiezza dello spettro per una certa frequenza, non si riuscirebbe ad apprezzare la distorsione relativa subita dalle altre componenti armoniche.

La causa di questo fenomeno è da cercarsi nella frequenza (3.3) scelte per disegnare il grafico. Si faccia riferimento alla figura 3.1(b) da cui si intuisce la distribuzione nel dominio lineare della frequenza delle componenti armoniche definite da una progressione geometrica. In gamma alta formano un *pattern* molto diradato e se si prova a mappare le righe delle FFT si rischia di “mancare” le componenti dominanti.

*Per fare un'analogia nel dominio del tempo, si campioni un segnale e poi lo si decimi come se fosse stato campionato con un periodo di campionamento che aumenta esponenzialmente con il tempo stesso. È come se si pretendesse comunque di poter ricostruire esattamente il segnale a tempo continuo.*

La distorsione dell'ampiezza in figura 3.7 avviene perché le due righe tra cui si interpola, secondo la (3.13), portano un contributo energetico modesto dovuto al

solo fenomeno di *leakage*. Questo non avviene in gamma bassa dove le  $f_k$  sono più vicine, oppure con segnali a banda larga come il rumore bianco in figura 3.6.

Appare evidente che la FFT, seppur veloce computazionalmente, fornisce componenti spettrali che non mappano efficientemente la frequenze musicali. Questo perché le frequenze  $f_k$  che sono state scelte per la scala temperata occidentale sono spaziate geometricamente mentre le componenti calcolate dalla FFT presentano una risoluzione costante e sono equispaziate. Mappare questi punti da una scala lineare in un dominio logaritmico fornisce poca informazione in bassa frequenza (poche componenti lineari mappate su un gran numero di punti logaritmici) e troppa informazione ad alta frequenza.

Se si vuole preservare tutta l'informazione fornitaci dalla FFT anche dopo l'operazione di decimazione, allora è necessario mappare in una sola componente spettrale  $X_k$  tutti i coefficienti discreti di Fourier  $X_m$  che cadono nell'intervallo:

$$\Delta f_k = \left[ \frac{f_k + f_{k-1}}{2}, \frac{f_k + f_{k+1}}{2} \right] \quad (3.14)$$

Si osservi che questa operazione può essere svolta calcolando l'energia all'uscita di un filtro di banda  $\Delta f$ : un filtro a 24-esimi d'ottava. Un'analisi con i filtri d'ottava, o a terzi d'ottava, sarebbe pur praticabile, ma un banco di oltre 240 filtri digitali a 24-esimi d'ottava, alcuni molto selettivi, richiederebbe IIR di ordine superiore al secondo e quindi un impegno computazionale e progettuale oneroso.

In un primo approccio si è pensato di mappare nel coefficiente discreto  $X_k$  il coefficiente della FFT che assume il valore massimo tra quelli compresi nell'intervallo  $\Delta f_k$ . Questa strategia permette di apprezzare correttamente l'ampiezza dei picchi di risonanza ma ha l'inconveniente di non rilevare i *notch* della funzione di trasferimento; dove si presentassero due picchi di risonanza

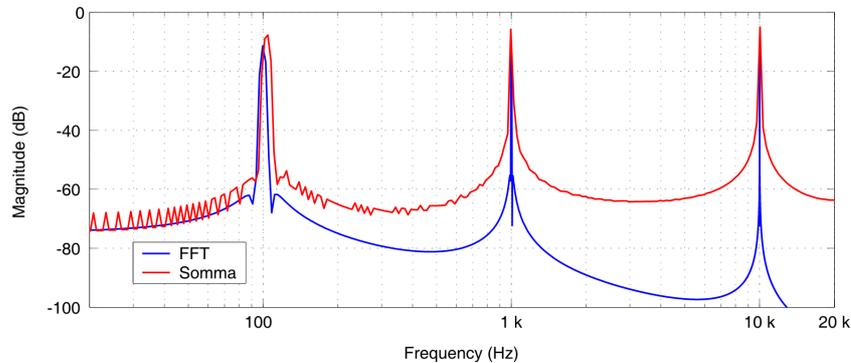


Figura 3.8.: Confronto FFT e metodo della somma.

all'interno dello stesso  $\Delta f_k$ , verrebbe rilevato solo quello di ampiezza maggiore.

In un secondo approccio si è pensato di mappare nel coefficiente discreto  $X_k$  la somma di tutti i coefficienti della FFT compresi nell'intervallo  $\Delta f_k$ . Come mostra la figura 3.8, non è corretto attribuire lo stesso peso a ciascun coefficiente che appartiene all'intervallo  $\Delta f_k$ . Per calcolare  $X_k$  bisognerebbe fare una somma pesata che privilegi i coefficienti  $X_m$  delle armoniche più vicine ad  $f_k$ : questo è ciò che fa la trasformata  $Q$  costante che verrà descritta nel prossimo paragrafo. Per rendere intuitivo il processo di equalizzazione e facilitare il settaggio dei parametri dei filtri, la risposta ideale dovrebbe apparire come piatta. Si osservi che se i punti del grafico rappresentano dei coefficienti discreti di Fourier, si dovrà usare il rumore bianco come segnale audio di misura. Se invece i punti rappresentano quantità integrali della frequenza, come in questo approccio con la sommatoria, si dovrà usare come segnale di misura il rumore rosa. Il rumore rosa, a parità di energia del rumore bianco, ha una densità spettrale di potenza minore in gamma alta: questo è un vantaggio per gli altoparlanti più fragili perché sollecita di meno i tweeter.

### 3.3. Constant-Q Transform

Una soluzione agli inconvenienti trattati nel paragrafo precedente è stata proposta da J.C. Brown nel 1991 ed è nota come Constant- $Q$  transform [9]. Seppure sia stata formalizzata analiticamente, la sua implementazione come codice sorgente C non è di dominio pubblico. Per scrivere la *routine* che implementa l'analizzatore di spettro è stata sperimentata la teoria di J.C. Brown e si sono cercate delle semplificazioni.

Ciascun quarto di tono viene analizzato dalla trasformata discreta di Fourier della sequenza d'ingresso troncata ad una lunghezza diversa:

$$X_k = \frac{1}{N_k} \sum_{n=0}^{N_k-1} w_k(n) \cdot x(n) e^{-j \frac{2\pi Q n}{N_k}} \quad (3.15)$$

dove  $w_k(n)$  è una sequenza di finestra della durata opportuna di  $N_k$  campioni, per esempio la finestra di Hamming definita in (3.12); mentre il fattore di merito  $Q$  è definito come il rapporto tra il valore nominale della frequenza e il passo di campionamento in frequenza:

$$Q = \frac{f}{\delta f} \quad (3.16)$$

Il fattore di merito della progressione delle frequenze  $f_k$  (3.3) assume un valore costante indipendente dalla frequenza:

$$Q = \frac{f_k}{f_{k+1} - f_k} = \frac{1}{2^{1/24} - 1} \approx 34 \quad (3.17)$$

Per semplicità di notazione si definisce la sequenza complessa di finestra:

$$\omega_k(n) = w_k(n) \cdot e^{j \frac{2\pi Q n}{N_k}} \quad (3.18)$$

La (3.15) diventa la sommatoria del prodotto di due segnali nel tempo che per il teorema di Parseval può essere valutata nel dominio della frequenza come:

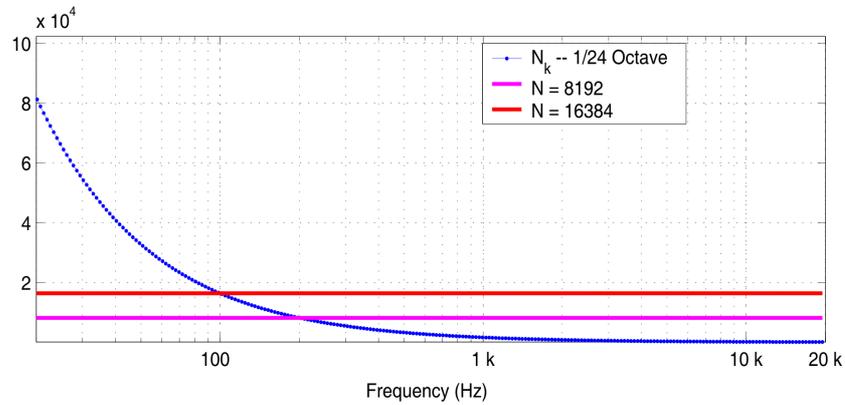
$$X_k = \frac{1}{N_k} \sum_{n=0}^{N_k-1} x(n) \cdot \omega'_k(n) = \frac{1}{N_k} \sum_{m=0}^{N_k-1} X_{m,k} \cdot \Omega'_{m,k} \quad (3.19)$$

dove  $X_{m,k}$  e  $\Omega_{m,k}$  sono i coefficienti calcolati con FFT delle sequenze  $x(n)$  e  $\omega(n)$  per  $n=0,1,\dots,N_k$ . Per analizzare un tono diverso è necessario ricalcolare le FFT su sequenze di un'altra lunghezza:

$$N_k = \frac{f_{SR}}{\delta_k} = \frac{f_{SR}}{f_k} \cdot Q \quad (3.20)$$

Questo metodo è computazionalmente inadeguato ad implementare un analizzatore di spettro che faccia un'analisi *real-time* della funzione di trasferimento del locale d'ascolto, e quindi non permette di percepire immediatamente l'effetto delle modifiche sui parametri dei filtri correttivi. Sarebbe più efficiente calcolare una sola operazione di FFT su un numero di campioni indipendente da  $f_k$ . La stessa J.C. Brown suggerisce [10] che è possibile calcolare l'operazione di FFT rispetto al caso peggiore  $N_{max} = f_{SR} Q / f_{min}$ , ma per un'analisi a 24-esimi d'ottava senza *overlapping*, la FFT verrebbe calcolata ogni 2,13 secondi.

Si è sperimentato il calcolo degli  $X_{m,k}$ ,  $\Omega'_{m,k}$  con un numero  $N$  arbitrario di campioni (comunque una potenza di due).



*Figura 3.9.: Lunghezza delle sequenze nell'analisi a quarti di tono*

La figura 3.9 illustra, al variare della frequenza, le lunghezze  $N_k$  definite dalla (3.20) per  $f_{SR}=44100$ . Nel paragrafo precedente è stata documentata la sperimentazione della FFT di una sequenza di 8192 campioni. Se si fissa nel grafico questo valore, si osserva che le sequenze relative alle  $f_k > 200$  Hz sarebbero più corte, mentre 8192 campioni non sono sufficienti a contenere le sequenze relative alle  $f_k < 200$  Hz. Come mostra la figura 3.10, questo significa che la risoluzione della FFT sarebbe migliore dell'intervallo di quarto di tono sopra i 200 Hz, ma peggiore al di sotto.

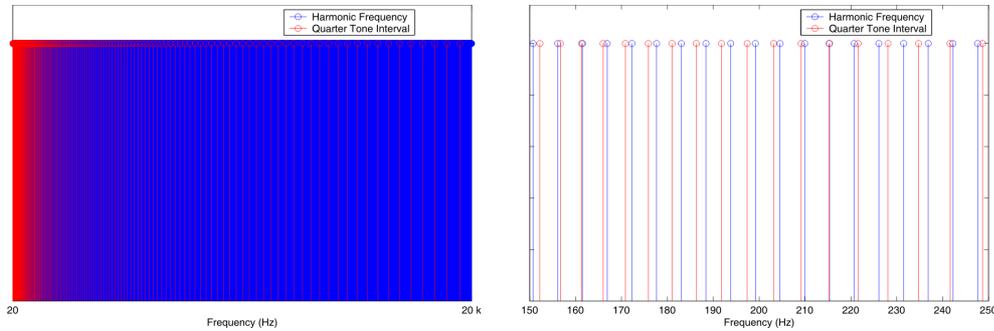


Figura 3.10.:

Tuttavia  $N$  va fissato scegliendo un compromesso tra la risoluzione e la periodicità con cui si intende aggiornare l'analizzatore di spettro. Si osserva che il valore  $N = 16384$  permette una risoluzione spettrale che è maggiore dell'intervallo di quarto di tono già sopra a 100 Hz e un *refresh* del grafico ogni 370 ms. La figura 3.11 mostra i valori di  $N_k$  per la sequenza dei semi-toni ( $Q \approx 17$ ) ed i toni ( $Q \approx 8.5$ ) della scala temperata.

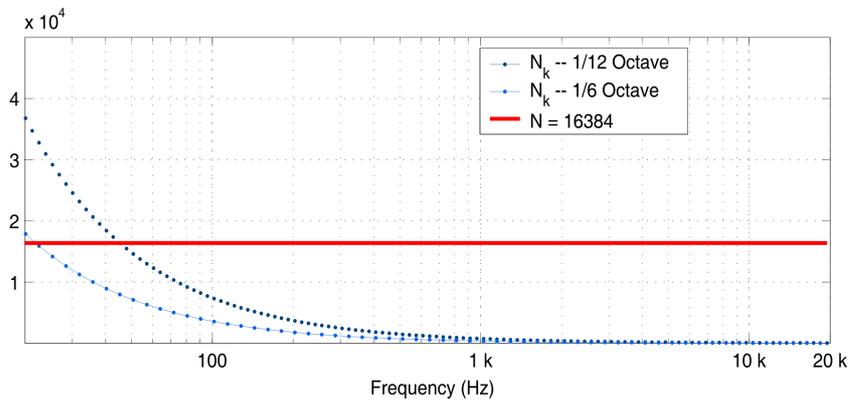


Figura 3.11.: Lunghezza delle sequenze per gli intervalli di tono e semi-tono.

Per sequenze finite di 16384 campioni, si osserva che la trasformata discreta

ha una risoluzione migliore di un quarto di tono a partire dalla terza ottava, migliore di un semi-tono a partire dalla seconda ottava e migliore di un tono su tutto la banda audio.

È stata scelta la lunghezza  $N=16384$ , indipendente da  $f_k$ , che permette di calcolare tutti gli  $X_k$  con una sola operazione di FFT sul segnale d'ingresso:

$$X_m = \sum_{n=0}^{N-1} x(n) e^{-j \frac{2\pi mn}{N}} \quad \Omega_{m,k} = \sum_{n=0}^{N-1} \omega_k(n) e^{-j \frac{2\pi mn}{N}} \quad (3.21)$$

La dipendenza da  $N_k$  è rimasta solo nella lunghezza della sequenza di finestrazione  $\omega_k(n)$  ma non nel calcolo della sua FFT. Tuttavia si osservi che gli  $\Omega_{m,k}$  non dipendono dalla sequenza  $x(n)$  da analizzare e si possono calcolare una sola volta, *off-line*, quando l'applicazione viene inizializzata. La sequenza viene troncata ad  $N$  campioni nell'analisi di quelle frequenze tali che  $N_k > N$ ; la sequenza viene estesa con l'operazione di *zero-padding* per quelle in cui  $N_k < N$ . Per inciso si fa notare che è stato necessario modificare la *routine* di FFT perché l'operazione di *zero-padding* cambia il valore medio del segnale e quindi anche la componente continua del suo spettro. La condizione per cui i coefficienti  $\Omega_{m,k}$  siano reali è che la sequenza nel tempo sia coniugata e simmetrica:  $\omega_k(n) = \omega_k^*(-n)$ . Questo si verifica se la sequenza di Hamming  $w_k(n)$  è centrata rispetto all'intervallo di osservazione e nulla al di fuori di  $[N/2 - N_k/2, N/2 + N_k/2]$ . Alla luce di queste considerazioni la (3.19) può essere riscritta nella forma:

$$X_k = \sum_{m=0}^{N-1} X_m \cdot \Omega_{m,k} \quad (3.22)$$

È stato rimosso anche il termine  $1/N$ : è stata calcolata la sommatoria anziché

la media ma si ottenuto lo stesso risultato perché se si è usato come segnale di test il rumore rosa anziché il rumore bianco, preferito per quanto già citato nel paragrafo precedente. La trasformata  $Q$  costante è stata calcolata dunque come la somma pesata del risultato della FFT: per dimostrare l'effetto dei pesi si esplicitano i coefficienti  $\Omega_{m,k}$  con l'aiuto della (3.18):

$$\begin{aligned}\Omega_{m,k} &= \sum_{n=0}^{N-1} w_k(n) \cdot e^{j\frac{2\pi Qn}{N_k}} e^{-j\frac{2\pi mn}{N}} = \\ &= \sum_{n=0}^{N-1} w_k(n) \cdot e^{-j2\pi(m - \frac{Q \cdot N}{N_k})\frac{n}{N}}\end{aligned}\quad (3.23)$$

e si definisce:

$$\frac{Q \cdot N}{N_k} = \frac{f_k}{f_{SR}} \cdot N = \frac{f_k}{\delta f} \quad \rightarrow \quad \hat{k} = \text{round}\left\{\frac{f_k}{\delta f}\right\}\quad (3.24)$$

Segue che:

$$\Omega_{m,k} = \sum_{n=0}^{N-1} w(n, k) \cdot e^{-j2\pi(m - \hat{k})\frac{n}{N}} = W_{m - \hat{k}}\quad (3.25)$$

dove  $W_m$  sono i coefficienti discreti di Fourier della sequenza di Hamming di durata  $N_k$ , mentre l'effetto dell'esponenziale è quello di sintonizzarne lo spettro attorno alla frequenza  $f_k$ . Si è dimostrato che la trasformata a  $Q$  costante è equivalente ad un banco di filtri con banda passante pari a 24-esimi d'ottava, ma più efficiente ed elegante computazionalmente. Nelle figure 3.12 e 3.13 sono state riportate solo le frequenze definite dalla progressione geometrica di ragione  $2^{1/3}$  per rendere la rappresentazione intellegibile, e le sequenze di  $N=8192$  campioni: si può verificare l'equivalenza con un banco di filtri passa-banda a terzi d'ottava.

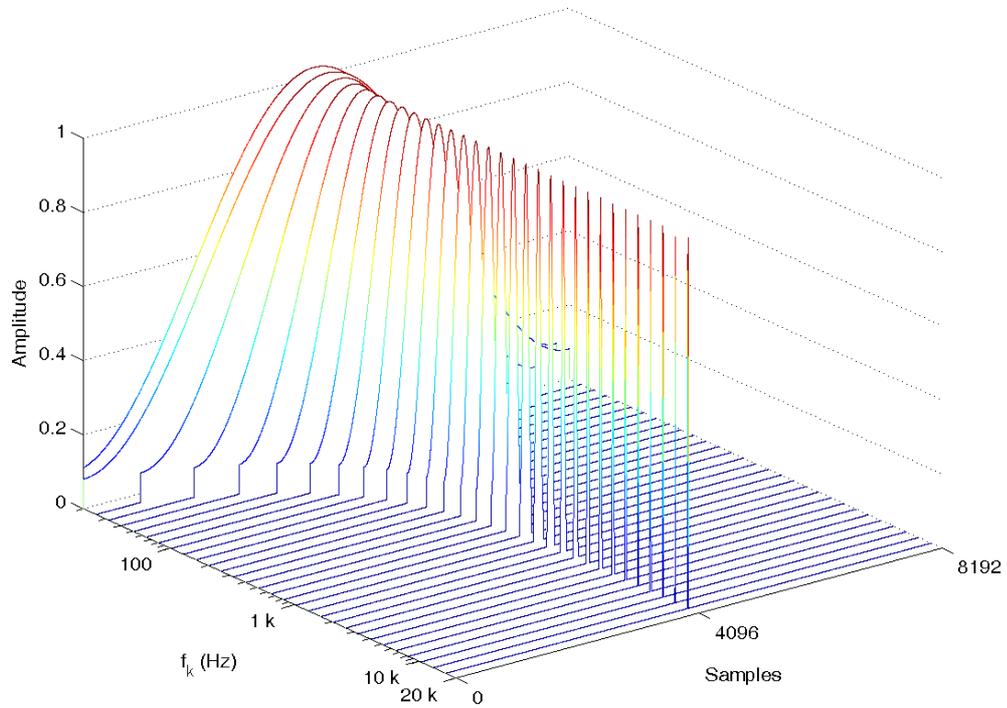


Figura 3.12.: Sequenze di Hamming nel dominio del tempo:  $w_k(n)$ .

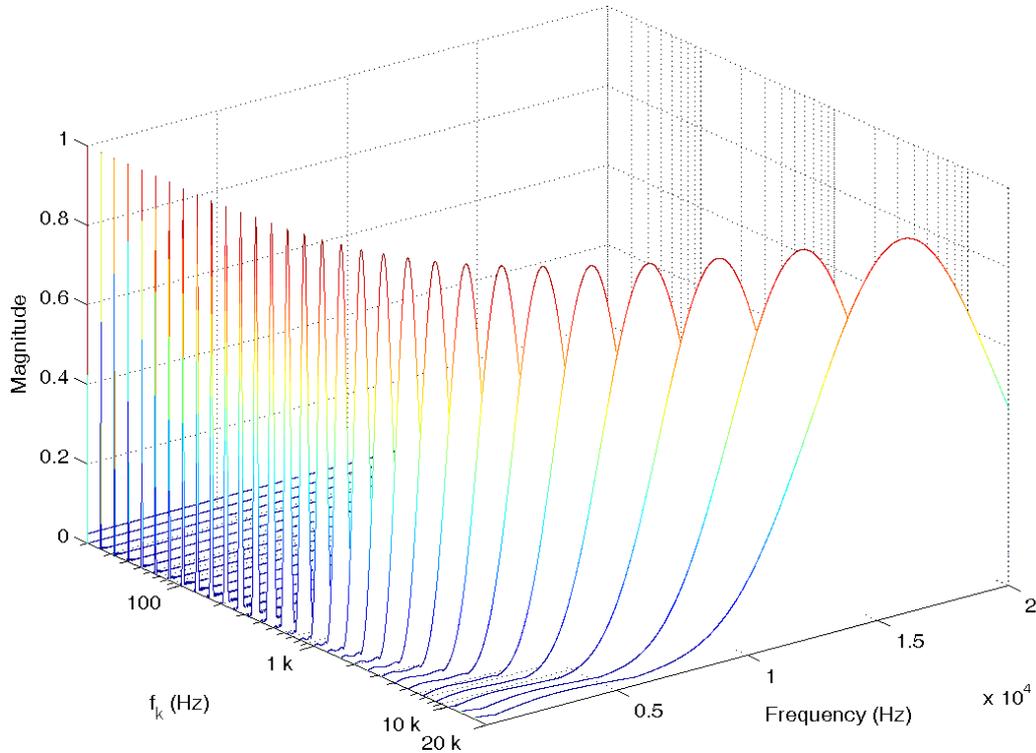


Figura 3.13.: Sequenze di Hamming nel dominio della frequenza:  $W_{m-\hat{k}}$ .

Sostituendo la (3.25) nella (3.22) si è giunti ad una formula che trae vantaggio dall'efficienza computazionale della FFT ma analizza lo spettro per le frequenze definite da una progressione geometrica:

$$X_k = \sum_{m=0}^{N-1} X_m \cdot W_{m-\hat{k}} \quad (3.26)$$

È stato possibile tuttavia ridurre ancora il numero di moltiplicazioni ed addizioni osservando che gli  $W_{m-\hat{k}}$  assumono un'ampiezza trascurabile all'aumentare di  $|(m-\hat{k})|$  e non danno contributo alla sommatoria. Sono stati trascurati i valori inferiori alla soglia 0.054 [11]; questo ha permesso di limitare

### 3.3. Constant-Q Transform

gli estremi della sommatoria per  $|(m - \hat{k})| < C_k$ , dove  $C_k$  rappresenta la metà del numero di operazioni sufficienti al calcolo di  $X_k$ . Come si può osservare dalla figura 3.14, la trasformazione (3.26) aggiunge un onere computazionale ridotto al numero di operazioni della FFT.

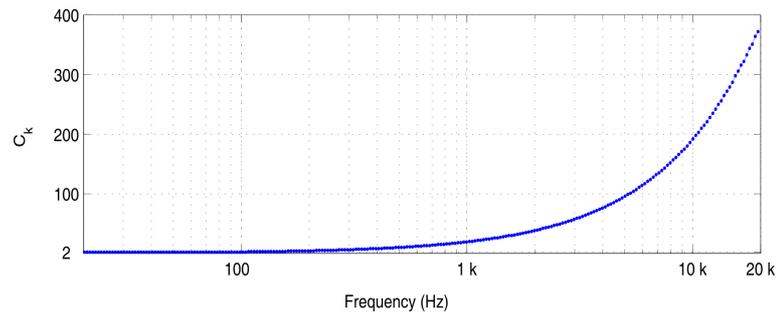


Figura 3.14.: Contributi non nulli nella 3.26.

Per sperimentare l'efficienza della trasformata  $Q$  costante si è posto l'analizzatore di spettro all'uscita dell'equalizzatore il cui ingresso è stato sollecitato da rumore rosa. Agendo sui parametri dei filtri è stato possibile simulare qualunque funzione di trasferimento. Come dimostra la figura 3.15

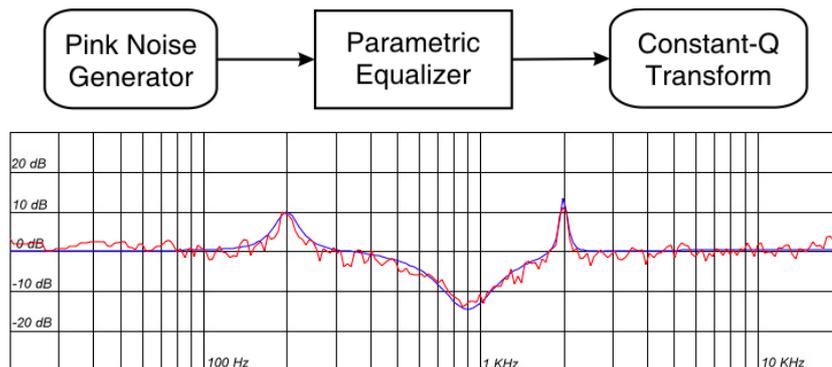


Figura 3.15.: Test d'efficienza della CQT mediante rumore rosa filtrato dall'equalizzatore parametrico.

l'analizzatore di spettro è in grado di misurare qualunque non linearità d'ampiezza nella risposta.

Si osservi che la figura mostra un'istantanea dell'analizzatore di spettro e le irregolarità della risposta sono fluttuazioni, dell'ordine di pochi dB, attorno al valore reale e sono da attribuirsi alla natura aleatoria del processo di rumore: lo *smooth* della risposta si realizza mediando nel tempo i valori che assume ciascun  $X_k$ .

Riguardo alla figura 3.7 si era fatto riferimento ad un segnale composto da cinque componenti armoniche tutte della stessa ampiezza. La figura 3.16 mostra come quello stesso segnale viene rilevato con la trasformata  $Q$  costante: si può osservare che fornisce un risultato più attendibile.

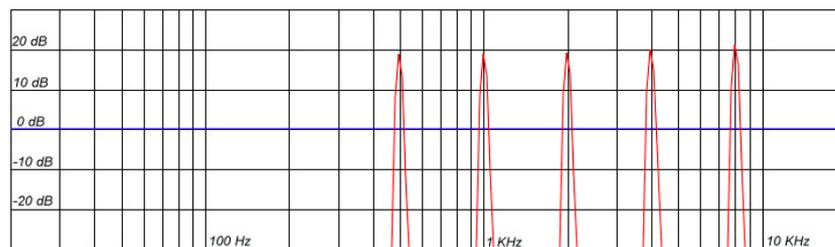


Figura 3.16.: Analisi di un segnale con cinque armoniche.

Terminata la descrizione della trasformata  $Q$  costante, può essere utile riepilogare il procedimento svolto per implementare l'analizzatore di spettro:

- per ciascun quarto di tono  $f_k$  è stata creata una sequenza di Hamming di lunghezza  $N_k$ , simmetrica rispetto ad un *buffer* di  $N$  campioni;
- è stata calcolata la FFT di ciascuna sequenza e lo spettro è stato sintonizzato attorno alla frequenza  $f_k$ ;
- in regime di analisi si calcola la FFT del segnale d'ingresso troncando la sequenza ogni  $N$  campioni;

- si calcolano i coefficienti  $X_k$  con la formula (3.26) trascurando i contributi non significativi al di sotto di una soglia;
- si traccia il grafico dello spettro sull'asse delle frequenze logaritmico interpolando gli  $X_k$ .

Si osservi che la trasformata  $Q$  costante non è invertibile: non esiste un'espressione per l'antitrasformata.

## Capitolo 4. Sviluppo Software per Macintosh

L'attività di progetto, svolta presso l'Istituto ISTI-CNR di Pisa, si è concretizzata nello sviluppo di un DSP software la cui catena di elaborazione audio è rappresentata dal sistema di analisi-correzione ambientale e dall'algoritmo di enhancement della dinamica delle registrazioni, che sono stati descritti nei capitoli che precedono.

La piattaforma hardware su cui è stato implementato è un computer Apple di ultima generazione, sebbene le stesse funzioni di processing possano essere integrate nel firmware di un dispositivo embedded.

Il Computer può essere considerato una sorgente musicale a tutti gli effetti che offre il vantaggio di poter archiviare e gestire intere librerie musicali. Una volta importato il brano, sarà l'hard disk a riprodurlo eliminando gli errori della meccanica di lettura tradizionale. Inoltre può essere connesso ad altri sistema di riproduzione esterni attraverso interfacce sia analogiche che digitali.

Sarebbe riduttivo pensare che la scelta dell'Apple Macintosh sia dettata solo dal contemporaneo contesto commerciale, dove ha saputo ritagliarsi un ruolo da protagonista grazie all'interazione tra Mac OS X, iTunes Music Store e iPod; è la piattaforma di riferimento per il download della musica da Internet. Tuttavia il Macintosh è la scelta ideale per l'integrazione tra musica e sistemi di home entertainment per l'efficienza delle soluzioni hardware di Apple e la robusta affidabilità offerta dal sistema operativo Mac OS X nell'elaborazione dell'audio *real-time*.

A differenza di altre piattaforme, dove al processo dei segnali audio viene data la stessa priorità di qualunque altro dato digitale, il Macintosh garantisce una latenza ridotta grazie a Core Audio: un “motore audio” dedicato, indipendente per

esempio dalla gestione della grafica. Supporta una risoluzione digitale fino a 32 bit / 192 kHz, ben cinque volte superiore alla frequenza di Nyquist per segnali audio. Un altro motivo che ha favorito la scelta del Macintosh è la dotazione standard di interfaccia FireWire che, come sarà descritto in seguito, permette la connessione di schede audio esterne.

In questo capitolo sarà fatta una panoramica su come viene gestito il suono dal sistema operativo Mac OS X e come si inserisce questa nuova applicazione all'interno della sua architettura. Successivamente verranno descritte le librerie e gli strumenti con cui è stato sviluppato il software; presentando poi la sua interfaccia grafica verrà descritto come accedere alle funzionalità descritte nei capitoli precedenti. Infine sarà proposto un setup Hi-end di futura destinazione commerciale.

#### **4.1. Sottosistema audio di Mac OS X**

Core Audio è l'architettura audio di Mac OS X, è stato progettato per gestire l'audio all'interno del sistema operativo. È in grado di riprodurre o registrare, mixare, generare o processare il suono sia di strumenti hardware reali che di strumenti software MIDI.

Core Audio combina un'interfaccia di programmazione in linguaggio C con una profonda integrazione con il sistema operativo. In particolare le caratteristiche utili allo sviluppo di questo lavoro sono state: il supporto plug-in per audio DSP, l'input-output multicanale e scalabile, un approccio standard alla programmazione verso qualunque dispositivo hardware esterno indipendentemente dalla connessione (USB, FireWire, PCI). L'architettura di Core Audio è descritta nella figura 4.1.

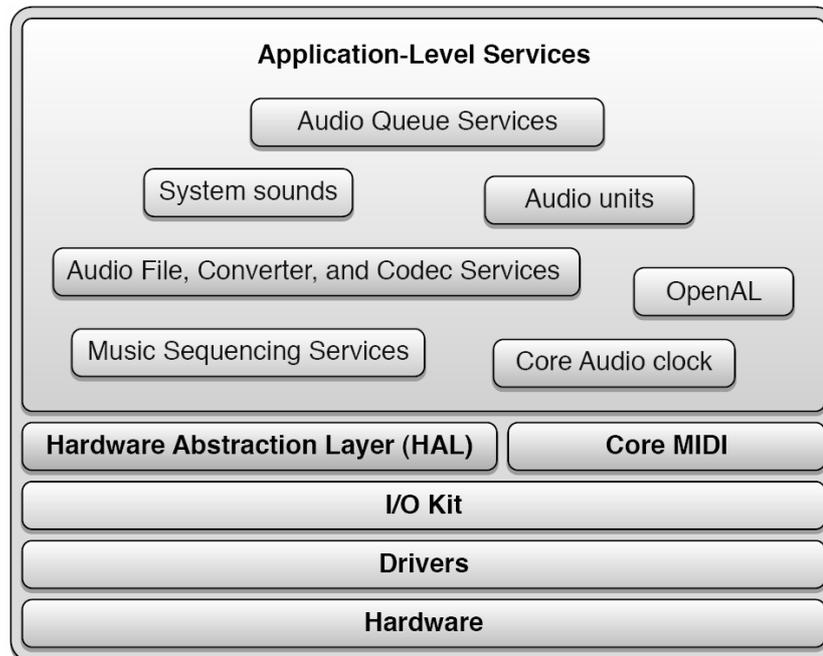


Figura 4.1.: Sottosistema operativo Core Audio di Mac OS X [12].

I segnali audio transitano dai dispositivi hardware attraverso l'HAL (Hardware Abstraction Layer), mentre CoreMIDI gioca lo stesso ruolo per segnali e dispositivi MIDI. L'HAL rappresenta per le applicazioni un'interfaccia con il livello hardware: scambia le informazioni che facilitano la sincronizzazione, garantendo così che il flusso audio non abbia mai latenza. Tuttavia il programmatore non interagisce direttamente con l'HAL bensì con un'altra applicazione AUHAL (Audio Unit HAL) che si occupa di fare il routing dell'audio verso l'HAL. Una catena di processing termina tipicamente con la chiamata dell'AUHAL che è la sola Audio Unit in grado di avviare o interrompere autonomamente il flusso dei dati audio.

Una Audio Unit è un plug-in che conferisce ad un'altra applicazione nuove

funzionalità o caratteristiche: può essere sia un'unità d'uscita che una sorgente audio digitale, come un sintonizzatore o un dispositivo virtuale; oppure entrambi, come un'unità DSP che trasforma uno stream audio. Una Audio Unit è un codice eseguibile che in Mac OS X non può girare autonomamente ma solo all'interno di un'applicazione *host*: le librerie di Core Audio sono il supporto di questa interazione.

Il software che viene presentato poteva essere compilato come plug-in di un'altra applicazione *host* (es. iTunes); invece è un'applicazione autonoma con la sua interfaccia, i suoi menu, etc. Tuttavia, nella catena di riproduzione di Core Audio, può essere pensata come un'applicazione in grado di procurare le capacità di elaborazione di un plug-in per all'applicazione sorgente.

Il sottosistema operativo audio gestisce i dati audio nella modulazione PCM a 32 bit (virgola mobile), che è il formato audio digitale non compresso; ciononostante Core Audio separa totalmente la riproduzione del suono dalla sua gestione consentendo al programmatore di gestire il flussi audio con la frequenza di campionamento e la risoluzione che ritiene opportuna, sarà compito del sottosistema audio adattare il flusso alle reali capacità dell'hardware. In fase di compilazione è sufficiente richiamare la funzione di libreria di Core Audio che implementa il seguente schema:

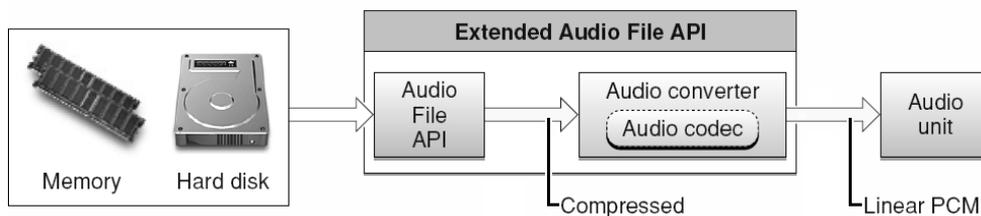


Figura 4.2.: Conversione del formato audio in Core Audio.

dove `AudioConverterAPI` permette di cambiare la risoluzione di

quantizzazione e la frequenza di campionamento attraverso avanzati algoritmi di interpolazione. Core Audio integra anche i codecs per la decodifica dei file compressi nei più comuni formati musicali: Mp3, AIFF, Apple Lossless, etc.

Lo schema a blocchi in figura 4.3 descrive nel dettaglio il percorso del segnale all'interno del computer; si farà riferimento al software con il nome commerciale di AD Charon DSP:

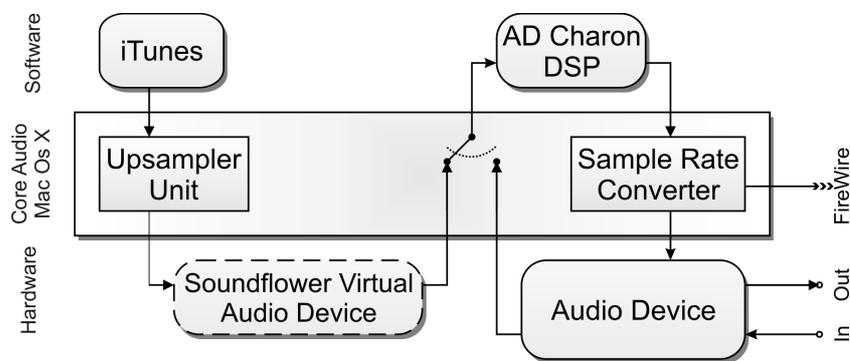


Figura 4.3.: Architettura del sistema d'elaborazione.

I file della libreria multimediale, o il CD precedentemente riversato nell'hard disk, vengono letti, ed eventualmente decodificati, da un software player come iTunes, integrato nell'ambiente Mac OS X. Il segnale viene sovra-campionato da Core Audio e indirizzato, tramite le preferenze di sistema, verso un dispositivo di input/output virtuale emulato da Soundflower.

Soundflower è un'applicazione freeware che permette il *routing* dell'audio sui sistemi Macintosh. Il programma infatti è un *wrapper*: un software in grado di simulare un dispositivo audio virtuale. È stato progettato per fornire ad altre applicazioni la possibilità di inviare e ricevere audio da e per altri software. Lavorando con bassissima latenza e con limitato uso del processore permette a ciascuna applicazione client di utilizzare la sua normale dimensione di buffer.

In Mac OS X Soundflower si presenta come uno o due dispositivi hardware Core Audio. Per la nostra elaborazione è sufficiente utilizzare il dispositivo a due canali anche se sarebbe disponibile un dispositivo a 16 canali per operazioni di *routing* più complesse.

L'applicazione AD Charon DSP legge dunque lo *stream* audio digitale di Soundflower su cui opera l'enhancement della dinamica e l'equalizzazione per correggere le distorsioni acustiche del locale d'ascolto. Durante la fase di acquisizione della risposta in frequenza dell'ambiente, l'ingresso va commutato sulla scheda audio integrata a cui viene connesso il microfono di misura.

Core Audio si occupa del sotto-campionamento e del condizionamento del segnale alle specifiche del dispositivo d'uscita: il segnale può essere inviato in formato numerico ad una scheda audio esterna mediante connessione FireWire, oppure alla scheda audio integrata del computer da cui uscirà in formato analogico per raggiungere l'amplificatore e gli altoparlanti.

#### **4.2. Strumenti di sviluppo per l'ambiente Mac.**

Xcode è l'ambiente di sviluppo che fa parte della suite di Mac OS X di Apple: permette di scrivere, fare debug e creare file binari universali sia per sistemi Macintosh con CPU Intel che PowerPC, indipendentemente dalla piattaforma host. Xcode integra Interface Builder: un *tool* WYSIWYG per la realizzazione di interfacce grafiche (GUI: Graphic User Interface), questi template sono memorizzati come file .nib a cui un'applicazione può accedere per creare finestre, menu, ed altri elementi. Xcode contiene una serie di librerie di codice, utilizzabili con linguaggi di programmazione diversi, che permettono allo sviluppatore di accedere alle funzioni del sistema operativo senza per questo doverne scrivere il codice. Queste librerie sono organizzate in *framework* quali Cocoa e Carbon, più

altri di terze parti.

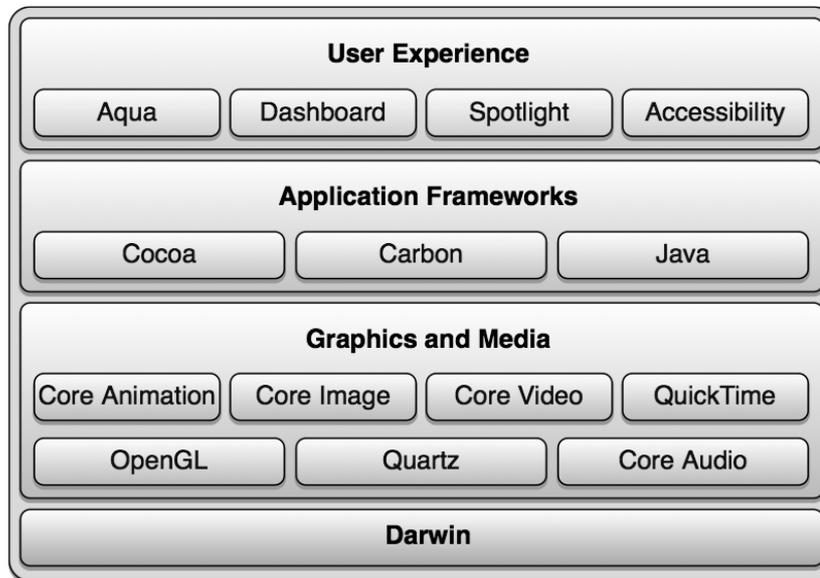


Figura 4.4.: Architettura gerarchica di Mac OS X [13].

Cocoa è il nome delle API (Application Programming Interface) che danno accesso allo sviluppo di applicazioni native per Macintosh, garantendo tutte le potenzialità del sistema operativo. Lo sviluppo che segue il *framework* di Cocoa utilizza il linguaggio di programmazione con cui è stato scritto il sistema operativo stesso: Objective-C. Objective-C è un'estensione del modello semantico del C per supportare la programmazione ad oggetti dinamici. Tuttavia il software presentato in questa tesi è stato scritto avvalendosi delle funzioni di libreria di Carbon.

Carbon è un'interfaccia procedurale verso il linguaggio C per Mac OS X, pensata per offrire agli sviluppatori un supporto nella transizione da Mac OS 9 [14]. Infatti si tratta di un *framework* utile a garantire la compatibilità con precedenti software, sia per sviluppare nuove applicazioni. Carbon permette agli

---

## 4.2. Strumenti di sviluppo per l'ambiente Mac.

sviluppatori C/C++ anche di fare il porting di applicazioni scritte per piattaforme diverse ed utilizzare le caratteristiche di Mac OS X senza per questo dovere riscrivere il codice sorgente da zero. Le librerie facilitano l'implementazione delle funzioni basilari svolte dall'applicazione: gestione file, interfaccia grafica, event handle, etc.

Il software AD Charon DSP fa appoggio anche sulle funzioni di un'altra libreria, Port Audio, che permettono di aprire stream audio sia di input che di output. Port Audio è una libreria multi-piattaforma sviluppata per Microsoft Windows, Apple Mac OS e il sistema operativo open source Linux. Questo permette allo stesso programma di elaborazione, scritto in linguaggio C, di girare su diversi sistemi operativi semplicemente ricompilando il codice sorgente per l'ambiente opportuno.

Rispetto alle interfacce di input/output native, Port Audio permette di richiamare funzioni più semplici senza tuttavia compromettere la latenza o sacrificarne le prestazioni. Consente di accedere anche alle funzioni d'utilità del sistema con un'interfaccia procedurale standard identica per ogni sistema operativo. Tra queste, sono state utilizzate le funzioni che richiamano la lista dei dispositivi audio installati e permettono di gestirli attraverso l'applicazione, specificando quale sia il buffer d'ingresso e quale quello d'uscita.

### **4.3. Descrizione dell'interfaccia utente**

L'interfaccia utente è molto semplice ed intuitiva: la schermata iniziale mostra il controllo del volume d'uscita espresso in decibel ed i VU-meter dei due canali.

La sezione “*Settings Reports*” permette di scegliere i dispositivi di ingresso-uscita e le curve di correzione ambientale tra i preset delle impostazioni avanzate. È possibile scegliere anche la frequenza di sovra-campionamento ed abilitare un

filtro passa-basso supersonico che reietta eventuali componenti armoniche al di sopra della banda audio.

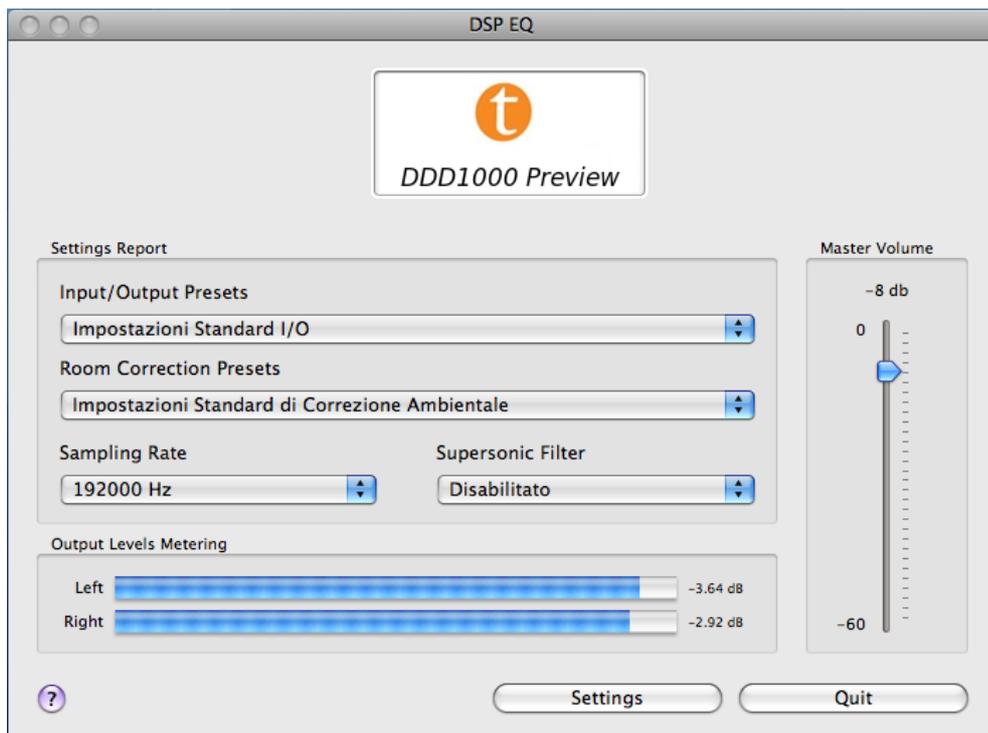


Figura 4.5.: Finestra principale di AD Charon DSP (versione preview).

Una volta selezionati questi preset, l'utente non dovrà più interagire in nessun modo con il programma e potrà godere al meglio del proprio impianto di riproduzione nella sua massima espressione. Infatti la registrazione dei preset è un'operazione avanzata da svolgere solo in fase di installazione, o comunque saltuariamente, alla quale si accede attraverso il pulsante “Settings”.

La sezione “Input / Output” permette di selezionare i dispositivi audio installati e configurati dal sistema operativo e di regolarne i rispettivi guadagni di ingresso ed uscita. In fase di acquisizione il dispositivo d'ingresso sarà quello a cui

### 4.3. Descrizione dell'interfaccia utente

è collegato il microfono di misura, mentre in regime di riproduzione musicale bisognerà selezionare il driver virtuale di Soundflower. Ogni possibile configurazione può essere salvata in un file di preset con le icone in basso, secondo le metafore dettate dalla Apple Human Interface Guidelines, e richiamata direttamente dalla schermata principale.

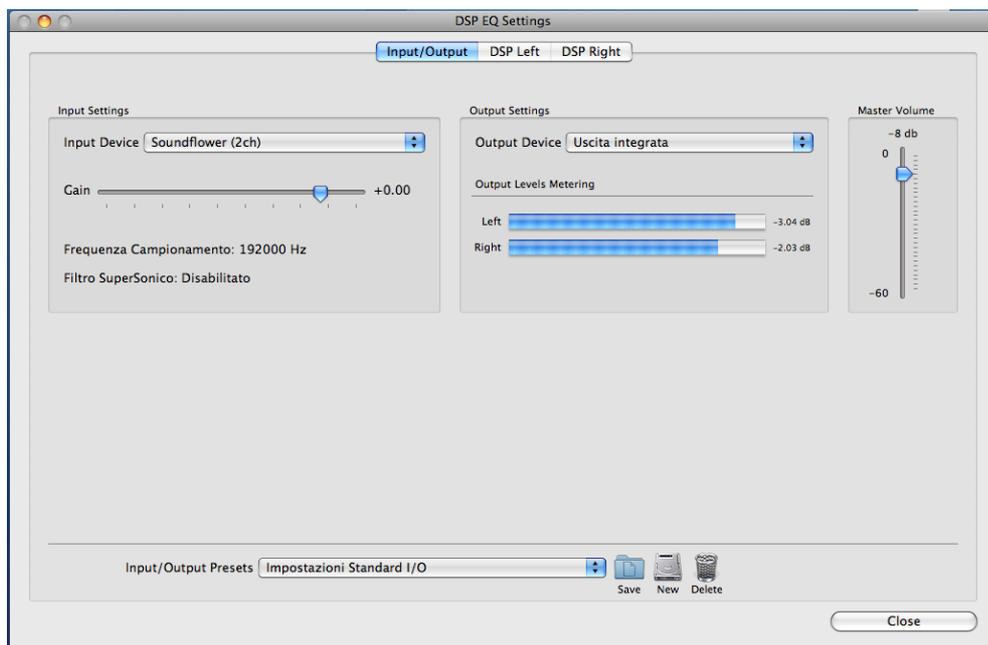


Figura 4.6.: Impostazioni avanzate: dispositivi di Input / Output.

Le sezioni “*DSP Left / Right*” danno accesso ai controlli dell'equalizzatore parametrico e all'analisi della risposta ambientale. Le due sezioni, identiche come interfaccia, agiscono su canali diversi; tuttavia selezionando la check box “*Link Channel*” è possibile intervenire contemporaneamente sulla curva di equalizzazione destra e sinistra. Al contrario per correggere individualmente il contributo di un solo diffusore, è sufficiente selezionare “*Mute*” e il rispettivo canale sarà silenziato.

### 4.3. Descrizione dell'interfaccia utente

Il grafico si riferisce sull'asse  $x$  alle frequenze della banda audio in scala logaritmica e la curva di colore rosso rappresenta la risposta in frequenza della sala d'ascolto. La procedura d'analisi, descritta nel capitolo 3, viene avviata premendo il tasto "Analyze". La riga di colore blu rappresenta la curva di equalizzazione per il canale in questione.

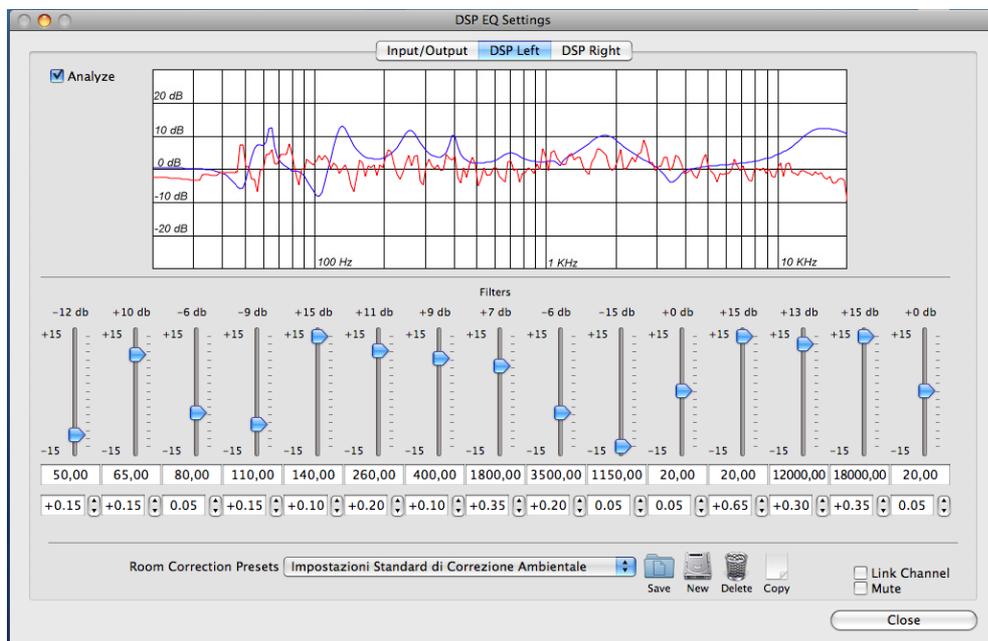


Figura 4.7.: Equalizzatore parametrico e analisi della risposta ambientale.

Ciascuno dei trenta filtri correttori parametrici può essere sintonizzato specificando, nella prima casella di testo, la frequenza di centro banda espressa in Hertz. Il cursore controlla il guadagno a centro banda nel range  $\pm 15$  dB. Il numero  $N$ , da inserire nella seconda casella di testo, non è il fattore di merito  $Q$  del filtro ma rappresenta la larghezza di banda normalizzata rispetto alla banda di un filtro d'ottava con la stessa frequenza centrale. Il fattore di merito si calcola di conseguenza come:

$$Q = \frac{\sqrt{2^N}}{2^N - 1} \approx \frac{1}{2} \cdot \frac{2^N + 1}{2^N - 1} \quad (4.1)$$

dove si è approssimato la radice quadrata con lo sviluppo in serie di Taylor arrestato al primo ordine.

Agendo sui pulsanti in basso si apre una finestra di dialogo in cui è possibile, anche per questa configurazione, definire dei preset. I parametri dei filtri vengono salvati in un file e la curva di equalizzazione di ciascun ambiente può essere richiamata comodamente dalla schermata principale.

#### **4.4. Una architettura Hi-End**

Una parte dell'attività di progetto svolta è stata proposta da TangerineTech Engineering con l'obiettivo di sviluppare un ambiente di elaborazione che potesse fungere da software di supporto all'impiego di una scheda audio esterna di prossima commercializzazione. Infatti nei sistemi audio digitali, siano essi computer o sistemi embedded, la qualità sonora è determinata dalla risoluzione della conversione del segnale da numerico ad analogico: è evidente che all'aumentare della risoluzione aumenta la precisione e quindi la fedeltà del segnale. La correzione delle distorsioni armoniche ed il ripristino della dinamica operate da AD Charon DSP si apprezzano meglio se il software può fare affidamento su una scheda Hi-End performante; la sinergia è biunivoca perché le prestazioni della scheda sminuiscono quando il segnale sorgente è distorto.

Al Macintosh viene dunque lasciato il compito di elaborazione prettamente digitale, nonché di archiviazione, gestione e decodifica dei file musicali; mentre l'onere di una precisa conversione viene demandata alla scheda esterna. La connessione tra i due dispositivi avviene secondo lo standard IEEE 1394 FireWire.



*Figura 4.8.: Configurazione con scheda audio esterna.*

Questa interfaccia di Apple permette la comunicazione su bus seriale a 400 Mbps nella configurazione FireWire 400 e fino a 800 Mbps nella configurazione FireWire 800, garantendo una connessione veloce ed affidabile.

## Conclusioni

L'affermarsi del computer come strumento multimediale ha facilitato la fruibilità della musica e ha permesso l'archiviazione, la trasmissione in rete e la gestione di intere discografie.

A livello professionale, il mondo degli esperti ha sviluppato complessi algoritmi di elaborazione del segnale audio implementati su sistemi HD/SW usando per lo più i processori per DSP.

Viste le attuali prestazioni dei processori per PC è possibile far eseguire le complesse procedure direttamente da un normale computer che diventa così, se opportunamente programmato, anche una potente piattaforma d'elaborazione numerica dei segnali.

Il software sviluppato e descritto nella tesi si inserisce in questo contesto e costituisce una soluzione efficace per linearizzare la risposta dei diffusori e degli ambienti di ascolto. Esso permette di aumentare la profondità e l'immagine sonora e garantisce un'esperienza d'ascolto coinvolgente, una fedeltà del suono che si avvicina all'evento *live*.

Come metodologia, tutti gli algoritmi sono stati prima sperimentati in Matlab e successivamente sviluppati come routine in linguaggio di programmazione C/C++. Il software finale è stato progettato per MAC OS X di Apple, permettendo di acquisire competenze anche nella programmazione con Xcode e Interface Builder.

Dal punto di vista formativo l'esperienza svolta ha dato un esito positivo perché le nozioni teoriche fornite dal Corso di Laurea in Ingegneria delle Telecomunicazioni sono state efficacemente applicate nell'elaborazione audio.

## Ringraziamenti

Desidero porgere i miei più sinceri ringraziamenti, con profonda stima e riconoscenza, a Graziano Bertini e a Massimo Magrini dell'Istituto ISTI-CNR per avermi dato la possibilità di svolgere questa attività di tesi, nonché per la pazienza e l'aiuto dimostrato nelle difficoltà. A loro sarò sempre grato per avermi introdotto all'elaborazione del suono e per aver contribuito alla mia crescita professionale.

Vorrei ringraziare Simone Bianchi della TangerineTech Engineering per avermi proposto di sviluppare la sezione d'analisi acustica ambientale che nella versione finale del software va ad affiancarsi all'equalizzatore parametrico di cui è il progettista.

Desidero infine ringraziare Leonello Tarabella che mi ha concesso l'utilizzo di alcune *routine* sviluppate nel ComputerART Lab. e per l'interesse dimostrato durante lo sviluppo del progetto.

L'ultimo pensiero corre alla mia famiglia che sempre mi ha sostenuto negli studi e a Laura: vera ragione di tutto il mio impegno.

## Bibliografia

- [1] Schroeder M.R., Diffuse Sound Reflection by Maximum Length Sequences, The Journal of the Acoustical Society of America, 1975
- [2] Sabine W. C., Collected Papers on Acoustics , Harvard University Press, 1922
- [3] Ballou, Glen, Ed., Handbook for Sound Engineers, Howard Sams, 1991
- [4] Bristow-Johnson R., The Equivalence of Various Methods of Computing Biquad Coefficients for Audio Parametric Equalizers, AES 97th Convention, Preprint, 1994
- [5] Goertzel G., An Algorithm for the Evaluation of Finite Trigonometric Series , American Math. - 65, 1958
- [6] Luise M, Vitetta G.M, Teoria dei Segnali, McGraw-Hill, 1999
- [7] Oppenheim A.V., Schafer R. W. , Digital Signal processing, Prentice-Hall, 1975
- [8] Gibbs J.W., Fourier Series, Nature - 58, 1898
- [9] Brown J.C., Calculation of a constant Q spectral transform, The Journal of the Acoustical Society of America - 89, 1991
- [10] Brown J.C., Puckette M.S. , An efficient algorithm for the calculation of a constant Q transform, The Journal of the Acoustical Society of America - 92, 1992
- [11] Blankertz B., The constant Q Transform, <http://wwwmath.uni-muenster.de/logik/Personen/blankertz/>,
- [12] Apple Inc., Core Audio Overview, <http://developer.apple.com/documentation/>, 2007
- [13] Apple Inc., Mac OS X Technology Overview,

<http://developer.apple.com/documentation/>, 2007

[14] Apple Inc., Carbon Overview, <http://developer.apple.com/documentation/>, 2005