

UNIVERSITA' DEGLI STUDI DI PARMA

**DOTTORATO DI RICERCA IN INGEGNERIA
INDUSTRIALE**

Ciclo XXII°



**Tecniche di analisi ed equalizzazione acustica
in campo automotive**

Coordinatore:
Chiar.mo Prof. Ing. Marco Spiga

Tutor:
Chiar.mo Prof. Ing. Angelo Farina

Dottorando: Ing. Marco Binelli

Alla mia famiglia

A me stesso

A chi mi ha dato fiducia

*Desidero ringraziare le persone
dell'Università degli Studi di Parma
e della ditta ASK Industries S.P.A.
che mi hanno aiutato durante
questi tre anni.*

*Ringrazio in particolare
Il prof. Angelo Farina e
Francesco Violi
per i preziosi insegnamenti,
il supporto e la collaborazione.*

*Invio un sentito
ringraziamento a Maurizio Arletti,
Claudio Becchi, Marco Romagnoli,
Monica Tesauri e Marco Vizzaccaro
per l'aiuto durante le
attività di laboratorio.*

Indice

Introduzione	6
1 Elaborazione digitale del suono	7
1.1 Elaborazione numerica del segnale	7
1.2 L'equalizzazione	16
1.3 Metodo di inversione di Kirkeby	20
1.4 Wavelet transform	22
2 Levigatura e decimazione spettrale	28
2.1 Lo smoothing	28
2.2 Test soggettivi	31
2.3 Phase unwrapping	38
3 Psicoacustica	41
3.1 Stimolo, sensazione, statistica dei dati	41
3.2 L'area di udibilità umana	43
3.3 Sistema di acquisizione delle informazioni	46
3.4 Il mascheramento acustico	50
3.5 Mascheramento di toni puri con rumore	51
3.6 Mascheramento dovuto a rumore a banda stretta	53
3.7 Mascheramento di toni puro da parte di toni puri	54
3.8 Mascheramento dovuto a una serie di toni	56
3.9 Effetti temporali	57
3.10 Mascheramento simultaneo	58
3.11 Post mascheramento	59
3.12 Bande critiche	61
3.13 Come determinare le bande critiche	62
3.14 La scala di Bark	64

3.15 Conclusioni	65
4 Analisi AQT	66
4.1 Il segnale AQT	66
4.2 La risposta del sistema allo stimolo	74
4.3 Metodo di misura della risposta all'impulso	75
4.4 Analisi della risposta	81
4.5 Articolazione	82
4.6 Andamento armonico	85
5 Sviluppi nell'analisi AQT	89
5.1 Burst con inviluppo a dissolvenza	89
5.2 Burst ad inviluppo esponenziale	93
5.3 Confronto tra inviluppo tradizionale, a dissolvenza ed esponenziale	95
5.4 Studi sulla fase	97
5.5 Il software di calcolo	97
6 Conclusioni	103
Bibliografia	104

Introduzione

Negli ultimi anni l'autovettura è diventata per molti un mezzo indispensabile nell'adempimento del proprio lavoro ed il tempo medio trascorso al suo interno è aumentato. Avere nel proprio autoveicolo una piacevole acustica ed un fedele impianto audio diventa quindi un aspetto sempre più importante del comfort complessivo richiesto dall'utente alle case automobilistiche. Questo giustifica la particolare attenzione che è stata riversata negli ultimi anni dalle grandi case costruttrici alla qualità degli impianti audio installati sulle proprie autovetture già a partire da modelli di fascia medio-bassa. Tecnologie di recente sviluppo hanno dato la possibilità ai produttori di installare nelle vetture alcune categorie di impianti detti di "*infotainment*". Tale neologismo anglosassone indica l'unione delle tecnologie di *information* (ad esempio GPS, GSM, WI-FI e la rete globale Internet) con quelle di *entertainment* (video LCD, impianti audio multicanale, interfacce per PDA o riproduttori musicali portatili). Riguardo a certe nicchie di mercato, come ad esempio le vetture destinate al pubblico giovanile, la disponibilità di tali tecnologie diventa importante ed in certi casi decisiva nella scelta dell'autovettura stessa. Questo fatto porta le case automobilistiche ad una concorrenza verso tale settore e spiega a sua volta spiega perché la ricerca è orientata allo studio dei sistemi audio tecnologicamente avanzati.

Già da alcuni anni sulle vetture di fascia medio alta del mercato europeo ed americano sono presenti sistemi di riproduzione multicanale dotati di elaboratori digitali del suono. Il tema dell'attuale ricerca è quello di sfruttare al meglio le potenzialità dei sistemi digitali indagando da una parte sulla psicoacustica per migliorare la qualità sonora percepita dall'utente finale e dall'altra quello di automatizzare il processo di messa a punto dell'impianto. La ricerca svolta è orientata alla coniugazione di entrambi gli aspetti.

CAPITOLO 1

Elaborazione digitale del suono

In questo capitolo sono introdotti i principali concetti, relativi allo studio dei sistemi di elaborazione digitale dei segnali e al dimensionamento dei filtri numerici. Vengono inoltre indagate le principali tecniche utilizzate per ricavare i filtri inversi necessari all'implementazione del sistema di equalizzazione.

1.1 Elaborazione numerica del segnale

Campionamento

Il concetto di segnale numerico nasce dalla possibilità di riprodurre completamente un segnale analogico attraverso un numero sufficiente di campioni. Se si indica con $^a x(t)$ un segnale a banda limitata B e con $^s x(n)$ i suoi campioni negli istanti nT si può scrivere [1]

$$^a x(nT) \stackrel{\Delta}{=} \sum_{n=-\infty}^{+\infty} ^s x(n) \cdot T \delta(t - nT) \xrightarrow{f} \frac{1}{T} \sum_{n=-\infty}^{+\infty} ^a X \left(f - \frac{n}{T} \right) \quad (1.1)$$

dove con \xrightarrow{f} si è indicata l'operazione di trasformazione di Fourier e con $^a X(f)$ la trasformata di $^a x(t)$. Filtrando passa basso il segnale $^a x(nT)$ così definito si ottiene $^a x(t)$ se e solo se $F_s \geq \frac{1}{T} > 2B$ dove $2B$ prende il nome di frequenza di Nyquist. Per evitare aliasing si deve scegliere una frequenza di campionamento F_s sufficientemente maggiore della frequenza di Nyquist ($F_s \geq 2B$), per permettere ad un filtro con pendenza non infinita di separare il segnale dagli alias.

Definizioni

Si indichi con $\bar{x}(n)$ la sequenza intera, mentre con $x(n)$ il campione n -esimo. Dato un sistema L che trasforma $\bar{x}(n)$ nella sequenza $\bar{y}(n)$ si dice che esso è :

- Lineare $\Leftrightarrow L[a\bar{x}_1(n) + b\bar{x}_2(n)] = aL[\bar{x}_1(n)] + bL[\bar{x}_2(n)]$
- Tempo invariante $\Leftrightarrow \bar{y}(n - n_0) = L[\bar{x}(n - n_0)]$
- Con memoria o senza memoria se il campione n -esimo dell'uscita dipende o no dai campioni precedenti dell'ingresso.
- Causale \Leftrightarrow per $\bar{x}(n) : \bar{x}(n) = 0 \quad \forall n < 0$ si ha che $y(n) = 0 \quad \forall n < 0$

Si definiscono

- $\bar{\delta}(n) = \begin{cases} 1 & \text{se } n = 0 \\ 0 & \text{altrimenti} \end{cases}$
- $\bar{u}(n) = \begin{cases} 1 & \text{se } n \geq 0 \\ 0 & \text{altrimenti} \end{cases}$

Inoltre si definiscono energia e potenza associate alla sequenza $\bar{x}(n)$ come segue:

$$E(\bar{x}(n)) = \sum_{n=-\infty}^{+\infty} |x(n)|^2 \quad P(\bar{x}(n)) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=-N}^N |x(n)|^2 \quad (1.2)$$

Considereremo in seguito solo sequenze per cui vale:

$$E(\bar{x}(n)) = \sum_{n=-\infty}^{+\infty} |x(n)|^2 < \infty \quad (1.3)$$

che sono dette a energia finita.

Sistemi lineari tempo invarianti

In base alla definizione di impulso discreto per una qualunque sequenza $\bar{x}(n)$ si può scrivere :

$$x(n) = \sum_{k=-\infty}^{+\infty} x(k)\delta(n-k) \quad (1.4)$$

Per sistemi lineari tempo invarianti si ha:

$$\begin{aligned} L[x(n)] &= L\left[\sum_{k=-\infty}^{+\infty} x(k)\delta(n-k)\right] = \sum_{k=-\infty}^{+\infty} x(k)L[\delta(n-k)] = \\ &= \sum_{k=-\infty}^{+\infty} x(k)h(n-k) \stackrel{\Delta}{=} x(n) * h(n) \end{aligned} \quad (1.5)$$

Dove $h(n) = L[\delta(n)]$ è detta risposta all'impulso del sistema e lo descrive completamente. Considero l'eccitazione $x(n) = e^{j2\pi fnT}$ la risposta del sistema è:

$$\begin{aligned} y(n) &= \sum_{k=-\infty}^{+\infty} h(k)e^{j2\pi f(n-k)T} = e^{j2\pi fnT} \sum_{k=-\infty}^{+\infty} h(k)e^{-j2\pi fkT} = \\ &= e^{j2\pi fnT} H(e^{j2\pi fT}) \end{aligned} \quad (1.6)$$

$H(e^{j2\pi fT}) \stackrel{\Delta}{=} {}^s H(f)$ è una funzione periodica in f con periodo $\frac{1}{T}$ detta funzione di trasferimento del sistema. ${}^s H(f)$ esiste se e solo se $h(n)$ è assolutamente sommabile. Dalla definizione di ${}^s H(f)$ segue che i campioni $h(n)$ sono i coefficienti dello sviluppo in serie di Fourier e si ha quindi

$$h(n) = \int_{f=-\frac{1}{2T}}^{\frac{1}{2T}} {}^s H(f)e^{j2\pi fnT} df \quad (1.7)$$

Di seguito si indicherà con ${}^a X(f)$ la trasformata del segnale analogico ${}^a x(t)$.

Se ${}^a x(t)$ è campionato per $t = nT$ si può scrivere

$${}^s X(f) = \frac{1}{T} \sum_{m=-\infty}^{+\infty} {}^a X\left(f - \frac{m}{T}\right) \quad (1.8)$$

La 1.8 indica il legame tra la trasformata del segnale discreto e quella del segnale analogico. Se inoltre $F_s = \frac{1}{T} > 2B$ le due coincidono su tutta la banda del segnale.

Trasformata Z

Per un segnale analogico è definita la trasformata di Laplace:

$${}^l X(s) = \int_{t=-\infty}^{+\infty} {}^a x(t) e^{-st} dt \quad (1.9)$$

Applicata ad un segnale campionato e valutata in $z = e^{sT}$ si ottiene

$${}^z X(z) = \sum_{n=-\infty}^{+\infty} {}^a x(nT) e^{snT} = \sum_{n=-\infty}^{+\infty} {}^s x(n) \cdot z^{-n} \quad (1.10)$$

che definisce la trasformata Zeta di $x(n)$ nella corona circolare $r_1 < |z| < r_2$ su cui converge. I punti sul bordo della corona possono appartenere alla regione di convergenza.

Si definiscono sequenze destre quelle per cui $r_2 = \infty$ cioè convergono sui punti del piano complesso esterni al cerchio di raggio r_1 , sequenze sinistre se $r_1 = 0$ cioè se il dominio di esistenza è la superficie interna al cerchio di raggio r_2 . Per definire univocamente la precedente serie ${}^z X(z)$ è necessario specificare la regione di convergenza. Infatti esistono serie con la stessa espressione di ${}^z X(z)$ ma domini

di esistenza diversi. Se la circonferenza di raggio unitario appartiene alla regione di convergenza si può valutare la 1.10 in $z = e^{j2\pi fT}$ ottenendo:

$$X(e^{-j2\pi fT}) = \sum_{n=-\infty}^{+\infty} x(nT)e^{-j2\pi fnT} \quad (1.11)$$

cioè la trasformata di Fourier della sequenza $\bar{x}(n)$. Inoltre si ha la seguente formula di inversione:

$$x(n) = \frac{1}{j2\pi} \oint_C X(z)z^{n-1} dz \quad (1.12)$$

dove C è una qualunque curva chiusa appartenente alla regione di convergenza e contenete l'origine.

Calcolando la trasformata di una sequenza ritardata di m campioni $x(n-m)$, si ottiene:

$$\begin{aligned} Z[x(n-m)] &= \sum_{n=-\infty}^{+\infty} x(n-m) \cdot z^{-n} = \\ &= z^{-m} \sum_{n=-\infty}^{+\infty} x(n-m) z^{-(n-m)} = z^{-m} X(z) \end{aligned} \quad (1.13)$$

Da cui si deduce che il ritardo di un campione produce nel piano Z la moltiplicazione per z^{-1} .

Trasformata discreta di Fourier (DFT)

Per rappresentare una sequenza finita di campioni, si può utilizzare la trasformata di Fourier discreta (DFT), essa è formata da un numero finito di campioni della trasformata di Fourier ugualmente spaziate in frequenza.

Le formule che descrivono la serie di Fourier discreta [2], definendo $W_N = e^{j2\pi/N}$ sono:

$$\tilde{X}(k) = \sum_{n=0}^{N-1} \tilde{x}(n) W_N^{-kn} \quad (1.14)$$

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{X}(k) W_N^{-kn} \quad (1.15)$$

Le sequenze $\tilde{X}(k)$ e $\tilde{x}(n)$ risultano periodiche di periodo N e in particolare la sequenza $\tilde{X}(k)$ può essere vista come la sequenza di campioni equispaziati sul cerchio di raggio unitario della trasformata Z di un periodo della sequenza $\tilde{x}(n)$. Questa relazione si ottiene considerando $\bar{x}(n)$ tale da rappresentare un periodo di $\tilde{x}(n)$, ovvero $x(n) = \tilde{x}(n)$ per $0 \leq n \leq N-1$ e $x(n) = 0$ altrimenti.

La trasformata Z di $\bar{x}(n)$ è data da:

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} = \sum_{n=0}^{N-1} x(n) z^{-n} \quad (1.16)$$

Pertanto la (1.14) può essere espressa nella seguente forma:

$$\tilde{X}(k) = X(z) \Big|_{z=e^{j(2\pi/N)k}=W_N^{-k}} \quad (1.17)$$

Questo dimostra come $\tilde{X}(k)$ sia uguale a N campioni di $X(z)$ sul cerchio di raggio unitario egualmente spazati.

L'espressione della DFT si può ricavare partendo dalla trasformata Z di una sequenza non periodica $\bar{x}(n)$

$$X(z) = \sum_{n=-\infty}^{+\infty} x(n) z^{-n} \quad (1.18)$$

Per cui la regione di convergenza comprenda il cerchio unitario (condizione sempre verificata se la sequenza è di lunghezza finita).

Campionando la trasformata Z in N punti equispaziati sul cerchio unitario si ottiene la sequenza periodica

$$\tilde{X}(k) = \sum_{-\infty}^{\infty} x(n)W_N^{kn} \quad (1.19)$$

Utilizzando la (1.15) e la (1.19) si può ricavare la relazione tra la sequenza periodica $\tilde{x}(n)$ e la sequenza considerata all'inizio $\bar{x}(n)$.

$$\tilde{x}(n) = \frac{1}{N} \sum_{k=0}^{N-1} \sum_{m=-\infty}^{+\infty} x(m)W_N^{km}W_N^{-kn} = \sum_{m=-\infty}^{+\infty} x(m) \left[\frac{1}{N} \sum_{k=0}^{N-1} W_N^{-k(n-m)} \right] \quad (1.20)$$

Ricordando:

$$\sum_{k=0}^{N-1} W_N^{-k(n-m)} = \begin{cases} N & \text{se } n = m \\ 0 & \text{altrimenti} \end{cases}$$

si ha:

$$\tilde{x}(n) = \sum_{r=-\infty}^{+\infty} x(n+rN) \quad (1.21)$$

La sequenza periodica $\tilde{x}(n)$ si ricava pertanto da sovrapposizioni successive della sequenza aperiodica $\bar{x}(n)$, se poi essa ha una lunghezza inferiore a N ogni periodo di $\tilde{x}(n)$ ne costituisce una replica, altrimenti si ha una sovrapposizione di valori non nulli che producono un fenomeno di “*time aliasing*”.

Nel caso in cui la sequenza $\bar{x}(n)$ sia di durata inferiore a N essa può essere ricostruita tramite $\tilde{x}(n)$ estraendone un periodo, questo implica anche che la

sequenza originale si può ottenere da N valori equispaziati della $X(z)$ sul cerchio unitario ovvero dalla serie di Fourier Discreta.

Si può riassumere quanto ricavato con le seguenti espressioni:

$$X(k) = \begin{cases} \sum_{n=0}^{N-1} x(n)W_N^{kn} & 0 \leq k \leq N-1 \\ 0 & \text{altrimenti} \end{cases} \quad (1.22)$$

$$x(n) = \begin{cases} \frac{1}{N} \sum_{k=0}^{N-1} X(k)W_N^{-kn} & 0 \leq k \leq N-1 \\ 0 & \text{altrimenti} \end{cases} \quad (1.23)$$

La (1.22) definisce la trasformata di Fourier Discreta $X(k)$ della sequenza $x(n)$ mentre la (1.23) definisce l'antitrasformata.

Dalle formule si ricava che il costo computazionale è proporzionale a N^2 , e per sequenze lunghe di campioni può diventare molto lungo. Inoltre il calcolo della DFT richiede di conoscere il valore di tutti i campioni della sequenza $x(n)$, che comporta l'introduzione di un ritardo elevato per sequenze lunghe.

L'utilizzo di algoritmi di calcolo FFT (più efficienti della DFT) e l'uso di algoritmi di "overlap and save" permettono rispettivamente di ridurre la complessità computazionale ed il ritardo.

Fast Fourier Transform (FFT)

Uno degli algoritmi matematici che calcola lo spettro a partire dai campioni quantizzati si chiama FFT (Fast Fourier Transform). Esso opera sugli N campioni reali nel dominio del tempo trasformandoli in $N/2+1$ campioni complessi nel dominio della frequenza con spaziatura costante. L'algoritmo produce un andamento lineare sull'asse delle frequenze. Una rappresentazione di questo tipo attribuisce la stessa importanza a tutte le frequenze.

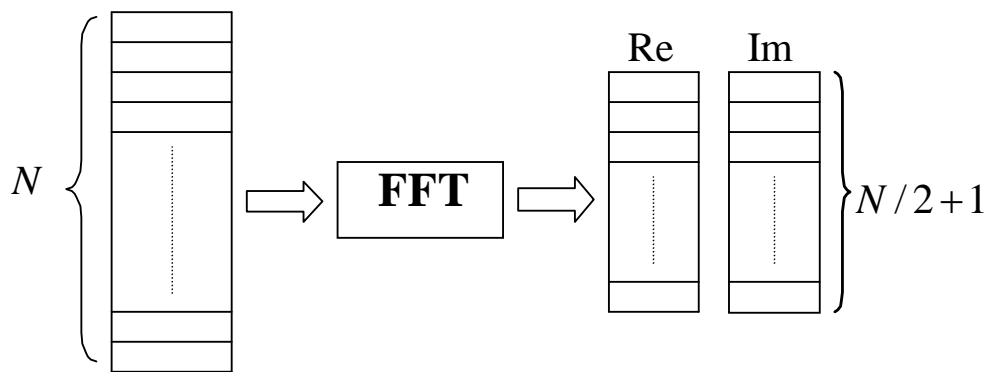


Figura 1.1 - Schema di algoritmo FFT

Il primo elemento che si ottiene è relativo alla frequenza nulla, ha valore reale e rappresenta il valore medio del segnale, cioè la sua componente continua. Il secondo elemento ha frequenza:

$$f_1 = \frac{1}{N \cdot T} \quad (1.24)$$

con T il periodo di campionamento. Le frequenze successive sono:

$$f_i = i \cdot f_1 \quad (1.25)$$

con $i = 1, 2, 3, \dots, N/2$ dove l'ultima frequenza è quella di Nyquist.

Dal punto di vista del calcolo della trasformata, l'algoritmo FFT è più efficace se nella sua esecuzione si utilizza la rappresentazione cartesiana invece che quella polare, e se viene scelto N tale che sia potenza di 2, ad esempio $N = 4096$. Applicando l'algoritmo si ottengono N linee spettrali; in realtà è possibile considerare solo le prime $N/2+1$ dato che le successive sono una replica coniugata delle precedenti.

1.2 L'equalizzazione

Il problema dell'equalizzazione di un ambiente acustico può essere schematizzato per semplicità con una rappresentazione a singolo canale.

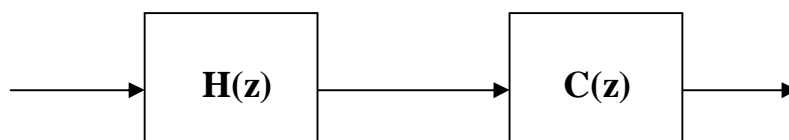


Figura 1.2 - Schema a blocchi di un canale del sistema acustico con equalizzazione

L'altoparlante e l'ambiente costituiscono nel loro insieme un sistema che, in prima approssimazione, può essere considerato lineare e tempo invariante (LTI) e del quale è dunque possibile misurare la risposta all'impulso. Siccome tutte le trasformazioni che opereremo avverranno nel dominio tempo-discreto, il sistema altoparlante-ambiente sarà indicato con la trasformata Z dei campioni della sua risposta impulsiva che chiameremo $C(z)$.

Altoparlante e ambiente (nel caso specifico l'abitacolo di un'autovettura) sono dunque il sistema da equalizzare nell'ipotesi che il segnale elettrico che pilota l'altoparlante sia privo di distorsioni (o che esse siano trascurabili). L'equalizzazione del sistema si ottiene antepoendo a $C(z)$ un sistema LTI denominato $H(z)$ per pre-distorcere il segnale elettrico che pilota l'altoparlante e rendere la risposta complessiva del sistema acustico quella desiderata.

E' opportuno sottolineare che i risultati che si ottengono utilizzando questa tecnica, sono buoni soltanto se il sistema altoparlante-ambiente si trova a lavorare in condizioni lineari o in presenza di piccole non linearità. Se viceversa ci si trova in presenza di un sistema fortemente non lineare, in generale non sarà possibile equalizzarlo tramite la pre-distorsione del segnale elettrico di pilotaggio, ma sarà necessario intervenire direttamente sui trasduttori, per aumentarne la dinamica, o sulle eventuali risonanze dell'abitacolo tentando di smorzarle. In generale gli equalizzatori sia di tipo analogico sia di tipo digitale sono uno strumento molto potente e versatile ma devono essere utilizzati con la consapevolezza che per

funzionare correttamente necessitano di sistemi di riproduzione con un adeguato livello di potenza e di ambienti acustici trattati per l'eliminazione dei fenomeni fortemente non lineari.

Non è corretto ritenere che con l'utilizzo dei sistemi di equalizzazione sia possibile ottenere sempre buone prestazioni indipendentemente dai trasduttori elettroacustici e dall'ambiente.

La risposta in frequenza o meglio la curva di risposta ottenuta tenendo in considerazione gli aspetti psicoacustici che influenzano la risposta percepita dagli ascoltatori (vedi capitoli 3 e 4) ha un peso molto elevato nella determinazione dell'indice di gradimento dell'impianto audio. I test di ascolto e valutazione effettuati [3], hanno evidenziato che gli impianti con una risposta in frequenza "piatta" nel settore automotive, non hanno le valutazioni più elevate, anzi in generale non sono particolarmente apprezzati. Da test acustici condotti in [4] e nella presente tesi (capitolo 2 e paragrafo 4.4) è stata ricavata la "curva target" cui si cercherà di arrivare durante la taratura dell'impianto per ottenere buoni risultati.

Filtri per l'equalizzazione della risposta in frequenza

Negli impianti di fascia alta che prevedono l'utilizzo di un DSP (i Digital Signal Processor sono microprocessori la cui architettura è ottimizzata per il calcolo dei filtri digitali) si possono utilizzare diverse soluzioni per ottenere una buona equalizzazione.

Principalmente esistono due categorie di filtri digitali:

- Filtri di tipo IIR (*Infinite Impulse Response*).
- Filtri di tipo FIR (*Finite Impulse Response*).

Il progetto di filtri IIR è, in generale, un problema di straordinaria complessità e non si conoscono soluzioni completamente soddisfacenti. Inoltre, essendo la risposta all'impulso di un filtro IIR di durata infinita, essa non può avere un punto di simmetria che è condizione necessaria per avere una fase lineare.

Nello sviluppo del progetto si è scelto di usare filtri FIR, quindi nella trattazione seguente si farà riferimento solo a filtri di questo tipo.

Filtri FIR

L'utilizzo di filtri FIR richiede un'elevata potenza di calcolo se si vuole ottenere l'equalizzazione della risposta alle basse frequenze (per avere una buona risoluzione in frequenza si devono usare filtri FIR con un elevato numero di coefficienti), però permette di utilizzare procedure automatiche di sintesi dei filtri e non presenta particolari problemi nell'implementazione su piattaforme DSP a virgola fissa (non essendo presente alcuna retroazione nella struttura del filtro non ci sono problemi d'instabilità).

Un filtro di tipo FIR si può definire attraverso la seguente equazione di tipo discreto:

$$y(n) = c_0x(n) + c_1x(n-1) + \dots + c_{N-1}x(n-N+1) \quad (1.26)$$

Dove i coefficienti $\{c_i\}$ sono chiamati "tap" del filtro.

La 1.1 può essere riscritta sotto forma di funzione di trasferimento come:

$$H(z) = \sum_{i=0}^{N-1} c_i z^{-i} \quad (1.27)$$

Da queste equazioni si ricava direttamente la struttura di calcolo dei filtri FIR sostanzialmente composta da una catena di ritardo (registro a scorrimento o meglio buffer circolare), da blocchi moltiplicatori e blocchi sommatore.

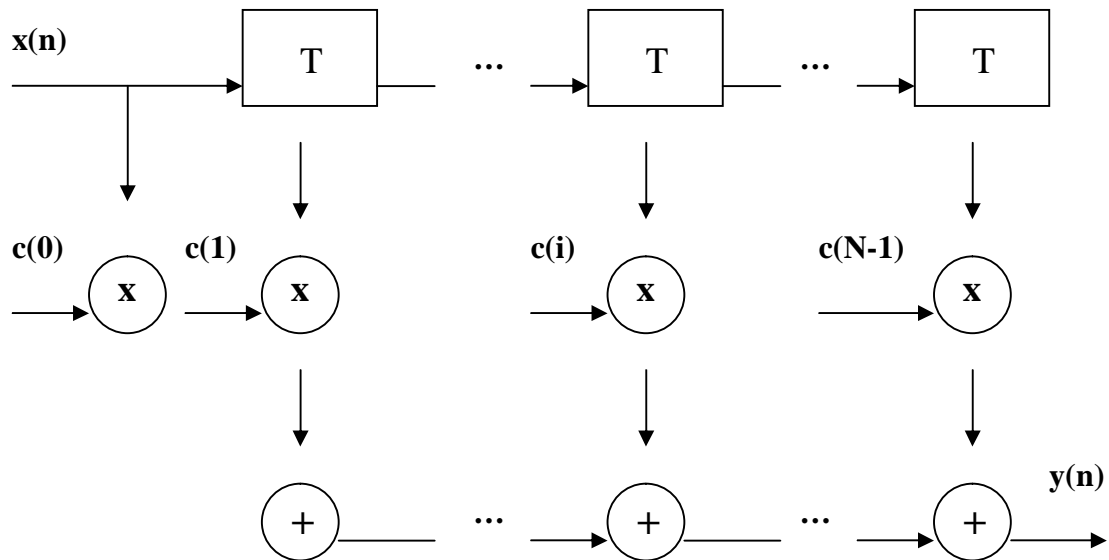


Figura 1.3 - Struttura di calcolo dei filtri di tipo FIR

Questa struttura ha le seguenti proprietà:

- E' sempre stabile anche se implementata su macchine con precisione finita, perché l'uscita dipende unicamente dagli ultimi N campioni dell'ingresso e quindi non è presente alcun fenomeno d'accumulo di errori di troncamento.
- Essa è realizzata in modo molto efficiente sui DSP, attraverso l'uso di operazioni di MAC (moltiplicazione e accumulo) tipicamente eseguibili in un solo ciclo di clock.
- Esistono algoritmi di sintesi soddisfacenti per la sintesi di filtri FIR di forma arbitraria.
- Se l'implementazione della convoluzione è fatta utilizzando una FFT partizionata si può ridurre la potenza di calcolo per l'implementazione dei filtri FIR con un elevato numero di "taps".

1.3 Metodo di inversione di Kirkeby

La sintesi di un sistema di equalizzazione ha come obiettivo quello di modificare la curva di pressione acustica portandola ad una curva piatta in modulo o più correttamente alla curva che da studi statistici su impianti di riproduzione musicale ha soddisfatto il maggior numero di ascoltatori (essa è denominata curva di comfort acustico o curva target).

Seguendo un approccio di questo tipo la sintesi dell'equalizzatore con struttura di tipo FIR (preferibili a filtri IIR per la semplicità di sintesi) si basa sulla operazione di inversione della curva di pressione acustica in funzione della frequenza per portarla a quella desiderata.

In questo capitolo verrà illustrato il metodo utilizzato per la sintesi di tali filtri nel caso di sistema single-input single-output (come un normale un normale impianto audio stereofonico o un sistema di cuffie).

Le tecniche disponibili per l'elaborazione dei filtri inversi FIR sono parecchie. Le più utilizzate risultano essere:

- filtri a fase minima
- filtri di Kirkeby
- filtri di N. Allen
- filtri di Morjopoulos

Ciascuna di queste tecniche di filtraggio presenta vantaggi e svantaggi che esulano da questa trattazione. In questa sede la tecnica di inversione utilizzata risulta essere quella di Kirkeby.

Inversione di Kirkeby single-input single output

La prima fase nell'elaborazione di un filtro inverso prevede la misura della funzione di trasferimento del sistema ottenuta utilizzando come sorgente l'impianto di riproduzione e come trasduttore un sistema microfonic. In un sistema cuffie questo si ottiene ponendo le stesse sopra una testa acustica e misurando separatamente la risposta da canale destro a canale destro e viceversa. In questo caso non è presente fenomeno di cross-talk acustico. Secondo la

notazione di figura 1.2, il risultato della misura è una singola risposta all'impulso stereo c . L'obiettivo è trovare una seconda risposta all'impulso h la cui convoluzione con la risposta c misurata restituisca una Delta di Dirac δ . Questo espresso nei domini del tempo e della frequenza risulta:

$$c(t) \otimes h(t) = \delta(t) \leftrightarrow C(f) \bullet H(f) = 1 \quad (1.28)$$

Dal momento che nel dominio della frequenza la convoluzione diventa una moltiplicazione che può essere condotta separatamente per ogni linea spettrale, il filtro inverso H ricercato si ottiene calcolando il reciproco della funzione spettrale complessa di C misurata.

$$H(f) = \frac{1}{C(f)} \quad (1.29)$$

Sfortunatamente questo semplice approccio risulta spesso non ottimale. Infatti, tipicamente, C non risulta essere a fase minima e pertanto non ammette inversione diretta. Solo una inversione approssimata può restituire un filtro inverso che sia stabile, causale ed a lunghezza finita. La tecnica elaborata da Kirkeby prevede di calcolare il reciproco nel dominio della frequenza aggiungendo un piccolo parametro di regolarizzazione al denominatore. Tale parametro ε permette inoltre, in corrispondenza delle bande estreme, di evitare una eccessiva regolarizzazione mantenendo il filtro stabile. Pertanto l'idea è quella di mantenere un piccolo valore del parametro nella zona centrale dello spettro dove si vuole una inversione accurata e un grosso valore agli estremi dove non si ha possibilità di inversione a causa di limiti fisici del trasduttore. Il valore di ε è calcolato attraverso un approccio a tentativi, e definito da un compromesso fra la lunghezza del filtro inverso che si vuole ottenere e l'accurata inversione di picchi spettrali e buchi. Tipicamente è difficile ottenere un valore ottimo per tutto lo spettro di frequenze e il risultato non è mai ottimale. Per tale motivo viene utilizzato un ε dipendente dalla frequenza:

$$H(f) = \frac{C'(f)}{C'(f) \cdot C(f) + \varepsilon(f)} \quad (1.30)$$

dove C' rappresenta il complesso coniugato di C .

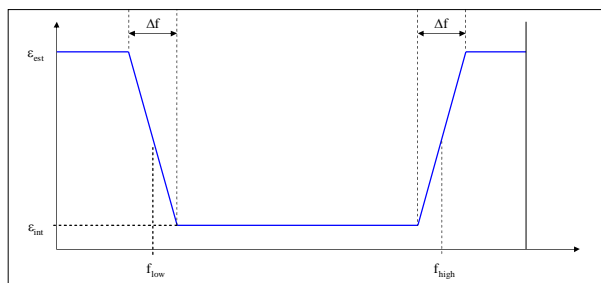


Figura 1.4 – Andamento tipico del parametro di regolarizzazione ε in funzione della frequenza

1.4 Wavelet Transform

La trasformata di Fourier, largamente utilizzata nell'analisi spettrale dei segnali, evidenzia la presenza delle componenti armoniche, ma non permette di ricavare facilmente informazioni su *quando* e *come* tali frequenze siano effettivamente presenti.

La trasformata di Fourier rappresenta infatti il prodotto scalare fra il segnale e una funzione sinusoidale complessa di durata temporale infinita, la quale è perfettamente locale in frequenza, ma assolutamente globale nel tempo. In altri termini, questa trasformazione è adatta a segnali stazionari perché permette di evidenziare nettamente fenomeni che avvengono nel dominio della frequenza ma non discrimina eventi che avvengono nel dominio del tempo.

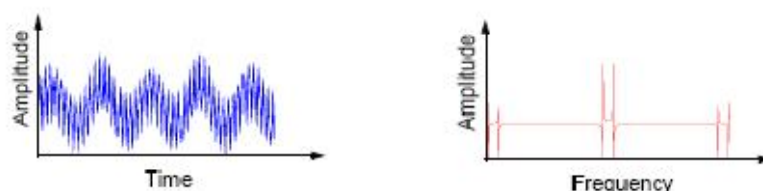


Figura 1.5 – Analisi di Fourier

Per i segnali non stazionari occorre quindi inserire nella trasformazione una dipendenza dal tempo. Il modo più immediato per ottenere questo risultato consiste nel rendere *locale* l'analisi operando su porzioni del segnale individuate da una finestra che trasla nel tempo. Nasce così la *trasformata di Fourier a breve termine* o *Short Time Fourier Transform*:

$$STFT_x(\tau, f) = \int_{-\infty}^{+\infty} x(t)g^*(t - \tau)e^{-j2\pi ft} dt = \langle g(t - \tau)e^{j2\pi ft}, x(t) \rangle \quad (1.32)$$

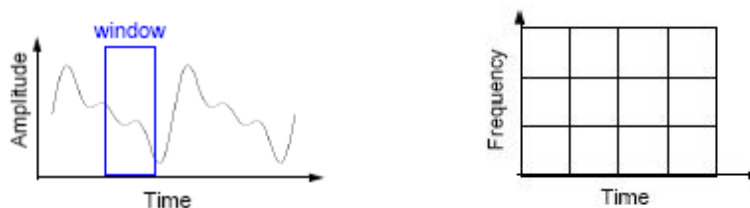


Figura 1.6 – Analisi STFT

Con questa nuova trasformazione si evidenzia lo spettro del segnale all'interno della finestra temporale definita dalla funzione $g(t-\tau)$, ottenendo un'informazione sul suo contenuto armonico in un intorno dell'istante di tempo τ . Questo approccio ha l'intrinseco difetto di moltiplicare nel dominio del tempo il segnale $x(t)$ con la funzione finestra $g(t)$ convolvendo quindi i relativi spettri nel dominio della frequenza. L'alterazione in frequenza dovuta al contributo spettrale della finestra $g(t)$ ci pone di fronte a una situazione di *trade-off* tra precisione dell'analisi temporale e dell'analisi in frequenza.

In molte applicazioni non è possibile permettersi di localizzare fenomeni lenti e fenomeni veloci con la stessa indeterminazione, nel senso che identificare una frequenza di 100MHz con precisione $\Delta f=1\text{kHz}$ è molto diverso dall'identificare 100kHz con la stessa precisione.

Per ottenere un'analisi a risoluzione variabile occorre fare in modo che le risoluzioni relative $\Delta t/t$ e $\Delta f/f$ risultino costanti e questo richiede che all'aumentare della frequenza f aumenti in modo proporzionale la banda Δf . A tale riguardo viene in aiuto una proprietà fondamentale della trasformata di Fourier: comprimendo nel tempo una funzione si ottiene una espansione in frequenza del suo spettro, e viceversa

$$\mathfrak{F}\left\{x\left(\frac{t}{a}\right)\right\} = |a|X(af) \quad (1.33)$$

Da questa considerazione nasce l'idea di sostituire l'operazione di *modulazione* con l'operazione di *scalamento*, ovvero anziché moltiplicare il segnale per la finestra $g(t)$ ad ampiezza temporale costante ed \mathcal{F} -trasformare si esegue direttamente il prodotto scalare con scalamenti e traslazioni di un unico prototipo. Quello che si ottiene prende il nome di *trasformata wavelet continua* (CWT):

$$CWT_x(a, b) = \int_{-\infty}^{+\infty} x(t)\psi_{ab}^*(t)dt = \langle \psi_{ab}(t), x(t) \rangle \quad (1.34)$$

con

$$\psi_{ab}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right) \quad (1.35)$$

Il prototipo $\psi(t)$ prende il nome di *wavelet madre*; a è il parametro di scalamento, b è il parametro di traslazione. La denominazione *wavelet* deriva dal fatto che, graficamente, il prototipo è una funzione che oscilla e si smorza come una piccola onda. Un esempio classico di ondina è riportato in fig. 1.7.

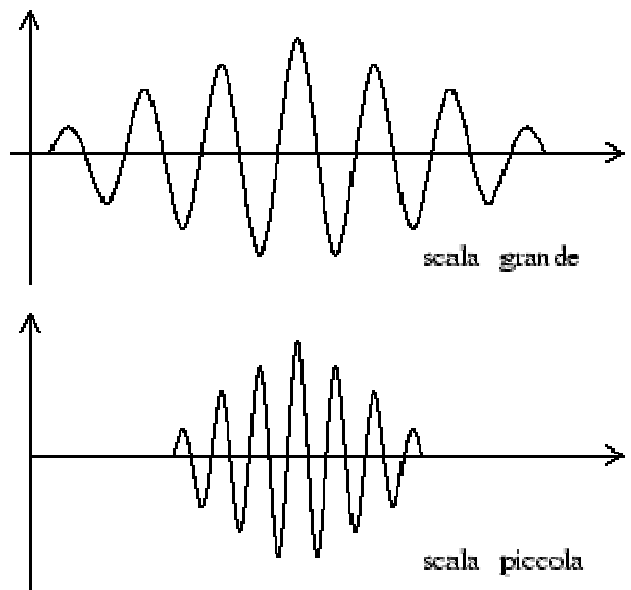


Figura 1.7 – Wavelet al variare del parametro di scalamento

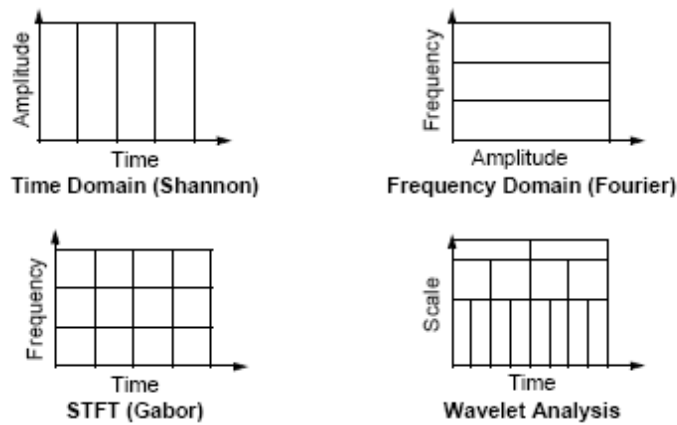


Figura 1.8 – Confronto tra i diversi tipi di analisi

Si osservi subito un particolare fondamentale; valutando la trasformata di Fourier dell'ondina, a meno del parametro di traslazione b , si ottiene

$$\mathfrak{F}\left\{\frac{1}{\sqrt{|a|}}\psi\left(\frac{t}{a}\right)\right\} = \psi(af) \tag{1.36}$$

essendo

$$\Psi(f) = F\{\psi(t)\} \tag{1.37}$$

Se $\psi(t)$ ha banda $\Delta(f)$ centrata in f_0 , allora $\psi(\frac{t}{a})$ ha banda $\frac{\Delta(f)}{a}$ centrata in $\frac{f_0}{a}$: la banda relativa rimane costante.

Come interpretare l'informazione fornita da una rappresentazione tempo-scala? Occorre svincolarsi dal concetto di frequenza, poiché confrontando la definizione di trasformata di Fourier con la definizione di trasformata *wavelet* continua (1.34), è immediato notare che il prodotto scalare del segnale $x(t)$ viene ora eseguito con una funzione non periodica e limitata nel tempo: il concetto di frequenza dell'armonica $e^{j2\pi ft}$ viene sostituito con il concetto di scala dell'ondina $\psi_{ab}(t)$. L'analogia è tuttavia immediata: valori piccoli di a significano ondine compresse nel tempo, quindi contenenti armoniche ad alta frequenza; effettuare il prodotto scalare con esse implica ottenere proiezioni che portano informazioni sui dettagli del segnale, cioè su fenomeni rapidamente variabili; d'altra parte, valori grandi di a comportano ondine lentamente variabili, con banda stretta, che invece colgono il comportamento del segnale a lungo termine. Queste fondamentali proprietà sono note come *localizzazione temporale* (più l'ondina è concentrata nel tempo, migliore è la risoluzione in questo dominio) e come *localizzazione spettrale* (più l'ondina è concentrata in frequenza, migliore è la risoluzione in questo dominio).

Tra le altre proprietà della trasformata *wavelet* è opportuno ricordare la *linearità*, *l'invarianza per traslazioni temporali* e *l'invarianza per scalamento e invertibilità*.

La definizione di trasformata *wavelet* continua è utilizzabile nei casi in cui si desideri una valutazione analitica della CWT di un segnale. Tuttavia nella maggior parte dei casi pratici, il segnale $x(t)$ non solo non è noto analiticamente (per esempio, quando è fornito da un sensore) ma risulta quantizzato, cioè è conosciuto esclusivamente in precisi istanti di tempo; se si considera inoltre il desiderio di una valutazione numerica della trasformata, allora essa dovrà essere valutata solo con un numero finito di valori delle variabili a e b .

La scelta non accurata dei parametri ovviamente può determinare da una parte una rappresentazione ridondante, ossia il numero di campioni della

trasformata risulta molto più grande di quello del segnale, dall'altra una rappresentazione non completa, cioè il numero di campioni non è sufficiente per la ricostruzione. La situazione ideale è la rappresentazione *ortogonale*, ovvero il segnale e la sua trasformata hanno lo stesso numero di campioni. Si considerino le seguenti osservazioni:

- al variare della scala a le ondine hanno banda relativa costante e costante è l'indeterminazione $\frac{\Delta(f)}{f}$ sulla frequenza. È quindi ragionevole effettuare una discretizzazione logaritmica di questo parametro.
- quando la scala è piccola le ondine hanno supporto temporale corto, quindi è logico che il passo di traslazione Δb sia piccolo per garantire la sufficiente copertura dell'asse-tempi da parte della serie di ondine.
- viceversa, quando la scala è grande le ondine sono lunghe, per cui il passo di traslazione può essere più grande.

A partire da questa forma di CWT a parametri discreti, una successiva quantizzazione nel dominio del tempo t permette di introdurre la versione definitiva completamente discretizzata nota come DWT (*Discrete Wavelet Transform*); l'aspetto più notevole è che tale rappresentazione è implementabile in modo estremamente efficiente mediante un banco di filtri discreti iterato che esegue una codifica a sottobande.

CAPITOLO 2

Levigatura e decimazione spettrale

La levigatura o smoothing spettrale è una tecnica di trattamento dello spettro che, applicata alle tecniche di filtraggio ed equalizzazione mediante filtri FIR viste in precedenza, permette sia il miglioramento della qualità sonora di equalizzazione, sia la riduzione del costo computazionale. Nel capitolo verranno descritte le tecniche utilizzate ed i risultati sperimentali ottenuti.

2.1 Lo smoothing

L'algoritmo in questione prevede l'utilizzo della rappresentazione polare di una DFT. Una volta ottenuti modulo e fase, è necessario eseguire un'operazione di unwrap sulla fase (vedi paragrafo 2.3).

In formule:

$$H[k] = DFT(IR[n]) \quad (2.1)$$

$$M[k] = |H[k]| \quad (2.2)$$

$$P[k] = unwrap(\angle H[k]) \quad (2.3)$$

A tal punto si procede ad un'operazione di media a finestra mobile. Ogni campione del modulo (o fase) risultante in uscita sarà la media del campione in ingresso stesso con i campioni adiacenti. E' necessario adottare una pesatura dei coefficienti di media al fine di considerare in maniera piu marcata i campioni centrali e dare meno peso ai campioni man mano che ci si allontana dal centro. Designando con W il vettore dei coefficienti di pesatura risulterà:

$$M'[k] = \sum_{i=0}^{L-1} M[k - \alpha + i] \cdot W[i] \quad (2.4)$$

dove

$$L = 2\alpha + 1 \quad (2.5)$$

è la lunghezza di W . In figura 2.1, è possibile osservare un confronto tra il modulo di una DFT prima e dopo l'applicazione dello smoothing.

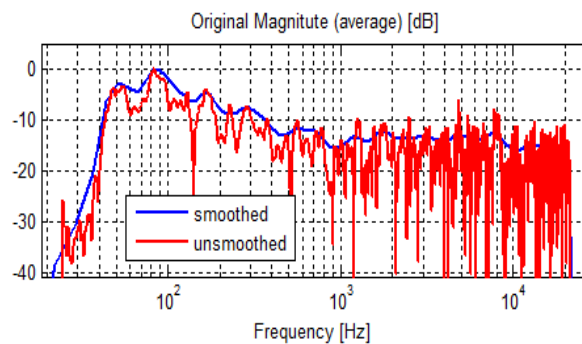


Figura 2.1: risposta in frequenza originale (unsmoothed) e levigata (smoothed).

La forma della finestra di pesatura è quella di Hanning. Per quel che riguarda la lunghezza L , sono state valutate diverse opzioni ispirate ai lavori precedentemente svolti nel campo dell'acustica domestica [5]. Seguendo tali esempi, sono state adottate finestre di lunghezza variabile in funzione della frequenza secondo varie leggi di dipendenza:

- Critical Bands (CB);
- Equivalent Rectangular Bandwidth (ERB);
- Double Octave Fraction (DOF);

le prime due seguono le note relative teorie psicoacustiche mentre l'ultima consiste nell'adottare bande di ventiquattresimi di ottava sotto la frequenza di Schroeder (nel caso dell'abitacolo di un'automobile circa 800 Hz) e bande di terzi d'ottava sopra tale frequenza. E' di seguito riportato in figura 2.2 l'andamento della lunghezza L in funzione della frequenza nei tre casi.

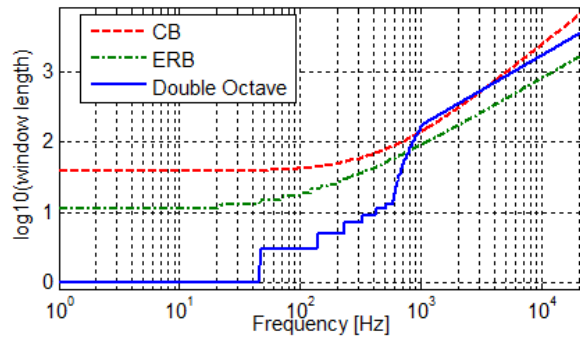


Figura 2.2: lunghezza della finestra di media in funzione della frequenza.

E' possibile notare nella linea blu (DOF) alcuni salti a bassa frequenza. Ciò non deve destare preoccupazione in quanto si tratta di salti di larghezza della finestra di pesatura, non del modulo della funzione di trasferimento.

Come affermato in precedenza è necessario eseguire un unwrap della fase prima dello smoothing. Tale operazione può essere svolta utilizzando vari algoritmi alcuni basati sul Cepstrum o sulla fattorizzazione polinomiale o una più semplice valutazione sulle discontinuità (Discontinuity Detect). Quest'ultimo algoritmo è stato scelto proprio per la sua semplicità e velocità.

Come già osservato da alcuni studi [6], lo smoothing spettrale apporta alcuni vantaggi se usato durante la sintesi di filtri numerici di equalizzazione. In primo luogo, lo smoothing permette un allargamento della zona di spazio d'ascolto entro cui la correzione stessa ha effetto; secondariamente permette di attenuare errori di misura dovuti alla debole non-stazionarietà del canale acustico.

Oltre a quanto sopra esposto, lo smoothing spettrale permette di ridurre il numero di campioni di uno spettro discreto. Com'è noto, è possibile aumentare la risoluzione dello spettro di una risposta impulsiva aggiungendo campioni nulli in coda a tale risposta. Per ridurre la risoluzione è invece necessario decimare lo spettro (operando separatamente su modulo e fase) prelevando da esso un campione ogni N . In questo contesto è necessario uno smoothing preliminare alla decimazione al fine di rendere il campione prelevato rappresentativo anche dei campioni adiacenti. E' da notare che l'operazione qui descritta è completamente differente da un classico sottocampionamento e che inoltre lascia inalterata la frequenza di campionamento.

La riduzione del numero di punti di uno spettro condotta secondo lo schema sopra citato permette di intraprendere un'equalizzazione con le tecniche già viste (ad es. Kirkeby) anche se la risposta impulsiva misurata è più lunga di quella del filtro di correzione. Mediante le tecniche che prevedono un'inversione punto punto del sistema si ottiene un filtro della stessa lunghezza (o maggiore se si interpolano le linee spettrali di partenza). Se si riduce lo spettro del sistema misurato prima di calcolare il filtro si ottiene quindi un filtro di lunghezza inferiore.

2.2 Test soggettivi

La tecnica dello smoothing spettrale applicata all'equalizzazione di impianti audio all'interno di automobili è stata investigata mediante test d'ascolto soggettivi. Sono stati coinvolti 9 soggetti dotati di capacità d'ascolto medio-alta.

Il test consistette nel far ascoltare, all'interno di un'automobile, 4 brani musicali di differenti generi (tabella 2.1).

n.	Artist	Title
1	Daugherty, Neiberg and Reynolds	I'm Confessin' (That I Love You)
2	Nirvana	Smells like teen spirit
3	Papete Disco Club	Room 5
4	A. Vivaldi	Winter

Tabella 2.1: brani musicali utilizzati nel test d'ascolto.

Attraverso un particolare software, i soggetti furono in grado di cambiare autonomamente ed in tempo reale il tipo di equalizzazione in esame nonché il brano in ascolto senza interruzioni di riproduzione. Le equalizzazioni sotto test si differenziavano sotto due aspetti: target curve e larghezza della finestra di media (smoothing) spettrale. Le target curve due (hard, con maggiore amplificazione delle basse frequenze e soft, con minore amplificazione delle basse frequenze) e sono mostrate in figura 2.3

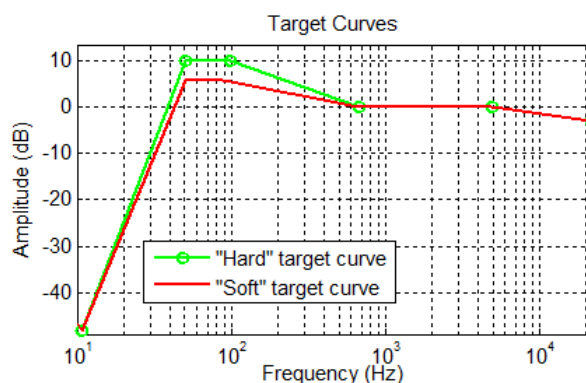


Figura 2.3: curve target utilizzate nella sintesi dei filtri di equalizzazione.

mentre le finestre di media sono quelle sopra citate (CB, ERB, DOF). Oltre a queste, fu utilizzata anche la configurazione relativa all'impianto audio originale dell'autovettura. Di seguito è riportato l'elenco delle 5 configurazioni utilizzate nate dalle combinazioni target curve – finestra di media.

- A – Soft + ERB;
- B – Hard + ERB;
- C – Originale (non filtrato);
- D – Soft + DOF;
- E – Hard + CB.

In figura 2.4 è possibile osservare alcuni spettri delle funzioni di trasferimento misurate con i filtri di equalizzazione inseriti.

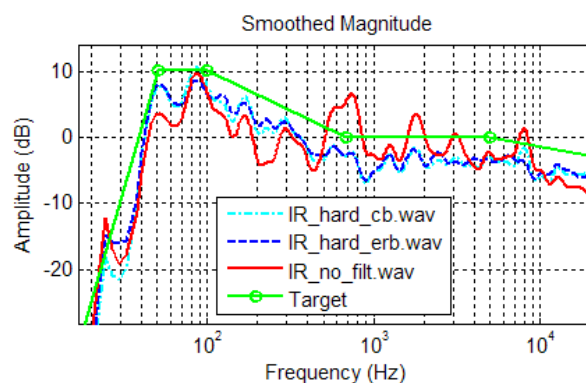


Figura 2.4: risposte in frequenza con filtri inseriti.

E' stata posta attenzione al livello di pressione sonora pesato A in quanto le differenze tra i vari filtri non superano il dB.

Il test fu eseguito all'interno di una vettura ospitante un impianto audio di fascia medio alta composto da un sub-woofer, un woofer ed un tweeter montati per ciascuna portiera anteriore, un medio centrale ed un amplificatore multicanale ASK. All'ingresso dell'amplificatore fu collegata una scheda audio per PC esterna.

Fu chiesto ai soggetti di compilare, durante il test, un questionario di valutazione relativo alle sensazioni d'ascolto di ogni singolo filtro proposto (figura 2.5). Per brevità è riportata solo la parte del questionario relativa ai filtri A e B.

Digital filtering questionnaire

Filter A									
Liking		Unpleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Pleasant
Treble		Too weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too loud
Bass		Too weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too loud
Voice		Undistinct	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Distinct
Stereo effect		Bad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Good
Distortion		Distorted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Undistorted
Warmth		Cold	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Warm
Colouring		Coloured	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	White

Filter B									
Liking		Unpleasant	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Pleasant
Treble		Too weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too loud
Bass		Too weak	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Too loud
Voice		Undistinct	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Distinct
Stereo effect		Bad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Good
Distortion		Distorted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Undistorted
Warmth		Cold	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	Warm
Colouring		Coloured	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	White

Figura 2.5: questionario di valutazione (filtri D, E ed F omessi).

Tale questionario comprende parametri di valutazione che possono aiutare non solo a giudicare il filtraggio ma anche quale aspetto cognitivo influenza il giudizio finale.

Segue una descrizione dei risultati statistici.

In figura 2.6 è mostrato il grafico relativo al gradimento dei filtri utilizzati.

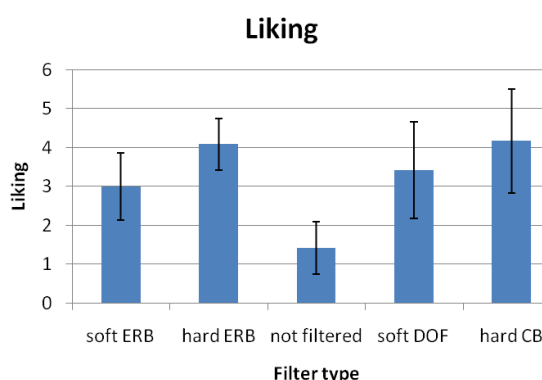


Figura 2.6: istogramma di gradimento dei filtri con indicatori di deviazione standard.

La risoluzione della scala utilizzata è di 1/5 (anche per le domande relative al livello di toni bassi e alti) e ciò rende difficile il confronto tra i filtri a causa dell'elevata deviazione standard. Per risolvere il problema è stato condotto un test t di Student per ogni coppia di filtri i cui risultati sono riportati in tabella 2.2.

Coppia di filtri	Percentuale di casualità
A B	0,22%
B C	< 0,01%
C D	0,01%
D E	16,00%
A C	< 0,01%
A D	34,80%
A E	1,83%
B D	11,54%
B E	84,87%
C E	< 0,01%

Tabella 2.2: test t di Student sul parametro "gradimento".

La seconda colonna riporta la probabilità che il risultato ottenuto sia dovuto a fattori casuali. In verde sono evidenziate le righe che riportano una percentuale inferiore al 5%, indicando che la differenza percepita tra le relative coppie di filtri è significativa.

Altri risultati sono ottenuti dalla ricerca di un legame tra parametri di valutazione soggettivi, in particolare tra il parametro di gradimento totale (Liking)

ed i parametri secondari. Nelle figure seguenti sono riportati i grafici relativi alle relazioni con coefficiente di correlazione maggiore di 0,8.

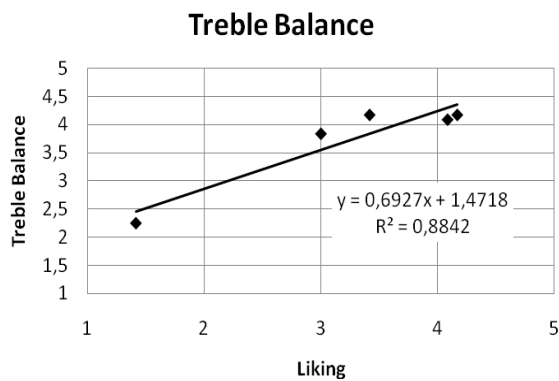


Figura 2.7: relazione "gradimento-bilanciamento alti" ("liking-treble balance").

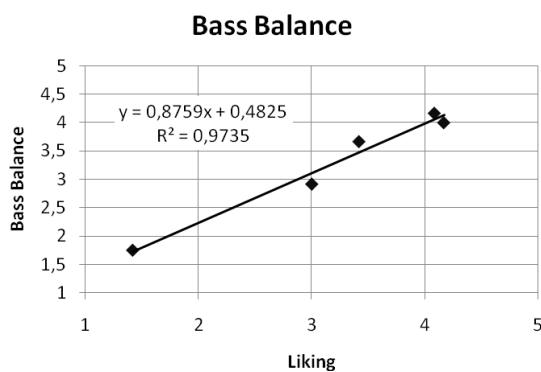


Figura 2.8: relazione "gradimento-bilanciamento bassi" ("liking-bass-balance").

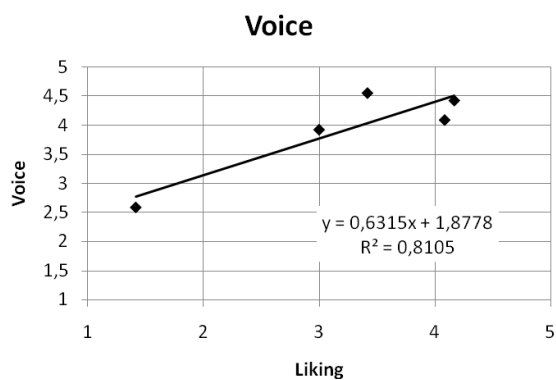


Figura 2.9: relazione "gradimento-voce" ("liking-voice").

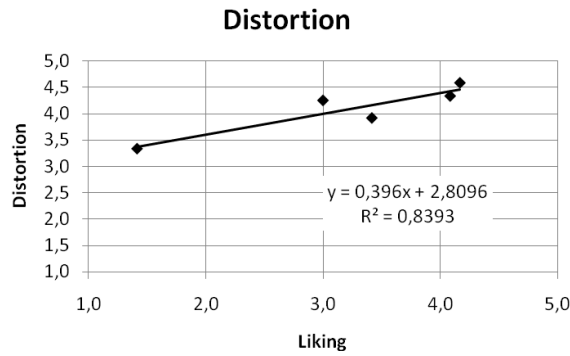


Figura 2.10: relazione “gradimento-distorsione” (“liking-distortion”).

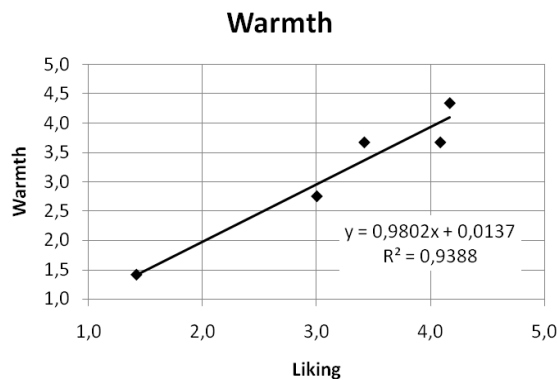


Figura 2.11: relazione “gradimento-calore” (“liking-warmth”).

Sono state individuate anche relazioni tra parametri soggettivi e oggettivi. Gli ultimi non sono altro che i livelli in bande d’ottava delle risposte impulsive misurate a filtri inseriti. Nelle figure si possono osservare i grafici relativi alle relazioni con coefficiente di correlazione maggiore di 0,7.

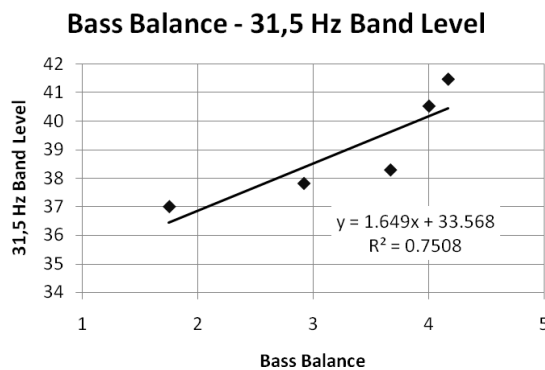


Figura 2.12: relazione “bilanciamento bassi-livello nella banda 31,5 Hz” (“bass balance-31.5 Hz band level”).

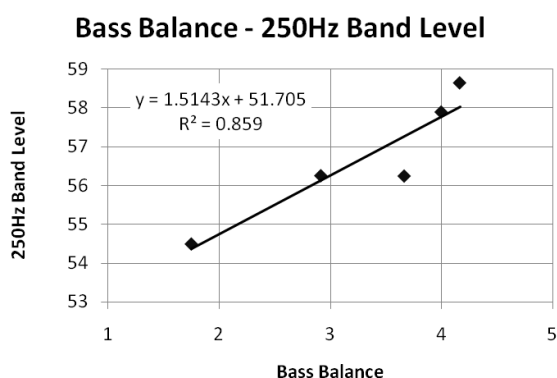


Figura 2.13: relazione “bilanciamento bassi-livello nella banda 250 Hz” (“bass balance-250 Hz band level”).

Utilizzando come parametri oggettivi la distanza spettrale (spectral distance) di uno spettro dalla curva target e la rugosità spettrale (spectral roughness), si ottengono i risultati evidenziati nelle figure 2.14 e 2.15.

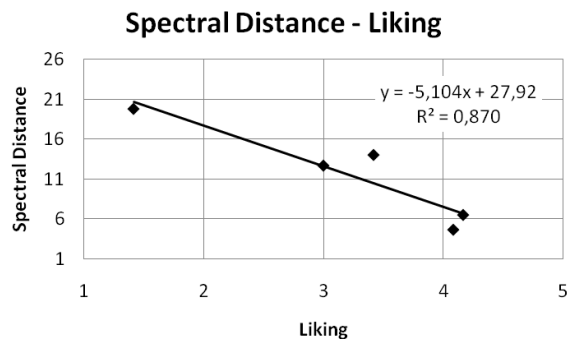


Figura 2.14: relazione “distanza spettrale-gradimento” (“spectral distance-liking”).

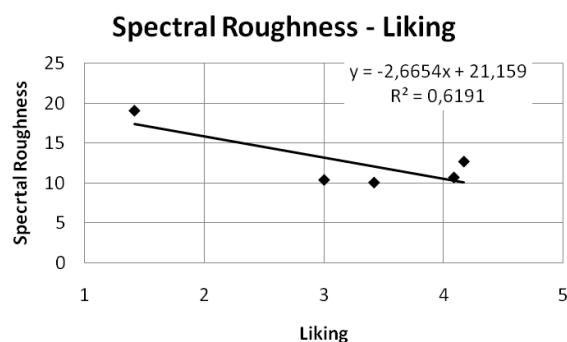


Figura 2.15: relazione “rugosità spettrale-gradimento” (“liking-spectral roughness”).

Seguono alcune considerazioni a conclusione del capitolo.

Osservando il test t di Student è possibile affermare che il gradimento d'ascolto aumenta attraverso i filtri digitali proposti. Emerge inoltre che i filtri migliori sono quelli relativi alla curva target "Hard", come evidenziato dalla colonna AB di tabella 2 e dalle relazioni "distanza spettrale-gradimento" di figura 2.14. Sfortunatamente non è stato possibile stabilire una preferenza dei soggetti relativamente al tipo di finestra di media spettrale ma solamente una relazione tra gradimento ed incremento della smoothness dei filtri (decremento della rugosità) come mostrato dalla figura 2.15. Risultati interessanti sono emersi dalla ricerca di relazioni tra i parametri soggettivi di gradimento globale ed i secondari. Sono state scoperte anche buone relazioni tra parametri soggettivi quali "bass balance" ed il livello sonoro in bande d'ottava.

2.3 Phase unwrapping

Nel presente paragrafo sono descritti i metodi utilizzati per il phase unwrapping (letteralmente "srotolamento della fase"), operazione necessaria per la decimazione spettrale sopra descritta.

Il metodo usato per eseguire l'operazione è denominato "discontinuity detect" (DD) e consiste nel riconoscere i salti di fase mediante il confronto con una soglia preimpostata ed eliminarli. Tale metodo è semplicemente realizzabile e di veloce esecuzione ma non assicura la correttezza del risultato. Operando nel dominio discreto, se la fase varia velocemente tra un punto e quello adiacente, c'è il pericolo che ciò venga confuso con un salto di fase e quindi rimosso. Al fine di scongiurare questo pericolo si è pensato di rimuovere il ritardo finito, corrispondente al tempo di volo acustico, che, come noto, contribuisce alla fase con una retta.

In un primo tempo si è pensato di stimare il tempo di volo relativo alla risposta impulsiva semplicemente osservando l'istante in cui l'involuppo della risposta impulsiva assume il valore massimo (figura 2.16).

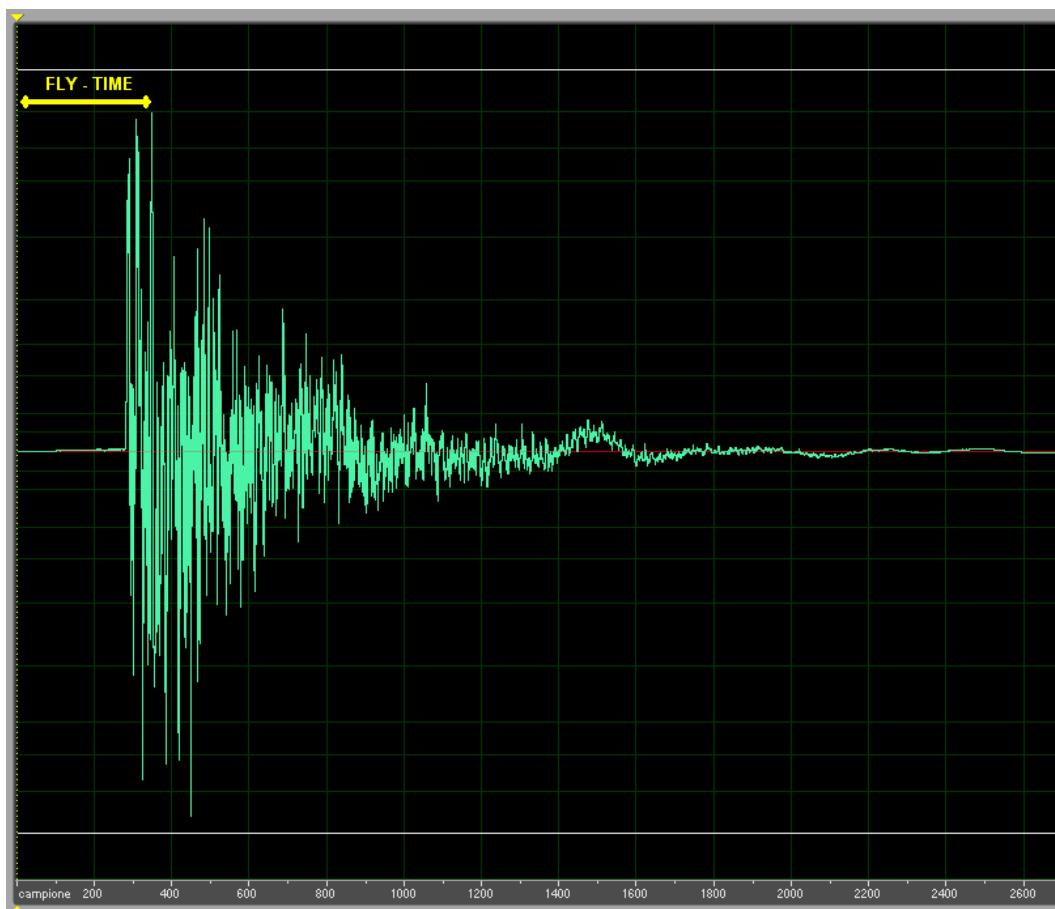


Figura 2.16: Esempio di calcolo del tempo di volo.

Osservando varie misure sperimentali è possibile affermare che quest'approccio non ha dato risultati soddisfacenti.

Il secondo metodo studiato si basa invece su un sovracampionamento della fase. E' possibile realizzare tale operazione effettuando uno zero-padding della risposta impulsiva, calcolando la relativa FFT, estraendone la fase ed infine applicando l'algoritmo DD. Il numero di punti, stimato empiricamente, necessario ad un'esecuzione corretta del DD è di 40'000 per risposte impulsive non superiori a 8192 campioni. Una volta eseguito lo srotolamento è possibile stimare il tempo di volo mediante una retta di regressione lineare e sottrarre alla fase la retta stessa. In tal modo è possibile osservare con più dettaglio l'andamento della fase (figure 2.17 e 2.18).

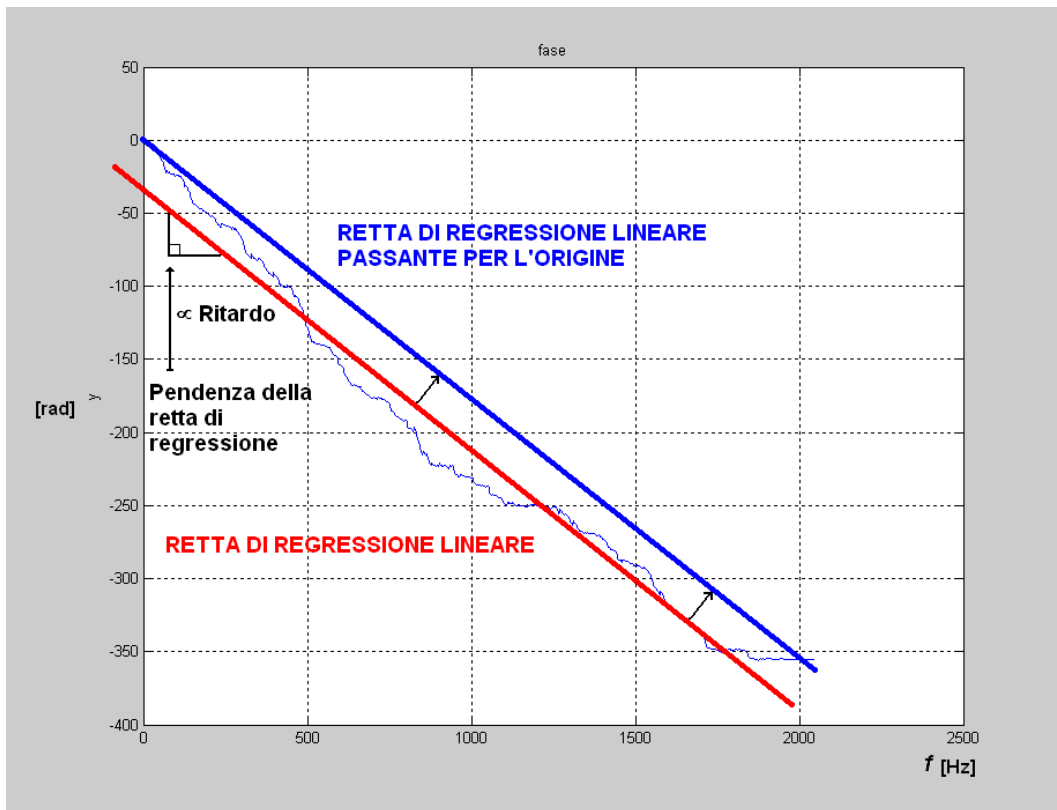


Figura 2.17: Esempio di calcolo della retta di regressione lineare dell'unwrap di fase.

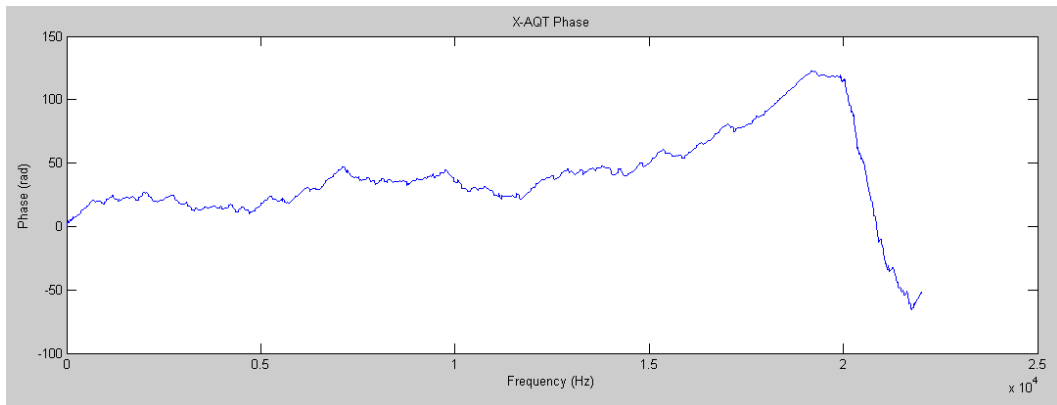


Figura 2.18: Fase srotolata compensata mediante retta di regressione lineare.

CAPITOLO 3

Psicoacustica

Dalle considerazioni sviluppate finora sicuramente è emerso che la particolare distribuzione energetica della risposta acustica in macchina , concentrata tutta nei primi millisecondi del fronte, non può essere studiata da un punto di vista stazionario, ma richiede un'analisi più approfondita, sviluppata a partire da considerazioni psicoacustiche sui fenomeni transitori dell'udito umano.

Il sistema non riesce in alcun modo a raggiungere una condizione di regime come avviene invece in sistemi acustici quali chiese e teatri, in cui la coda sonora rappresenta l'assestamento dell'energia sonora e come tale non può essere studiato da un punto di vista statico. Innanzitutto si introducono alcune nozioni fondamentali necessarie alla comprensione dei fenomeni che saranno descritti.

3.1 Stimolo, sensazione, statistica dei dati

Tutti i fenomeni psicoacustici di cui si è a conoscenza sono stati scoperti mediante l'uso dei test costruiti ad hoc, in cui uno stimolo, costituito da un segnale in pressione, viene inviato al sistema uditivo umano o in etere o direttamente nel canale uditivo, e provoca una sensazione uditiva ben precisa che viene preso come risultato del test. I test sono prettamente soggettivi, in quanto sono realizzati su campioni addestrati all'ascolto, e quindi per essere verificati hanno bisogno di una base statistica molto ampia, in modo da costruire un risultato medio che sia valido in generale per ogni individuo a meno di malfunzionamenti dell'organo uditivo stesso. I test si basano su metodologie ben distinte. Le più importanti sono riportate di seguito:

Metodo di aggiustamento: in questo metodo, il soggetto ha il controllo dello stimolo. Per esempio, il soggetto può variare il livello di un tono puro fino a quando diventa udibile. Tale metodo viene utilizzato per determinare la soglia minima di udibilità di un suono. In altri casi il controllo invece è sulla frequenza.

Metodo del tracciamento: il soggetto ha ancora il controllo dello stimolo. Tuttavia, in contrasto con quanto avveniva nel metodo precedente, il soggetto ha il controllo unicamente della direzione in cui varia lo stimolo. Quello che si ottiene è un tracciamento a zigzag in cui si passa ad esempio da tono udibile a tono non udibile.

Stima della grandezza: in questo metodo agli stimoli sono assegnati dei numeri corrispondenti alla grandezza percepita. Per esempio ad una sequenza di stimoli possono essere assegnati dei numeri che corrispondono alla potenza dello stimolo. In particolare viene utilizzata per determinare il rapporto fra la potenza percepita di due suoni, uno a potenza costante a cui viene assegnato il valore 100, e un altro che lo segue a potenza diversa, di cui si deve valutare il rapporto.

- Procedura sì o no: all'ascoltatore sono presentate unicamente due opzioni. Ad esempio sento o non sento il suono.
- Scelta forzata fra due intervalli : la scelta verte unicamente fra due intervalli di valori prefissati
- Procedure adattative: gli stimoli presentati durante la prova all'ascoltatore dipendono dai risultati ottenuti dai test precedenti. E' una sorta di procedura up-down.

Lo stimolo può essere descritto in termini di pressione sonora, frequenza, durata. Tali grandezze fisiche sono correlate con le sensazioni psicofisiche di forza, pitch, durata soggettiva e le relazioni esistenti vengono ripotate tramite equazioni o grafici. Ad esempio esistono grafici che rappresentano l'andamento della potenza sonora percepita al variare dell'ampiezza dello stimolo. Da notare che non sempre ad una variazione dello stimolo corrisponde a una variazione della sensazione uditiva perché la risoluzione umana è finita ed esiste una "soglia" oltre la quale la percezione non cambia. Un esempio estremo di soglia, è la soglia assoluta di udibilità, ovvero il livello necessario ad un tono puro per essere udito. Non sempre però i risultati hanno un andamento univoco, ma dipendono

fortemente dalle condizioni in cui vengono presentati. Quindi le condizioni di test devono essere rigorosamente tenute sotto controllo.

Alla fine di ogni test effettuato su un certo set di ascoltatori si ha una quantità rilevante di dati. Di essi se ne calcola un valore medio che viene assunto come risultato del test. Da notare che anche il tipo di media da utilizzare influenza molto l'attendibilità finale del risultato, in quanto ognuna tende a dar maggior peso ad alcune zone di valori rispetto ad altre. Ci sono medie lineari e geometriche su valori presentati in db o in lineare. Tale scelta tipicamente non è chiara a priori ma va definendosi nel corso della elaborazione dei dati.

Il primo concetto psicoacustico rilevante che si può definire tramite questa serie di test è, come è già stato accennato, la determinazione della zona in cui avviene la percezione uditiva ovvero il livello minimo che permette ad suono di essere rilevato dal nostro sistema uditivo. Faremo riferimento a tale soglia come "threshold in quiet" ovvero soglia in quiete.

3.2 L'area di udibilità umana

L'area d'ascolto rappresenta un piano in cui vengono rappresentati i suoni che risultano udibili all'orecchio. Essa viene presentata su una scala logaritmica per quanto riguarda le frequenze riportate in ascisse e in db per le ampiezze riportate in ordinata. Inoltre in ascissa è riportato anche il rapporto fra le bande critiche, in quanto tale scala descrive meglio il comportamento del sistema uditivo umano di quanto non faccia la scala classica delle frequenze. Queste curve sono ricavate utilizzando come toni di test toni puri in condizioni di regime di durata superiore ai 100 ms. In figura sono rappresentate la soglia d'udibilità, la soglia di dolore e le aree tipiche della voce e della musica.

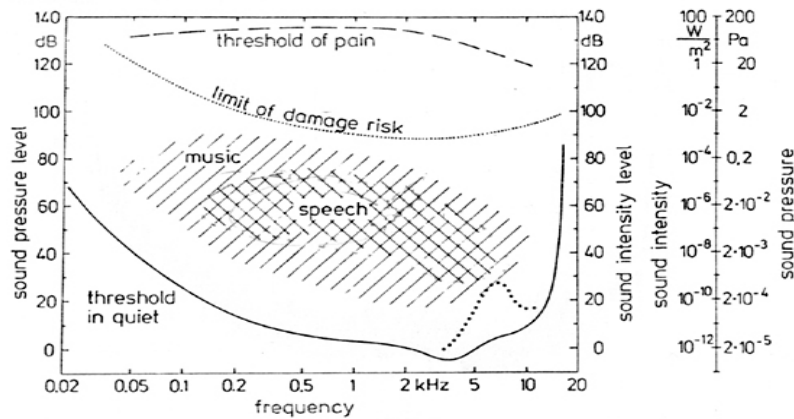


Figura 3.1 : Area di udibilità umana.

Quello che interessa notare è che tali curve sono utili unicamente nei casi in cui il suono venga riprodotto in condizioni stazionarie, ovvero non risultano utili nel caso che si sta sviluppando in questa sede. La cosa interessante che emerge è però che la sensibilità umana non è uniformemente distribuita lungo tutte le frequenze ma esistono aree in cui è maggiormente accentuata. Ad esempio la soglia di udibilità, rappresentata in funzione della frequenza, risulta molto maggiore alle basse frequenze rispetto alle alte. La sua dipendenza dalla frequenza può essere misurata precisamente e velocemente usando il metodo di tracciamento. In tale metodo il soggetto usa uno switch con cui cambia la direzione di variazione della pressione sonora decidendo se incrementarla o decrementarla. Allo stesso tempo la frequenza è fatta variare lentamente dal basso all'alto (o viceversa) in maniera automatica. Quando il tono da udibile diventa non udibile (o viceversa) lo switch viene commutato. Si ottiene in questo modo un andamento continuo a zigzag della soglia di udibilità dell'individuo. L'andamento medio di tale soglia viene riportato in figura 3.2.

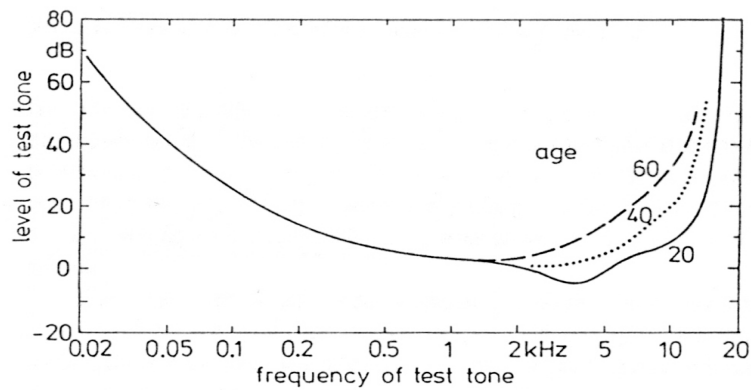


Figura 3.2 : soglia di udibilità.

A bassa frequenza, la soglia richiede un livello di pressione sonora relativamente alto che raggiunge circa 40db a 50Hz. Il livello a 200Hz è già sceso però a 15db. Per frequenze che vanno dai 500Hz ai 2KHz la soglia per il soggetto rimane indipendente dalla frequenza, un effetto che raramente si presenta. Dai 2 ai 5KHz invece si raggiunge una sensibilità molto spiccata che si estende fino agli zero db. Dai 16 ai 18KHz la sensibilità scende rapidamente, e conseguentemente la soglia sale.

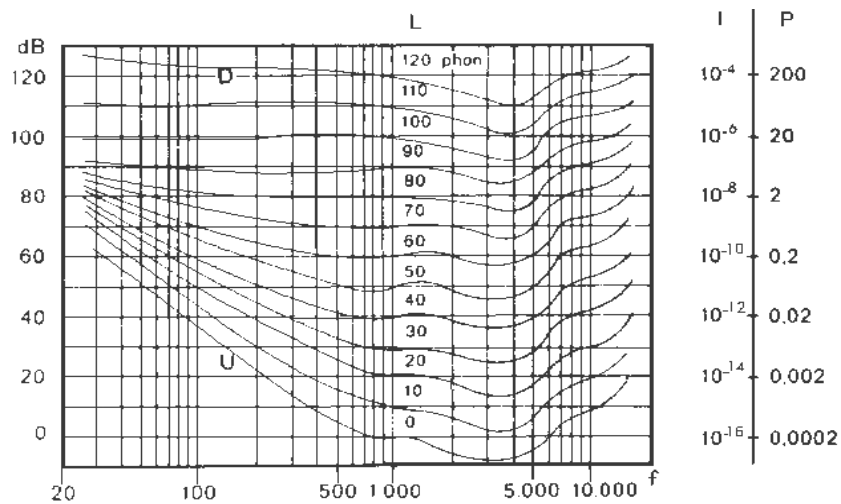


Figura 3.3 : Audigramma di Fletcher e Munson

3.3 Sistema di acquisizione delle informazioni

In questo paragrafo ci si propone di chiarire la genesi di quei fenomeni che saranno utilizzati in seguito e giustificare i risultati che saranno esposti. Per chiarezza e completezza risulta interessante considerare la fisionomia del sistema uditivo umano perché permette di capire la complessità di quei fenomeni su cui si basa la nascita della sensazione uditiva fornita dal cervello umano partendo semplicemente da un segnale in pressione.

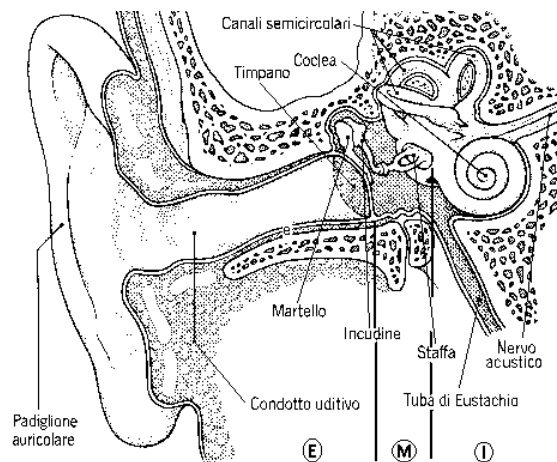


Figura 3.4: Struttura del sistema uditivo umano.

Innanzitutto è fondamentale capire che rispetto ad un campo sonoro ideale, libero e progressivo, vengono introdotte nel caso reale delle distorsioni dovute alla presenza della testa e del corpo umano. In particolar modo per frequenze sotto i 1,5KHz si risente di questi fattori. Esiste una funzione di trasferimento fra il suono presente in campo libero e quello effettivamente percepito che viene indicata con la sigla HRTF, acronimo di Head Related Transfer Function. Nel caso trattato non si tiene in considerazione tale funzione di trasferimento perché l'attenzione è spostata verso i fenomeni veri e propri di decodifica e trasduzione del messaggio sonoro. Il sistema uditivo può essere diviso in tre sezioni distinte:

L'orecchio esterno, costituito dal padiglione auricolare, dal condotto uditivo e dal timpano, che ha il compito di raccogliere i suoni provenienti dall'ambiente, di convogliarli all'interno del condotto uditivo. Al suo interno i suoni vengono amplificati di 12 dB e vengono convogliati verso il timpano.

Orecchio medio che funge da adattatore di impedenze. Infatti trasforma un segnale di piccola forza, ma dispiegato su una grande area, in un segnale con grande forza ma localizzato, evitando in tal modo sprechi di energia.

Orecchio interno, costituito dalla coclea, dai canali semicircolari e dal nervo acustico.

I canali semicircolari in realtà non servono all'udito ma sono la sede del senso dell'equilibrio. La coclea invece è costituita da due condotti paralleli, il canale vestibolare e il canale timpanico, separati da una parete chiamata membrana basale. I due condotti sono riempiti di endolinfa e sono in comunicazione tra loro tramite un'apertura alla fine della membrana basale. La coclea riceve il segnale dalla staffa tramite la finestra ovale, una membrana che separa l'orecchio medio da quello interno, e lo trasmette all'endolinfa nel canale vestibolare. Le variazioni di pressione del fluido nei due canali vengono quindi percepite dalle cellule cigliate presenti sulla membrana basale che trasmettono l'informazione al nervo acustico. La membrana basale funge da filtro molto selettivo per quanto riguarda le frequenze percepibili: essendo più tesa e fine vicino alla finestra ovale da cui entra il segnale e via via sempre più spessa mano a mano che ci si allontana da essa, ha una risposta differente al variare delle frequenze. Per la sua conformazione permette dunque di percepire prima le componenti del suono ad alta frequenza e solo più tardi quelle a bassa frequenza. Grazie a questa particolare fenomeno è possibile distinguere all'ascolto toni riprodotti contemporaneamente a frequenze diverse. L'idea che le basse frequenze producano oscillazioni della membrana basale vicino all'elicotrema e le alte frequenze vicino alla finestra ovale, è stata confermata sperimentalmente. Un esempio di come vengono distinte tre frequenze campione è riportato in figura. Si nota che esiste un andamento tipo: l'ampiezza cresce gradualmente dalla finestra ovale verso l'elicotrema, raggiunge un massimo, e diminuisce rapidamente sotto tale massimo.

Si nota come la separazione delle diverse frequenze dello stimolo è molto netta in corrispondenza dei massimi. Ogni tono fa vibrare una regione distinta della membrana basale. Si può schematizzare allora la percezione umana con una serie di filtri passabanda selettivi di determinata forma centrati su frequenze

particolari, che corrispondono alla divisione naturale dei toni su diverse zone della membrana basale. Tale scala viene detta scala di Bark.

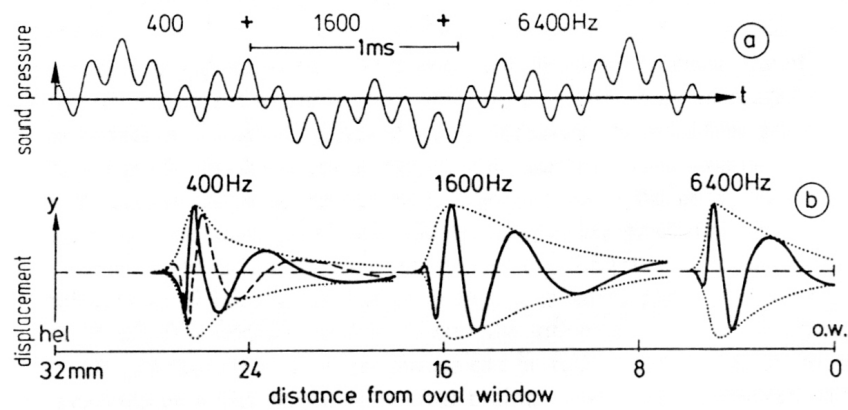


Figura 3.5 : distinzione all'interno della membrana basale di tre toni sovrapposti.

I filtri coprono l'intero spettro audio sovrapponendosi tra loro. Quindi un tono puro tipicamente non solo fa vibrare la zona corrispondente al proprio filtro ma anche le zone vicino. Tale effetto però, di norma non viene tenuto in considerazione e si ipotizza una suddivisione netta fra i filtri. Tutte le informazioni estratte dal segnale dalle cellule cigliate vengono allora elaborate da una rete neurale che ritrasmette al cervello la distribuzione energetica del suono alle varie frequenze. A causa però della limitata lunghezza del canale di trasmissione, in presenza di troppe informazioni i suoni meno intensi con frequenze nelle vicinanze di altri suoni con intensità molto maggiori vengono trascurati e non percepiti: si ha perciò il fenomeno detto del mascheramento.

Del mascheramento e della scala di Bark si parlerà in un paragrafo a se stante.

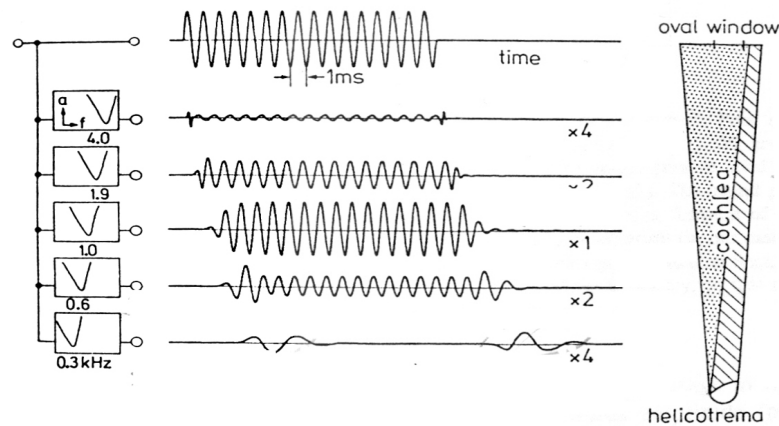


Figura 3.6: schematizzazione del sistema uditivo attraverso un banco di filtri.

Un altro aspetto rilevante dell'udito umano è la sua capacità di adattare la sua sensibilità al livello di stimolazione esterna. Esiste una specie di retroazione infatti che, nel caso di segnali molto deboli, provoca una maggiore sensibilità del sistema uditivo e permette di percepire segnali anche di bassissimo livello. Questo permette di capire quanto il sistema uditivo umano sia complesso. La retroazione è portata talmente al limite che esistono fenomeni di auto-oscillazione del sistema che danno vita a emissioni sonore anche in assenza di stimoli esterni. Tale comportamento caratteristico del sentire umano viene da origine alle emissioni otoacustiche. Si tratta di suoni prodotti all'interno del sistema uditivo che vengono percepite come oscillazioni in aria. La cosa importante da dire è che è possibile trovare una relazione fra i tipi di emissione otoacustica e l'andamento in funzione della frequenza della soglia d'udibilità. Questo fa pensare che entrambe i comportamenti siano due aspetti distinti dello stesso fenomeno. Questo non fa altro che avvalorare tutta la teoria sviluppata sui fenomeni psicoacustici. Ad esempio in corrispondenza dei minimi delle emissioni spontanee cadono i massimi della soglia di udibilità. Tali punti corrispondono anche alle frequenze per cui si ha emissione spontanea. Le emissioni otoacustiche assumono dunque un ruolo fondamentale perché tramite il loro studio è possibile ricavare un modello dell'udito umano e ricavare la forma tipica dei filtri di cui si è parlato in precedenza. Inoltre tramite le emissioni spontanee è possibile centrare questi filtri in corrispondenza di frequenze tipiche che corrispondono alle zone di massima sensibilità.

Per gli scopi preposti in questa sede è sufficiente un'analisi condotta ad un livello meno approfondito. Ma tre concetti devono essere tenuti in considerazione.

Mascheramento

Bande critiche

Andamento della scala di Bark

3.4 Il mascheramento acustico

Il mascheramento acustico [7] è un effetto che ha un'influenza enorme sulla sensazione uditiva umana, al punto da trovare una delle sue applicazioni più importanti nel sistema di compressione audio MPEG, il cui successo è dovuto alla elevata qualità della codifica che rende i file compressi spesso indistinguibili da quelli originali.

Con il termine mascheramento si intende la capacità di un tono o di un rumore a banda stretta (detti di 'masking') di rendere un altro tono di 'test' impercettibile all'orecchio umano sotto determinate condizioni. Tipicamente la soglia di mascheramento di un tono si eleva molto al di sopra della soglia di udibilità; le due curve sono invece identiche quando la frequenza del tono di mascheramento e di test sono molto lontane fra loro. Questo è spiegabile con il fatto che il fenomeno del mascheramento avviene unicamente per frequenze che vanno a stimolare zone della membrana basale che sono spazialmente vicine tra loro. E' allora possibile costruire una maschera di filtraggio psicoacustico che schematizza il fenomeno. Vengono percepiti unicamente quei toni che cadono al di sopra di tale maschera.

Oltre al mascheramento totale che rende un tono udibile o non udibile esiste anche un fenomeno di mascheramento parziale in cui il tono di masking affievolisce l'effetto provocato dal tono di test.

Da quanto detto emerge che il mascheramento deve analizzato sotto due aspetti distinti:

- Da un punto di vista temporale per capire per quali successioni di toni si ha l'effetto.

- Dal punto di vista della frequenza per capire per quali zone dello spettro e in che misura un tono di masking di frequenza prefissata ha influenza sul tono di test.

Una successiva classificazione può essere effettuata per entrambe i campi. Dal punto di vista temporale il mascheramento viene distinto in:

- pre-mascheramento: si ha quando il tono di test viene ‘oscurato’ da toni di masking che vengono generati in istanti seguenti l’inizio del tono di test; questo tipo di mascheramento non ha particolari effetti sull’ascolto e tiene unicamente conto del fatto che è necessario un piccolo lasso di tempo prima che l’orecchio possa reagire allo stimolo di pressione corrispondente al tono di test, dando quindi il tempo a successivi stimoli di oscurarlo.
- mascheramento contemporaneo: si ha quando il tono di test viene generato durante la riproduzione dei toni di masking ed è da esso oscurato;
- post-mascheramento : si ha quando il tono di test viene generato in istanti successivi alla fine dei toni di masking.

Dal punto di vista della frequenza il mascheramento viene distinto in mascheramento di un tono dovuto:

- ad un altro tono
- a rumore a banda stretta

Una breve digressione viene sviluppata su tali tipi di mascheramento, soffermandosi maggiormente sui casi di interesse alla trattazione.

3.5 Mascheramento di toni puri con rumore

Svariati tipi di rumore sono comunemente usati in psicoacustica:

- Il classico rumore bianco, che è un rumore a banda larga, con spettro piatto su tutta la gamma di frequenza audio. La densità spettrale del rumore bianco è indipendente dalla frequenza.

- Rumore rosa, che ha uno spettro di potenza piatto su una scala frequenziale a terzi di ottava. Questo provoca una maggiore sollecitazione del sistema alle basse frequenze e risulta di fondamentale importanza nella determinazione della risposta di un ambiente in quanto presenta caratteristiche prossime alla risoluzione in frequenza dell'orecchio.
- Rumore di mascheramento uniforme: fortemente dipendente dalla frequenza.

Nel caso del rumore bianco la curva di mascheramento che si ottiene andando a testare diversi toni è riportata in figura 3.7.

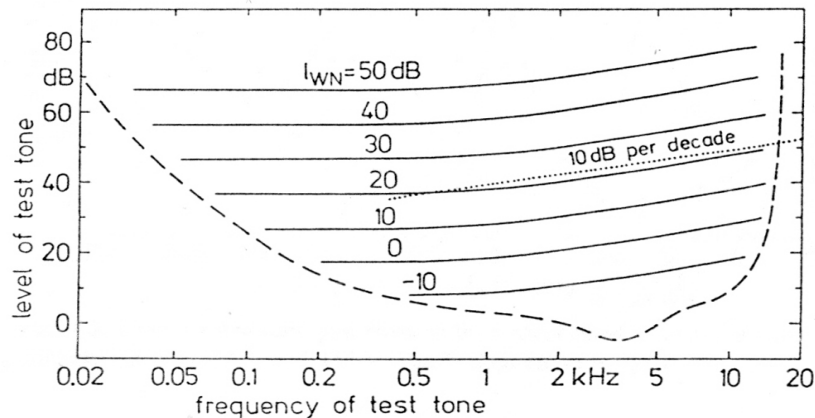


Figura 3.7: soglia di mascheramento provocata da rumore bianco.

Nonostante il rumore bianco abbia una densità spettrale indipendente dalla frequenza, la soglia di mascheramento risulta piatta unicamente in bassa frequenza. Oltre i 500 Hz la soglia cresce con il crescere della frequenza con una pendenza di 10 dB per decade. Il comportamento in ampiezza della soglia è però lineare in quanto, a una crescita di 10 dB del livello del rumore di masking, corrisponde una crescita di pari intensità della soglia di mascheramento. Da tali considerazioni risulta semplice ottenere un andamento della curva di masking piatta. Infatti basta pre-distorcere il livello del rumore di mascheramento in modo che abbia un andamento inverso a quello della curva presentata in figura 3.7. Molto più interessante risulta il mascheramento che si ottiene tramite un rumore a banda stretta.

3.6 Mascheramento dovuto a rumore a banda stretta

Per rumore a banda stretta si intende un rumore che si estende su una banda di larghezza pari o inferiore a quella delle bande critiche. Quindi si estenderà per circa 100Hz se composto da rumore di frequenze inferiore ai 500 Hz e pari a $0.2f$ per quelle superiori. In figura 3.8 è mostrata la soglia di mascheramento di toni puri centrati a 250 Hz, 1 KHz e 4 KHz mascherati da rumore a banda stretta. Il livello di ogni rumore di mascheramento è 60 dB mentre le loro larghezze di banda sono rispettivamente 100,160 e 700 Hz. Come si nota l'andamento tipico è quello di filtri passabanda con pendenze elevatissime (dell'ordine di 200 dB/ottava).

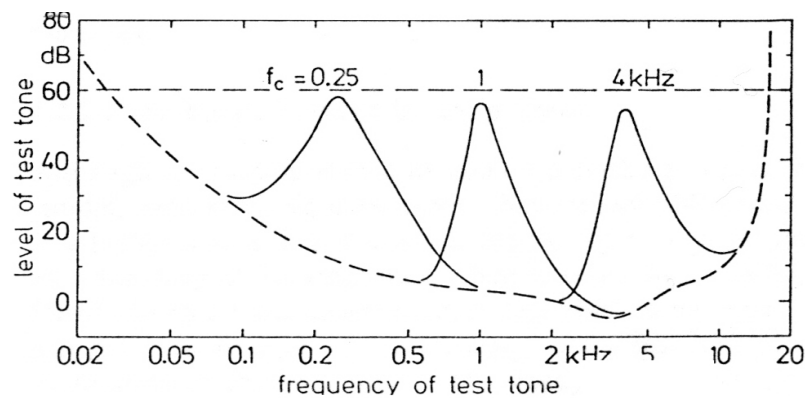


Figura 3.8: soglia di mascheramento provocata da rumore a banda stretta per diverse frequenze.

In figura 3.9 invece è rappresentata la dipendenza della soglia di mascheramento dal livello di un rumore centrato a 1KHz. Tutte le curve, a qualsiasi livello mostrano una elevata pendenza del tratto di curva che conduce verso il massimo. Dopo il massimo invece per livelli elevati di SPL la discesa risulta essere molto più dolce. Questo è un comportamento fortemente non lineare del sistema uditivo.

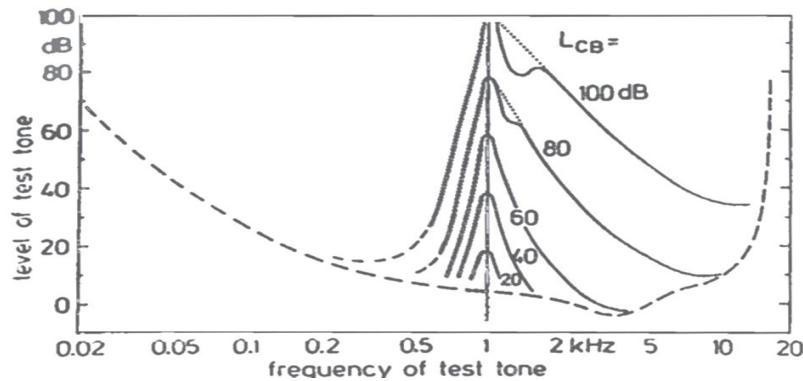


Figura 3.9 : soglia di mascheramento provocata da rumore a banda stretta per diversi livello di rumore

3.7 Mascheramento di toni puro da parte di toni puri

In figura 3.10 viene riportata la soglia del tono di test come funzione della sua frequenza in presenza di un tono di frequenza pari a 1 KHz a livello pari a 80 dB. Oltre agli effetti visti nella sezione precedente si ha in questo caso un effetto aggiuntivo. Si nota infatti in figura la presenza di un beat in corrispondenza degli intorni frequenziali in cui cade il tono di mascheramento. Per evitare la sensazione di beat viene utilizzato in quelle regioni un tono di test che si trova sfasato di 90° rispetto a quello di masking. Con questo artefatto si ottiene il risultato presentato in figura 3.11.

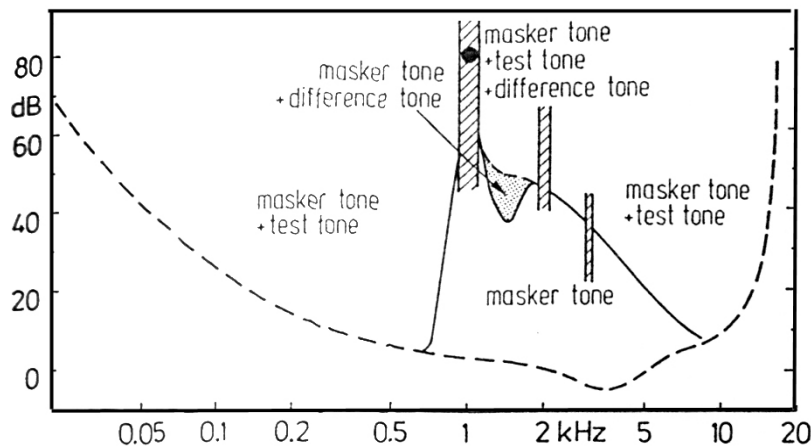


Figura 3.10: mascheramento provocato da toni puri in fase con il tono di test.

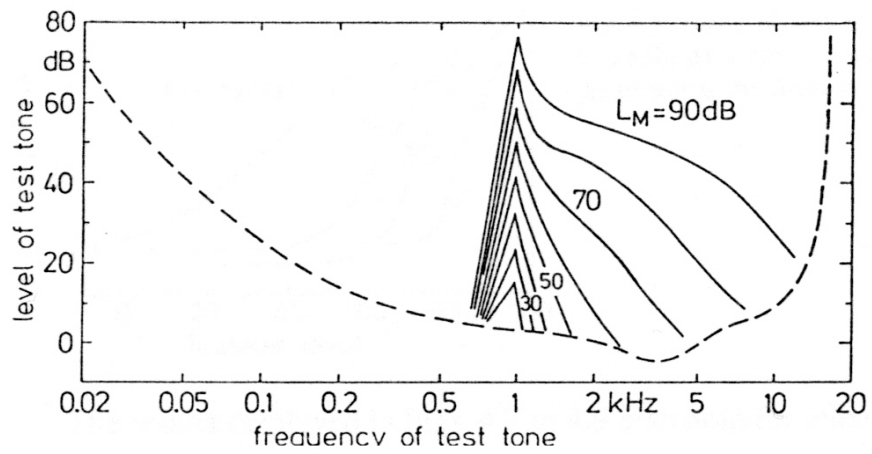


Figura 3.11: mascheramento provocato da toni puri in controfase rispetto al tono di test.

3.8 Mascheramento dovuto a una serie di toni

Le prove sperimentali eseguite in questo ambito sono essenziali per determinare il mascheramento prodotto da strumenti, quali la tromba, il cui suono è costituito non solo dall'armonica fondamentale, ma da una serie di armoniche multiple della stessa. Il mascheramento prodotto da una serie di armoniche è presentato in figura 3.12. Si nota come la risoluzione richiesta per descrivere la soglia di masking in bassa frequenza sia maggiore rispetto alle alte, in quanto alle alte risulta essere all'incirca costante.

Sono stati fatti diversi esperimenti al fine di determinare, con l'utilizzo di una serie di toni, la stessa soglia di mascheramento prodotta da un rumore a banda stretta esteso all'interno di una banda critica. La sovrapposizione delle due curve si è ottenuto generando in maniera randomica all'interno della banda critica di interesse, 5 toni puri. Questo permette di affermare che, nella costruzione della curva di mascheramento all'interno di una banda critica, è sempre possibile determinare un rumore che provochi lo stesso effetto di mascheramento di una serie di toni. Tale proprietà viene sfruttata nella codifica MP3 dove, durante la valutazione della soglia di mascheramento, i toni vengono valutati in base al loro rumore equivalente. Si capisce quanto sia rilevante la divisione dello spettro nelle bande critiche corrispondenti, in quanto esse rappresentano, in maniera molto accurata, la distinzione armonica praticata dal sistema uditivo umano.

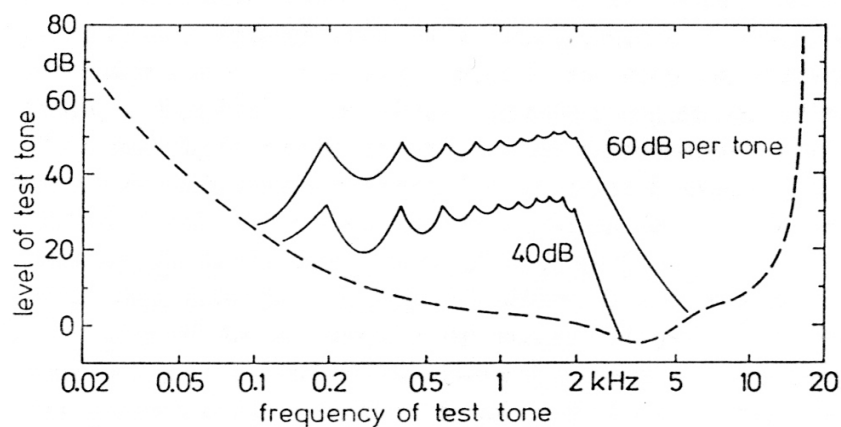


Figura 3.12: soglia di mascheramento prodotta da una serie di toni.

3.9 Effetti temporali

Si è posta finora in evidenza l'importanza fondamentale che l'analisi svolta in bande critiche assume nella determinazione della soglia di mascheramento e questo sarà sviluppato più in dettaglio nel paragrafo successivo.

Si rende ora necessario andare oltre. I risultati ottenuti infatti descrivono il mascheramento da un punto di vista prettamente statico, in quanto i segnali di test e mascheramento utilizzati negli esperimenti sono costituiti da segnali duraturi nel tempo che permettono al sistema uditivo di portarsi a regime. Tali condizioni però non rispecchiano la situazione reale d'ascolto umano, in quanto i segnali che tipicamente giungono all'udito sono segnali, quali la voce e la musica, fortemente non stazionari. Nasce l'esigenza di introdurre informazioni aggiuntive che tengano in considerazione il comportamento della sensazione uditiva nella situazione reale. In particolar modo assumono un ruolo fondamentale l'attacco e lo stacco di un tono in quanto ne determinano la sensazione di dinamica associata e il livello sonoro effettivamente percepito dall'udito umano. Questo non corrisponde al valore di regime, ma al picco che cade nei millisecondi seguenti l'inizio del fronte, ovvero nel lasso di tempo in cui per effetto Haas l'orecchio integra l'informazione sonora.

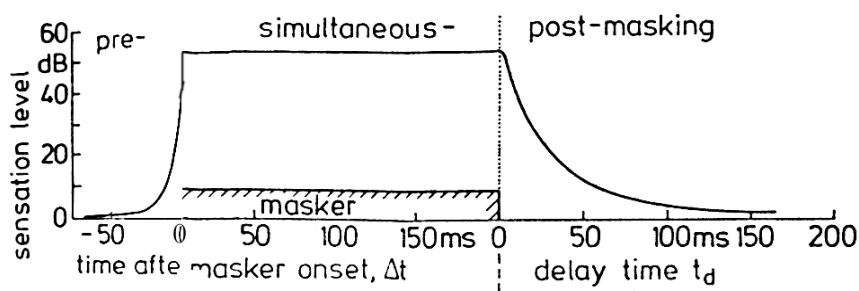


Figura 3.13: andamento temporale del mascheramento.

Tipicamente in natura suoni forti si susseguono a suoni deboli dando vita a fenomeni di mascheramento. Un esempio di tale comportamento si ha considerando la voce in cui i suoni forti sono rappresentati dalle vocali, mentre le consonanti rappresentano i suoni deboli. Spesso una consonante attenuata è

mascherata da una forte vocale pronunciata in precedenza. Per misurare tali effetti quantitativamente si utilizza un tono di masking di durata limitata e una serie di toni di test costituiti da burst. Il tono di test viene generato nell'intervallo di tempo che comprende la generazione del tono di masking in istanti che interessano i millisecondi immediatamente precedenti lo stesso, il periodo di riproduzione e gli istanti successivi ottenendo in tal modo una valutazione relativa dei suoi effetti nel tempo. In figura 3.13 è presentato un tono di mascheramento di 200 ms e la conseguente curva di masking che si ottiene. La distinzione che viene effettuata fra i vari tipi di masking, come già detto all'inizio, si riallaccia alla relazione temporale esistente fra i due toni. Solo due di essi però sono rilevanti dal punto di vista pratico ovvero il mascheramento simultaneo e il post-mascheramento.

3.10 Mascheramento simultaneo

La soglia di udibilità e la soglia di mascheramento dipendono dalla durata del suono di test e in particolar modo due dipendenze possono essere differenziate: una è la dipendenza della soglia dalla durata di un singolo tono di test, l'altra è la dipendenza da una ripetizione di burst.

In figura 3.14 è riportata la dipendenza per un tono di test costituito da un burst sinusoidale. La soglia di udibilità è presentata con una linea tratteggiata mentre la soglia mascheramento del tono di test dovuta a un rumore uniforme di livello pari a 60 dB e 40 dB è presentata in linea continua

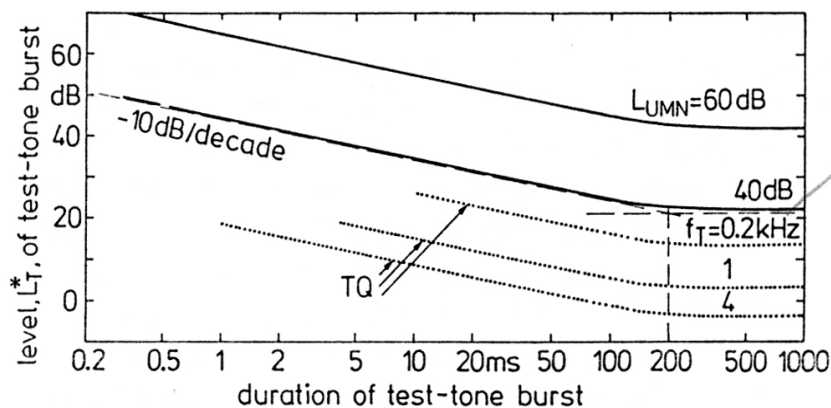


Figura 3.14: mascheramento simultaneo dovuto a un tono di test.

La dipendenza dalla durata del tono di test mostra che per durate maggiori di 200 ms la soglia assume un andamento costante, ovvero oltre i 200 ms il sistema uditivo si trova praticamente in condizioni di regime. Per durate inferiori invece si ha uno slope delle due curve di 10 dB per decade. Da tali considerazioni si capisce che il sistema uditivo integra l'intensità sonora per un intervallo di 200 ms e quindi la soglia di mascheramento risulta essere indipendente da ripetizioni periodiche di burst di periodo superiore ai 200 ms.

3.11 Post mascheramento

Tenendo presente gli scopi prefissati in questa sede, risulta chiaro che l'interesse verte verso il post-masking perché permette di stabilire in che termini il sistema uditivo è in grado di percepire gli effetti causati dall'ambiente, in particolar modo dalle code associate alle riflessioni precoci.

Abbiamo detto che un fronte d'onda nel momento in cui giunge al sistema uditivo comincia ad essere integrato per un periodo pari a 25 ms. Per ripristinare la sensibilità piena necessaria a percepire l'arrivo di un nuovo fronte è necessario aspettare un certo lasso di tempo che è dipendente dalla durata stessa dell'evento sonoro e dal contenuto armonico del fronte iniziale.

E' stata studiata da Zwicker [7] una curva che descrive l'andamento del post-mascheramento in funzione della durata del tono di masking. Qui è riportata in figura 3.15. Toni di test che cadono al di sotto di tale curva non vengono percepiti ma vengono oscurati da un messaggio sonoro precedente.

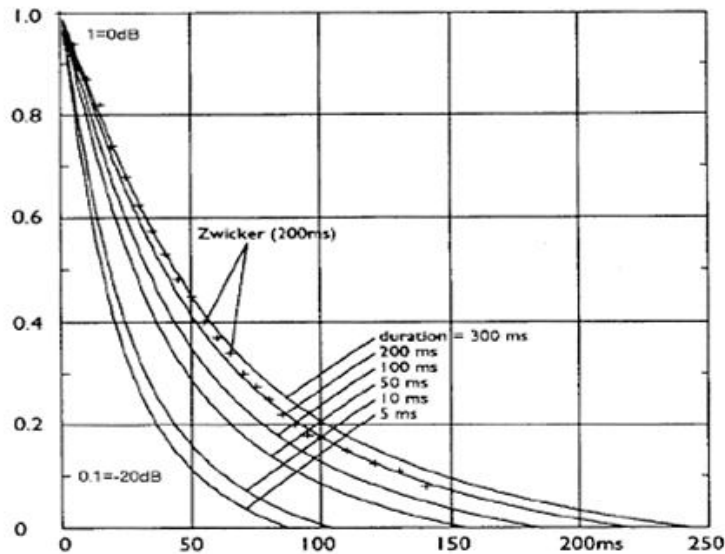


Figura 3.15: Curva di Masking di Zwicker.

Pensando alle riflessioni che arrivano all'orecchio dopo la fine del fronte diretto come a toni di 'test' che seguono il tono di masking, si capisce come queste vengano percepite e quindi influenzino l'ascolto solo se il loro livello si trova al di sopra della curva di mascheramento. Se si trovano appena al di sopra di tale curva contribuiscono a dare senso di spazialità al suono. Se invece il loro livello è molto forte (circa 10 dB sopra il livello del suono diretto) vengono percepite come un'eco distinta e quindi provocano effetti indesiderati.

Il fenomeno del post-mascheramento è fortemente legato alla frequenza del tono di masking. Infatti si ha mascheramento solamente se la frequenza del tono di test è vicina a quella dell'evento che lo ha preceduto. Questo è stato ampiamente detto nei paragrafi precedenti riguardanti la dipendenza del masking dalla frequenza. In particolar modo si è messo in evidenza come sia possibile dividere lo spettro in aree di frequenza chiamate 'bande critiche' che costituiscono una divisione rispondente alla classificazione delle frequenze effettuata dal sistema uditivo umano.

3.12 Bande critiche

Il concetto di banda critica fu proposto per la prima volta da Fletcher. Egli assunse, considerando un tono puro mascherato da un rumore a banda stretta, che la parte di rumore che effettivamente contribuisce al mascheramento è la parte che spettralmente cade vicino al tono. Al fine di ricavare una soglia assoluta da considerare come soglia di mascheramento assunse che il mascheramento è raggiunto quando la potenza del tono e la potenza prodotta dalla componente di rumore che cade vicino ad esso e che provoca il mascheramento sono le stesse. Allora l'estensione spettrale delle bande caratteristiche definite in questo modo è quella che permette al tono e al rumore, una volta che si abbia mascheramento, di avere la stessa potenza. Questo come già si è avuto modo di dire è dovuto al fatto che gli effetti provocati da un rumore a banda stretta e da un tono sono i medesimi a causa della tendenza dell'orecchio a processare il suono in bande di una determinata larghezza. Questo fenomeno viene sfruttato dal formato MP3 in quanto componenti tonali che non possono essere percepite dal sistema umano vengono trasformate in rumore di quantizzazione che necessita di un numero di bit molto inferiore per essere descritto rispetto alla componente tonale.

L'analisi svolta per bande critiche risulta un mezzo potente per la valutazione della qualità acustica di un ambiente. Infatti è noto che analizzare e correggere la risposta armonica di un ambiente frequenza per frequenza non è un buon modo di procedere e spesso non apporta risultati soddisfacenti. Una analisi e successiva correzione invece effettuata per bande critiche può costituire un'alternativa molto efficace da tenere ben presente. La divisione dello spettro in bande critiche viene anche a coincidere con l'andamento della scala di Bark, che rappresenta l'andamento della risoluzione umano nel campo delle frequenze.

Come evidenziato in figura 3.7, il rumore bianco produce soglie di mascheramento che oltre i 500 Hz sono dipendenti dalla frequenza, sebbene il rumore bianco non lo sia, e dopo tale frequenza crescono con una pendenza di circa 10 dB per decade. Tali considerazioni permettono di determinare approssimativamente la larghezza delle bande critiche.

3.13 Come determinare le bande critiche

La pronunciata selettività del sistema uditivo umano è stata messa in evidenza in questo capitolo parlando del mascheramento. Come detto, si può assumere che il sistema uditivo umano processi i suoni in bande di frequenza relativamente strette. Considerando la curva presentata in figura 3.7, si capisce come tali bande risultino avere una estensione spettrale costante sotto i 500 Hz mentre al di sopra di tale frequenza la larghezza di banda delle stesse cresce di un fattore 10 ogni qual volta la frequenza cresce dello stesso fattore 10. Assumendo, come fece Fletcher, che un tono immerso nel rumore sia udibile all'interno della banda critica dove il rumore è centrato, quando la potenza acustica del tono eguaglia quella del rumore, risulta che la larghezza di banda in questione può essere stimata come segue.

Dalla figura 3.7 si nota che per frequenze inferiori ai 500 Hz, la soglia di mascheramento è 17 dB superiore alla densità di livello del rumore bianco che sta mascherando il tono. Ipotizzando che la frequenza del tono sia ininfluente sulla soglia di mascheramento e che, all'interno della banda in questione, in corrispondenza della soglia la potenza acustica del rumore eguagli quella del tono, si può pensare di stimare la larghezza di banda come $10^{17/10}$, valore che risulta essere all'incirca pari a 50 volte un 1 Hz. Allora la banda in bassa frequenza risulta essere 50 Hz. In realtà però l'ipotesi che il sistema uditivo umano produca una soglia di mascheramento la cui potenza è indipendente dalla frequenza del tono è scorretta. Infatti la potenza del tono al livello della soglia di mascheramento è solo circa la metà di quella del rumore che cade nella banda in questione. Usando questa informazione addizionale, la larghezza della banda in questione può essere stimata con precisione. A basse frequenze, le bande critiche mostrano una larghezza di banda di circa 100 Hz, mentre a frequenze superiori ai 500 Hz la larghezza di banda è all'incirca il 20% della frequenza di centro banda. Esistono anche metodi pratici per la determinazione di tali bande e forniscono gli stessi risultati ottenuti tramite queste considerazioni teoriche. I valori delle larghezze di banda sono riportati in tabella 3.1.

Il concetto delle bande critiche è importante per descrivere la sensazione uditiva in quanto come detto questa divisione è basata sul fatto che il sistema uditivo umano analizza uno spettro a banda larga suddividendolo in parti corrispondenti alle bande critiche. Ogni banda critica è centrata in un punto corrispondente alla cosiddetta scala di Bark.

z	f1,fu	fc	z	Delta fg	z	f1,fu	fc	z	Delta fg
Bark	Hz	Hz	Bark	Hz	Bark	Hz	Hz	Bark	Hz
0	0				12	1720			
		50	0,5	100			1850	12,5	280
1	100				13	2000			
		150	1,5	100			2150	13,5	320
2	200				14	2320			
		250	2,5	100			2500	14,5	380
3	300				15	2700			
		350	3,5	100			2900	15,5	450
4	400				16	3150			
		450	4,5	110			3400	16,5	550
5	510				17	3700			
		570	5,5	120			400	17,5	700
6	630				18	4400			
		700	6,5	140			4800	18,5	900
7	770				19	5300			
		840	7,5	150			5800	19,5	1100
8	920				20	6400			
		1000	8,5	160			7000	20,5	1300
9	1080				21	7700			
		1170	9,5	190			8500	21,5	1800
10	1270				22	9500			
		1370	10,5	210			10500	22,5	2500
11	1480				23	12000			
		1600	11,5	240			13500	23,5	3500
12	1720				24	15500			
		1850	12,5	280					

Tabella 3.1: larghezza delle bande critiche

3.14 La scala di Bark

Tale scala è molto importante nella descrizione del sistema uditivo umano. E' costituita da ventiquattro punti che coprono l'area di frequenze che va dai 20 Hz ai 16 KHz. Ogni punto della scala di Bark tipicamente corrisponde ai punti in cui la stimolazione della membrana basale raggiunge il suo massimo. Tale scala è essenziale anche per descrivere la percezione delle minime variazioni in frequenza che il sistema uditivo riesce a cogliere. E' riportata in figura 3.16 la curva che rappresentano l'andamento della scala di Bark in funzione della frequenza. Da notare che la risoluzione dell'orecchio umano verso le basse frequenze è molto più fine rispetto alle alte in quanto in tale zona dello spettro si riescono meglio a distinguere come distinte frequenze molto ravvicinate fra loro. E' essenziale dunque valutare in maniera preponderante l'andamento armonico di un sistema verso le basse frequenze e pesare maggiormente gli indici che danno una valutazione della risposta in tale area dello spettro.

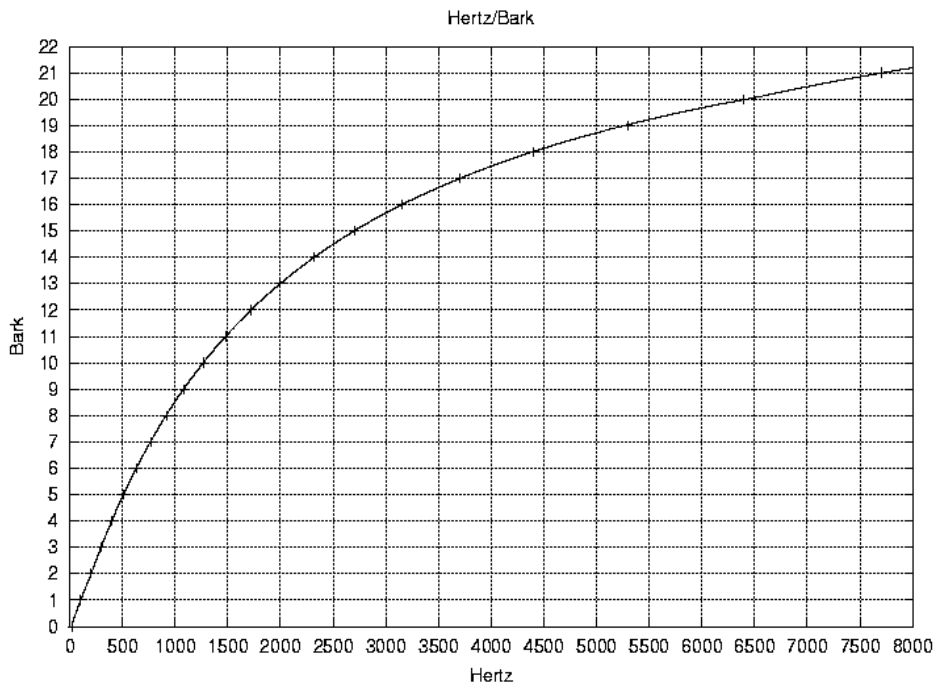


Figura 3.16: Andamento della scala di Bark.

3.15 Conclusioni

Il metodo AQT rivisto alla luce delle considerazioni appena svolte costituisce un mezzo molto potente. Si è appurato che ai fini di una valutazione acustica che restituisca un indice di gradevolezza l'analisi svolta a bande critiche non risulta necessaria, ma è uno strumento utile nel momento in cui si debba intervenire sulla correzione acustica dell'ambiente.

CAPITOLO 4

Analisi AQT

Nel capitolo precedente si sono chiariti i fenomeni che determinano la sensazione uditiva finale nel caso di segnali non stazionari. Rimane da stabilire ora in che modo raggruppare tali nozioni per riuscire ad ottenere una valutazione dell'ascolto all'interno delle autovetture. E' necessario allora avere a disposizione un metodo che fornisca una valutazione degli effetti di propagazione sonora in una forma tale che renda possibile una successiva analisi psicoacustica dei risultati. Quello che fa il metodo AQT è appunto questo: stimolare il sistema con segnali non stazionari, rilevarne la risposta, analizzare l'incidenza sulla risposta dell'ambiente e valutare infine tramite post processing psicoacustico il peso effettivo di tale incidenza sull'ascolto. Vedremo in dettaglio che elaborazioni matematiche sono legate a ciascun blocco del sistema. La rivoluzione essenziale che permette al metodo di sussistere è l'introduzione del nuovo segnale di stimolo AQT. Esso infatti offre la possibilità di rilevare la risposta dinamica dell'ambiente e di valutarla in base alle considerazioni sviluppate nei capitoli precedenti.

4.1 Il segnale AQT

Il segnale AQT, utilizzato per stimolare il sistema, permette come detto di evidenziare la risposta dell'ambiente a segnali transitori. In particolar modo pone in evidenza in che misura l'ambiente trattiene l'energia fornita dagli altoparlanti e in corrispondenza di quali frequenze l'effetto delle code riverberanti è maggiormente evidente. Questo è utilissimo nel momento in cui si vuole

determinare l'influenza delle riflessioni precoci, principali responsabili delle code sonore, sulla valutazione acustica complessiva.

Il segnale AQT non è costituito da un impulso classico ma da una serie di burst formati da toni puri sinusoidali. Questi si susseguono intervallati da un periodo di silenzio che varia in relazione alle dimensioni dell'ambiente in cui avviene la misura. Un esempio di segnale AQT costituito da successione di cinque burst è riportato in figura 4.1.

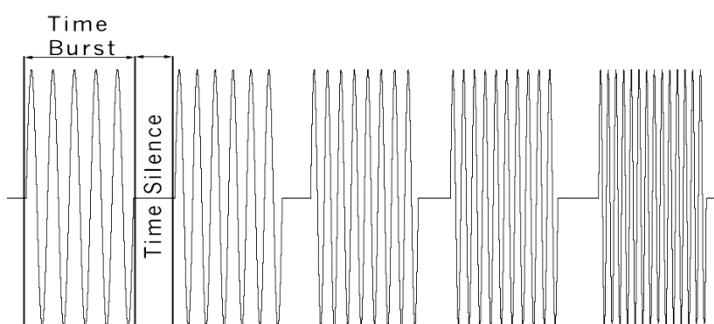


Figura 4.1 Esempio di segnale AQT.

Le impostazioni riguardanti la durata del burst e del successivo periodo di silenzio sono il primo problema da affrontare. Infatti si tratta di determinare quale tipo di stimolo sia più idoneo ad ottenere informazioni attendibili sul comportamento del sistema in risposta a segnali non stazionari. Le considerazioni seguenti chiariscono in che modo è stata effettuata la scelta.

Come detto il primo passo da compiere è procurarsi uno stimolo che permetta di valutare il comportamento del sistema in regime non stazionario. Per questo si utilizza come segnale un burst sinusoidale che evidenzia le reazioni dell'ambiente a regime, in fase di attacco e di stacco. La durata temporale del burst è un parametro che va impostato. La scelta ricade sul valore 200 ms perché, come emerge dal capitolo 3, si può ritenere tale valore come il valore oltre il quale il nostro sistema uditivo ha raggiunto la condizione di regime. Tale valore del burst è lo stesso utilizzato da Zwicker nella determinazione delle sue curve e lo stesso adottato da Liberatore nell'analisi AQT di piccole stanze.

Dopo il burst è presente un periodo di silenzio durante il quale, cessando la stimolazione del sistema, è possibile determinare in che modo si estingue l'energia all'interno dello stesso. Viene quindi messa in evidenza la coda riverberante provocata dalle riflessioni all'interno della stanza. Nel caso di piccoli ambienti questa coda non corrisponde a una vera e propria situazione di regime del sistema in quanto tale condizione in piccole cavità non viene mai raggiunta. E' però possibile da essa estrarre in ogni modo informazioni sulla presenza delle riflessioni precoci. In particolar modo, anticipando quello che verrà detto in seguito, la coda riverberante permette di determinare in quali casi l'andamento temporale della risposta al burst è tale da provocare il mascheramento di successivi messaggi sonori. Intuitivamente questo è possibile applicando alla coda la maschera studiata da Zwicker presentata in figura 3.15. Se la coda ha un decadimento lento e il suo livello sonoro cade al di sopra di tale curva avrà il mascheramento del segnale successivo. Altrimenti questa non inciderà sull'ascolto complessivo in quanto al di sotto della curva la sensibilità del sistema uditivo non è ancora stata ripristinata e quindi la coda non viene percepita. Da notare che nella generazione del treno di burst, se si lascia troppo poco tempo fra un burst e l'altro non si dà il tempo al sistema di estinguere in maniera apprezzabile l'energia e quindi non si ottengono informazioni sulla coda, anzi si rischia di raggiungere la densità di regime. Al contrario, lasciare troppo tempo fra due burst successivi permette al sistema di estinguere completamente la sua energia e quindi non ottengo informazioni utili sul mascheramento che la coda provoca su messaggi sonori successivi, in quanto essa si è estinta completamente.

Da un'analisi comparativa fra diversi risultati ottenuti variando la durata temporale del periodo di silenzio è stato possibile appurare che per spazi ristretti come l'autovettura i valori ottimali da utilizzare come impostazioni del segnale AQT sono:

- durata del Burst: 200 ms
- durata del silenzio: 33 ms

In figura sono riportati una serie di dati a supporto di tale affermazioni. Si nota che utilizzando periodi di silenzio troppo elevato si ottengono risultati che risultano essere poco attendibili.

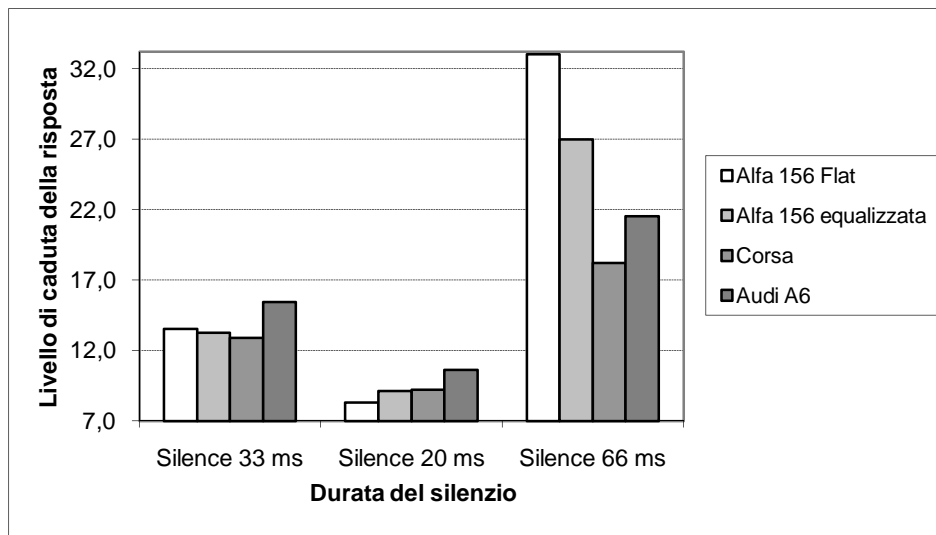


Figura 4.2: Grafico del livello di caduta della risposta alla coda con diversi tempi di attesa fra i burst.

Commento: la scelta è ovviamente da far ricadere sul silente 33 ms. Infatti mantenendo fra un burst e l'altro un periodo di silenzio di 20 ms i valori di caduta dei segnali non sono attendibili. Questo è dovuto al fatto che esiste una discordanza fra i risultati ottenuti in fase di ascolto e risultati riportati in figura 4.2. Ad esempio il recupero che si ottiene nell'abitacolo della Corsa è molto superiore rispetto a quello che si ottiene nel caso dell'Alfa156. Da prove d'ascolto però risulta che la dinamica della risposta è molto migliore in quest'ultima risentendo meno della presenza delle code riverberanti. Non si può dunque tenere in considerazione una durata temporale di 20 ms. Il valore di 66 ms al contrario restituisce risultati troppo poco coerenti per avere un significato attendibile. Questo è dovuto al fatto che la ripresa della coda in corrispondenza dei 66 millisecondi è molto accentuata. Quindi si è scelto il valore 33 millisecondi. In realtà si dovrebbero verificare valori di silenzio intermedi, in quanto probabilmente esiste un valore compreso fra 33 ms e 66 ms (ad esempio 45 ms) che potrebbe dare indicazioni ancor più significative sulla ripresa della risposta. Il

valore di 33 ms è stato scelto anche per motivi di coerenza con la letteratura già esistente che considera tale valore come ottimale [8].

Stabilite le impostazioni temporali da assegnare ai vari burst, è necessario determinare le frequenze a cui generarli. Dalle considerazioni psicoacustiche che sono state fatte nel capitolo 3 è emerso che il sistema umano tende a processare i segnali che arrivano al condotto uditivo dividendo lo spettro in aree frequenziali corrispondenti alle bande critiche. Si è inoltre mostrato che la sensibilità nel campo delle frequenze del sistema uditivo umano è molto più accentuata verso le basse frequenze. Essa è descritta in maniera molto fedele dalla scala di Bark. Basandosi su tali considerazioni si è deciso, nella scelta delle frequenze a cui generare i vari burst, di non coprire uniformemente l'intera gamma dello spettro audio ma di adottare una risoluzione maggiore a bassa frequenza e minore alle alte. La generazione dei burst avviene quindi a frequenze molto ravvicinate nella zona bassa dello spettro, mentre salendo in frequenza è richiesta ai fini dell'analisi AQT una risoluzione molto minore. La scelta adottata da Liberatore è:

Area di frequenze	step
20Hz – 300Hz	2Hz
300Hz – 1000Hz	4Hz
1000Hz – 2000Hz	logaritmica

Tabella 4.1

Nel caso sviluppato in questa sede si è preferito ricorrere a una scala delle frequenze che ripercorra in maniera fedele l'andamento della scala di Bark rappresentato in figura 3.16. Ciò è stato possibile grazie a formulazioni matematiche ricavate empiricamente che ne descrivono l'andamento in maniera esatta. Diverse formule empiriche esistono al riguardo.

Zwicker & Terhardt (1980)	(1) $B = 13 \tan^{-1}(0.76f/1000) + 3.5 \tan^{-1}(f/7500)^2$
	(2) $B = 8.7 + 14.2 \log_{10}(f/1000)$

Terhardt (1979)	(3) $B = 13.3 \tan^{-1}(0.75f/1000)$
	(4) $B = 12.82 \tan^{-1}(0.78f/1000) + 0.17(f/1000)^{1.4}$
Wang, Sekey & Gersho	(5) $B = 6 \sinh^{-1}(f/600)$
Schroeder	(6) $B = 7 \sinh^{-1}(f/650)$
Traunmüller	(7) $B = 26.81/(1+(1960/f)) - 0.53$

Tabella 4.2

L'andamento della scala di Bark approssimato da tali formule viene riportato nella tabella presentata di seguito:

Bark scale (rounded)	Frequency (Hertz)	Approximations						
		Z & Te		Te		W, Se (5)	Sc (6)	Tr (7)
		(1)	(2)	(3)	(4)			
0	0	*0	-	*0	*0	*0	*0	-
1	100	0.987	-5.500	*0.996	1.005	0.995	1.073	0.771
2	200	1.963	-1.225	1.980	*2.002	1.965	2.121	1.952
3	300	2.920	1.275	2.943	*2.978	2.887	3.126	3.029
4	400	3.847	3.049	3.876	3.924	3.751	4.074	*4.014
5	510	4.823	4.548	4.859	4.920	4.627	5.044	*5.006
6	630	5.830	5.851	5.871	5.944	5.498	6.016	*5.991
7	770	6.920	7.088	6.965	7.052	6.409	7.043	*7.032
8	920	*7.985	8.186	8.033	8.131	7.279	8.028	8.034
9	1080	9.006	9.175	9.055	9.164	8.103	8.968	*8.995
10	1270	10.079	10.174	10.122	10.246	8.968	9.960	*10.011
11	1480	11.106	11.118	11.139	11.280	9.809	10.927	*11.005
12	1720	12.109	12.045	12.121	12.289	10.653	11.901	*12.001
13	2000	13.093	12.975	13.071	13.278	11.513	12.897	*13.010
14	2320	14.028	13.890	13.954	14.218	12.371	13.892	*14.003
15	2700	14.943	14.825	14.791	15.137	13.256	14.920	*15.004
16	3150	15.830	15.776	15.566	16.030	14.162	15.973	*15.997
17	3700	16.714	16.768	16.292	16.923	15.113	17.079	*16.996

18	4400	17.628	17.837	16.978	17.856	16.141	18.277	*18.018
19	5300	18.582	*18.985	17.614	18.851	17.249	19.568	19.042
20	6400	19.531	20.148	18.160	19.889	18.375	20.880	*19.994
21	7700	20.452	21.288	18.611	*20.984	19.480	22.168	20.840
22	9500	21.482	22.584	19.037	22.392	20.738	23.635	*21.695
23	12000	*22.587	24.024	19.420	24.285	22.137	25.267	22.516
24	15500	*23.711	25.603	19.750	26.967	23.671	27.056	23.270

Tabella 4.3

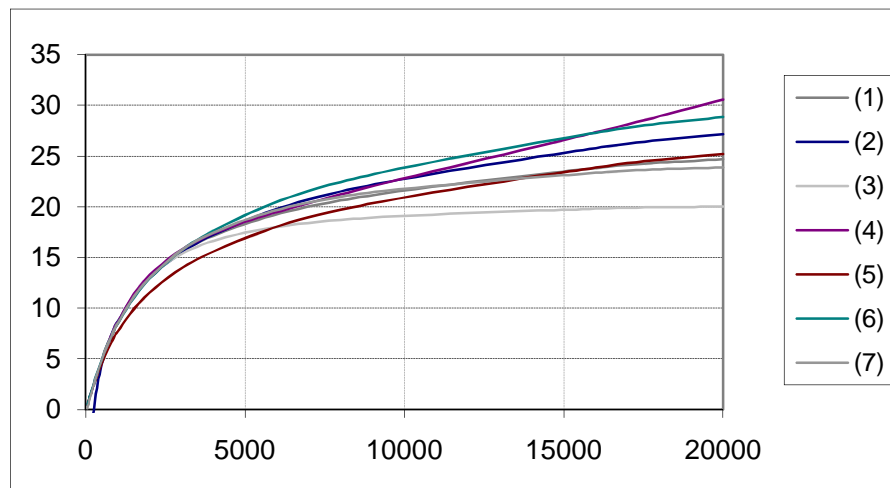


Figura 4.3: andamento delle equazioni che rappresentano la scala di BARK.

Dalla tabella emerge che le formulazioni empiriche che meglio descrivono la scala di Bark sono la formula (3) fino ai 300 Hz e la (7) per le frequenze successive. Per tali formule si è calcolata la formula inversa e si sono presi intervalli di B pari a 0.02 nei primi 300 Hz e 0.05 nei successivi. L'intero spettro in tal modo viene descritto considerando all'incirca trecento punti caratteristici.

L'altra novità di rilievo che è stata introdotta nella generazione dello stimolo è dovuta alla necessità di trattare con segnali a banda stretta che vadano ad agire su una zona limitata dello spettro. Nella versione classica infatti il segnale AQT viene troncato in corrispondenza dei 200 ms assunti come durata temporale del burst. Questo provoca delle terminazioni a gradino che notoriamente hanno uno spettro a larga banda. Vengano in tal modo sollecitate frequenze del sistema che

non dovrebbero essere interessate dallo stimolo. Idealmente infatti l'analisi deve essere condotta per toni puri. Per evitare il problema si è deciso di terminare tutti i burst con uno zero crossing. In particolare le soluzioni adottabili per raggiungere tale scopo sono fondamentalmente due:

- Estendere la lunghezza del Burst al primo zero crossing che si incontra dopo la fine dei 200 ms
- Generare il segnale partendo dalla fine del burst e fare poi all'inizio del burst un fade-in.

La soluzione che si è adottata è la seconda in quanto permette di ottenere risultati sicuramente più accurati soprattutto in bassa frequenza dove aspettare il successivo zero-crossing equivale ad allungare la durata dello stimolo di qualche millisecondo.

In figura 4.4 è riportato il confronto fra due spettri ottenuti rispettivamente da un burst in cui è stato implementato lo zero crossing (presentato in grigio scuro), e da burst troncato in corrispondenza dei 200 ms (presentato in grigio chiaro).

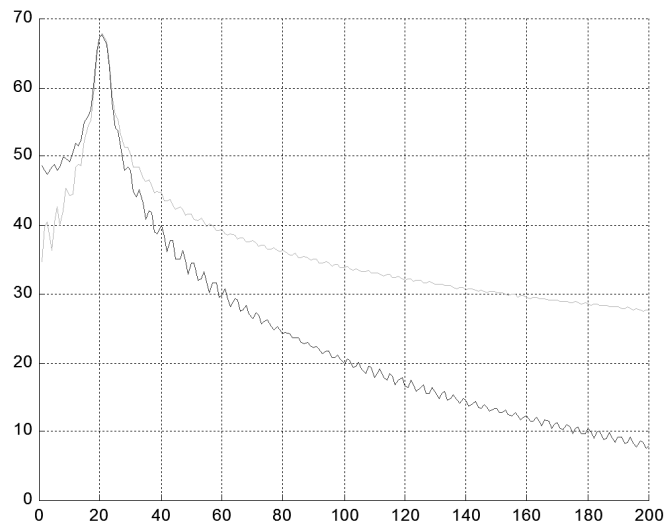


Figura 4.4: confronto fra burst terminati con e senza zero crossing.

4.2 La risposta del sistema allo stimolo

Originariamente, per ottenere la risposta AQT di un sistema acustico, il segnale di prova veniva mandato in riproduzione all'interno dell'ambiente da caratterizzare e la risposta a tale stimolo veniva registrata [9]. Ultimamente è stato sviluppato un nuovo metodo chiamato virtual AQT [8] che fornisce risultati analoghi a quelli ottenuti con il metodo tradizionale attraverso un procedimento molto più semplice. Infatti la curva virtual AQT viene ottenuta dalla convoluzione fra la risposta all'impulso di un sistema e il segnale di test AQT di cui si è parlato. Per maggiori informazioni e per la verifica dell'aderenza dei due metodi si rimanda a [8].

Come si comprende l'introduzione del sistema virtual AQT permette al sistema di calcolo di essere completamente automatizzato in quanto basta la risposta all'impulso dell'ambiente per poter poi ottenere la risposta AQT.

Un esempio di risposta AQT completa è riportato in figura 4.5. E' possibile ottenere indicazioni sulla qualità della risposta AQT semplicemente analizzandola graficamente. Naturalmente in sè e per sè il grafico non rappresenta un metodo di analisi esaustivo in quanto non fornisce informazioni numeriche esatte sugli aspetti della risposta. Per questo poi al grafico bisogna associare una analisi accurata della risposta. Tuttavia risulta comunque importante perché fornisce a livello visivo una prima idea sull'andamento di tutti i parametri che saranno valutati in seguito. Ad esempio l'escursione della curva, rappresentata dalla differenza fra la curva superiore e inferiore, è legata alla ripresa energetica del sistema.



Figura 4.5: esempio di risposta AQT a un treno di burst e AQT completa.

Si nota come in bassa frequenza l'escursione della risposta sia molto piccola. Questo indica che il sistema in tali zone di frequenze trattiene molta energia e non dà il tempo alla risposta di riprendersi prima dell'avvento del successivo burst. Il sistema quindi in bassa frequenza è fortemente affetto da code riverberanti dovute alle riflessioni precoci. In alta frequenza invece il comportamento è sostanzialmente corretto. Da notare che la risposta ideale sarebbe una curva in cui l'escursione è costantemente pari al valore di ripresa ottimo e in cui l'involuppo della stessa sia aderente alla curva target d'ascolto [3]. Visto che il metodo Virtual AQT necessita della risposta all'impulso per funzionare, seguirà una breve digressione sui metodi ottimali di misura della risposta all'impulso.

4.3 Metodo di misura della risposta all'impulso

La risposta impulsiva di un ambiente è molto importante in quanto il suo andamento è indipendente dal tipo di segnale di eccitazione ed è una caratteristica intrinseca del sistema in analisi. Una corretta metodologia di misurazione di tale

funzione risulta pertanto indispensabile per ogni forma di equalizzazione o *processing*.

Un ambiente interno, dal punto di vista acustico, può essere pensato come un sistema il cui ingresso è una sorgente sonora. Tale segnale sarà modificato e l'uscita sarà il suono che ogni ascoltatore (o equivalentemente ogni simulatore) riceverà all'interno della stanza. Grazie alla teoria dei segnali è noto che, dato un sistema, ad un ingresso corrisponde una ed una sola uscita corrispondente alla convoluzione del segnale di input con la funzione $h(t)$, ovvero la risposta all'impulso.

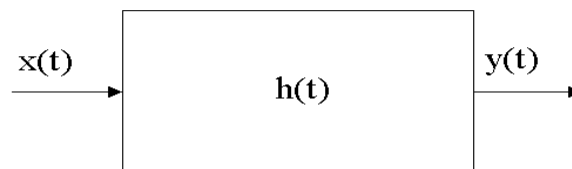


Figura 4.6: schema base di un sistema.

$$y(t) = x(t) \otimes h(t) \quad (4.1)$$

La risposta impulsiva di una stanza, in ambito energetico, è composta da numerosi singoli impulsi, il primo dei quali rappresenta il suono diretto, mentre gli altri rappresentano le riflessioni del suono per mezzo delle pareti

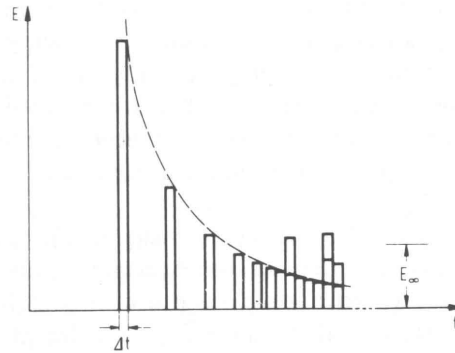


Figura 4.7: rappresentazione energetica della risposta impulsiva.

La relazione precedente mostra come il segnale risultante in uscita è dato dall'integrale di convoluzione del segnale d'ingresso con una serie di impulsi cui danno contributo diversi fattori: il primo è il suono diretto; quelli immediatamente susseguenti sono detti "riflessioni precoci".

Sotto il profilo temporale l'andamento della risposta impulsiva è ottenibile, ad esempio, come antitrasformata del rapporto fra le trasformate dei segnali d'uscita e d'ingresso

$$h(t) = f^{-1}[Y(f) / X(f)] \quad (4.2)$$

L'andamento temporale di una risposta all'impulso tipica è riportato in figura.

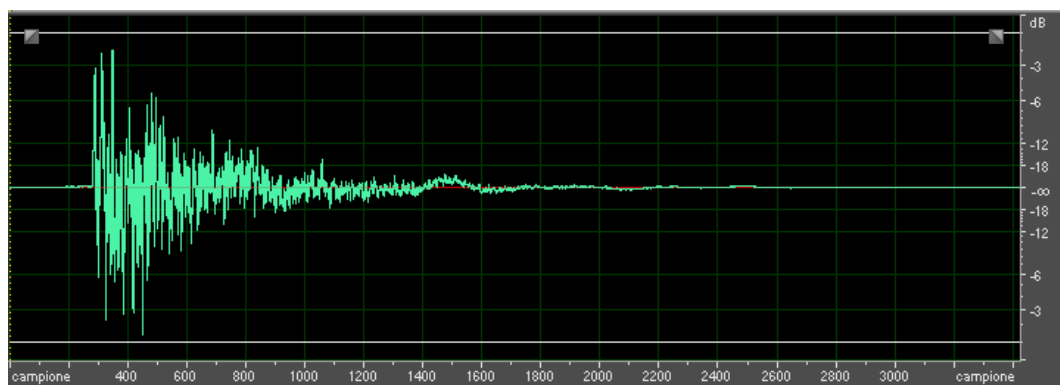


Figura 4.8: risposta impulsiva di un ambiente.

Misura della risposta all'impulso con la tecnica SWEEP

Esistono diversi metodi di misurazione della risposta all'impulso (MLS, TDS, SWEEP, ecc.). Tra questi si è scelto di utilizzare la tecnica SWEEP in quanto in grado di massimizzare il rapporto segnale/rumore e restituire anche in condizioni difficili una misura della risposta all'impulso ottima. Questo tipo di segnale risulta inoltre particolarmente idoneo a velocizzare il calcolo della risposta all'impulso.

Il segnale SWEEP, è un segnale composto da un tono puro la cui frequenza aumenta nel tempo con un determinato andamento come mostrato in Figura 4.9 e in Figura 4.10.

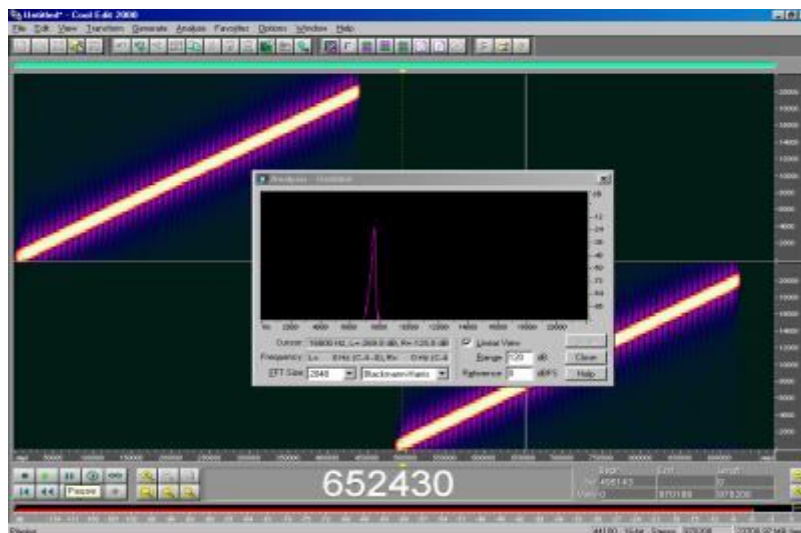


Figura 4.9: analisi di uno SWEEP lineare con Cool Edit.

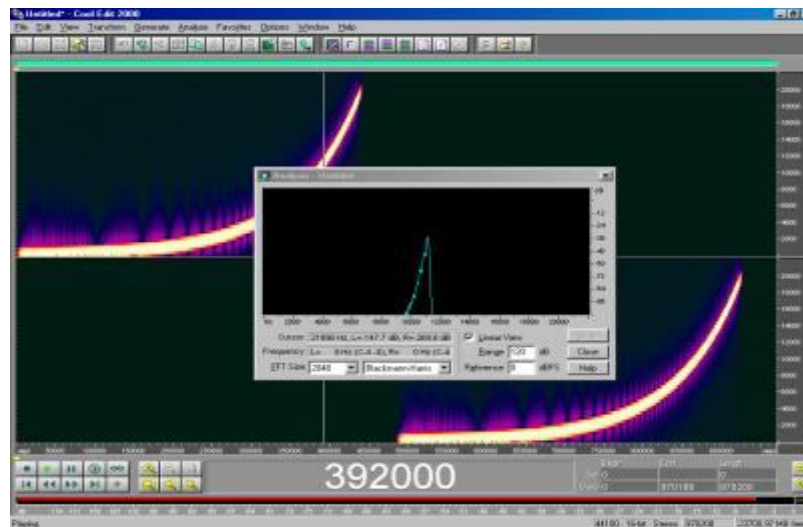


Figura 4.10: analisi di uno SWEEP logaritmico con Cool Edit.

Riprendendo il sistema di partenza di figura 4.6, con in ingresso un segnale SWEEP (lineare o logaritmico), esiste un filtro inverso $x'(t)$ tale che

$$x(t) \otimes h(t) \otimes x'(t) \quad (4.3)$$

produce la risposta $h(t)$ cercata. Ciò è vero se vale la relazione

$$x(t) \otimes x'(t) = \delta(t) \quad (4.4)$$

In questo caso, il filtro inverso cercato non è altro che il segnale SWEEP lineare stesso rovesciato sull'asse dei tempi. Per calcolare la risposta impulsiva si procede quindi convolvendo la misura della risposta allo SWEEP con lo SWEEP inverso ottenuto con l'operazione di *Time Reversal Mirroring*.

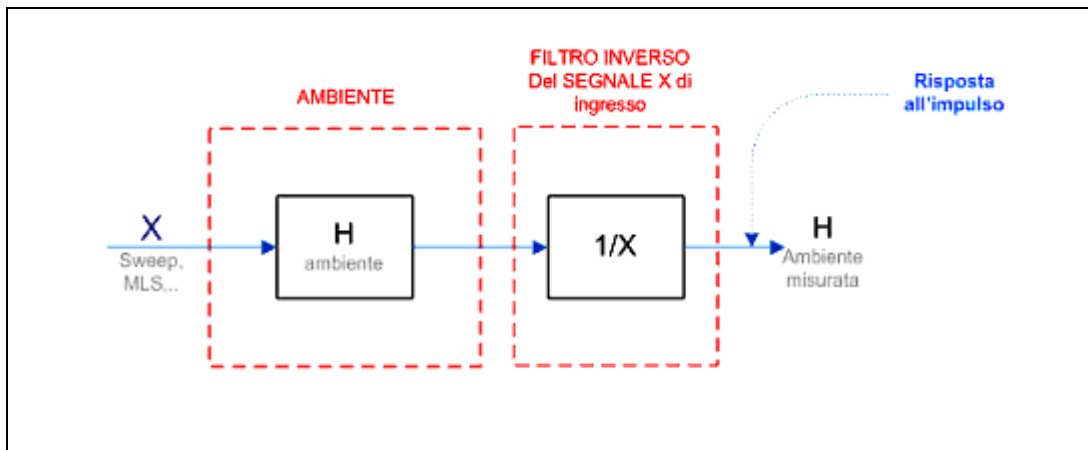


Figura 4.11: schema a blocchi della tecnica di misura della risposta all'impulso.

Evoluzione del metodo AQT: overlapping

Finora è stata accertata la coerenza di risultati fra il metodo AQT e il metodo virtual AQT. Un'ulteriore innovazione però è stata introdotta negli ultimi anni a riguardo. Infatti si è deciso di spezzare il treno di burst, costituito da circa 300 elementi, e fare il virtual AQT, non più sul segnale complessivo, ma burst per burst. Questo significa convolvere singolarmente i burst del treno con la risposta dell'ambiente e poi ricostruire solamente in un secondo momento il segnale utile al calcolo dell'articolazione. In tal modo è possibile rendere la stimolazione completamente localizzata in un'area di frequenze.

Una volta effettuata la convoluzione per tutti i 300 burst che costituiscono il segnale AQT, si ricostruisce la risposta complessiva sovrapposendo le risposte singole in corrispondenza della fine del periodo di silenzio. Tale modo di procedere risulta efficace e molto attendibile. Implementare questa nuova tecnica risulta molto semplice utilizzando delle matrici bidimensionali, in cui viene riportata lungo le righe la successione di bit che descrive la risposta temporale, e lungo le colonne il numero identificativo del burst.

4.4 Analisi della risposta

Abbiamo ottenuto tramite il metodo di virtual AQT in maniera molto semplice la risposta al segnale AQT del sistema acustico che si vuole caratterizzare. Come è stato accennato si tratta ora di estrarre informazioni utili da essa. A questo riguardo si è già detto come lo scopo ultimo dell'analisi sia determinare in quale misura le code che seguono la riproduzione dei burst possano influire su un successivo messaggio sonoro oscurandolo. Qui vengono in aiuto tutte le considerazioni psicoacustiche sviluppate nel capitolo 3. Innanzitutto si tratta di stimare quale parte della risposta al singolo burst bisogna considerare e in che maniera stimarla. Visto che l'intenzione è determinare quanto velocemente si estinguono le code, un buon modo di procedere è andare a considerare di quanto si smorza la risposta al burst in corrispondenza dei 33 ms che vengono dopo la fine dello stesso. Se la curva cade velocemente si ha una buona risposta del sistema e le riflessioni non influenzano i successivi messaggi sonori. Altrimenti c'è perdita di informazioni sul messaggio trasportato dal fronte originale e questo pregiudica la risposta dinamica del sistema.

L'altro aspetto fondamentale della risposta è l'andamento armonico della curva ovvero il livello di risposta che si ottiene in corrispondenza di ciascun burst. Infatti la risposta ad ogni frequenza dipende molto dall'andamento dei coefficienti d'assorbimento dei materiali presenti all'interno dell'autovettura e dalla presenza di onde stazionarie. Esse come sappiamo provocano una distribuzione di pressione non uniforme ma caratterizzata da zone ad elevata pressione e zone a pressione nulla. Questo altera la risposta armonica del sistema e diventa così fondamentale ottenere una curva che dia una stima dell'andamento armonico del sistema e determinare il suo scostamento rispetto ad una curva di massima piacevolezza, indicata con curva target (vedi paragrafo 1.2).

4.5 Articolazione

L'articolazione fornisce una misura della dinamica del sistema ottenuta da considerazioni prettamente psicoacustiche. Infatti dall'esperienza degli acustici emerge che ai fini di una piacevole percezione di segnali fortemente non stazionari, quali sono i brani musicali in genere, assumono un'importanza rilevante le fasi di attacco e di stacco. Un sistema che risponde agli stacchi e agli attacchi con una costante di tempo elevata e quindi non prontamente pregiudica fortemente la dinamica del segnale originale. Riconducendosi al discorso fatto sulle riflessioni precoci, un sistema in cui le riflessioni precoci abbiano una energia elevata e non si estinguano rapidamente è un sistema pessimo dal punto di vista della dinamica e ad esso corrisponde un basso valore di articolazione.

Oltre a pregiudicare la dinamica del sistema, le forti riflessioni, a causa del mascheramento, provocano un oscuramento delle frequenze successivamente emesse dal diffusore e quindi parte dell'informazione musicale viene perduta e il messaggio sonoro risulta 'povero' di dettagli. In tal senso l'articolazione diventa fondamentale nella valutazione acustica di un sistema.

Nella valutazione dell'articolazione si considera unicamente la fase di stacco del burst che viene interessata dal fenomeno del post-mascheramento, che come detto è di gran lunga il maggior responsabile della sensazione dinamica restituita da un sistema. La valutazione dell'articolazione viene effettuata partendo dall'istante in cui finisce il burst di stimolo e calcolando di quanti db si attenua la risposta ad esso entro un intervallo di tempo definito che noi abbiamo già stimato essere di 33 ms. Sistemi con poca dinamica avranno una coda nella risposta al burst che si estingue con una pendenza molto dolce e quindi un'articolazione molto bassa. Viceversa sistemi dotati di un'ottima dinamica permettono alla coda di estinguersi con una derivata molto maggiore. In figura 4.12 è riportata una schematizzazione di come viene effettuato il calcolo dell'articolazione.

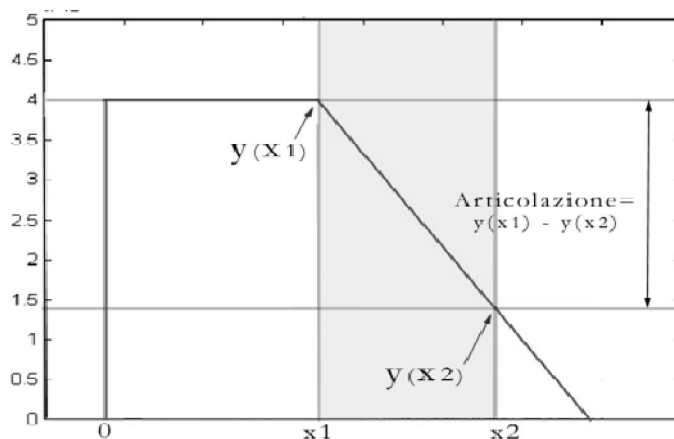


Figura 4.12 : calcolo dell'articolazione.

Di seguito invece viene riportata la formula per il calcolo della stessa:

$$\text{Art} = y(x_1) - y(x_2) \quad (4.5)$$

In realtà tale modo di procedere applicato ad alcuni casi particolari ha evidenziato un problema fondamentale del metodo. Infatti con tale procedimento si sono ottenute in alcune circostanze valori negativi di articolazione. Questo ovviamente non ha senso per la definizione stessa che ne è stata data. Il motivo che sta alla base di questi risultati poco attendibili è che, soprattutto in bassa frequenza si hanno casi in cui il massimo della risposta al burst non cade all'interno dei 200 ms che costituiscono la durata temporale del burst. Questo è dovuto alla presenza delle onde stazionarie e all'andamento dei coefficienti d'assorbimento dei materiali all'interno dell'autovettura. In figura 4.13 se ne dà un esempio da cui è possibile estrarre considerazioni interessanti: all'inizio del burst è presente un picco che corrisponde all'arrivo del fronte diretto. Successivamente in seguito all'arrivo di onde in controfase il livello sonoro si abbassa. Alla fine dei 200 ms, quando termina il burst, si estingue il contributo del fronte diretto e quindi rimane solo l'onda in controfase con il corrispondente picco. Quello che noi sentiamo però a causa dell'effetto Haas è il picco che giunge in seguito alla fine dello stacco e quindi da questo punto l'orecchio comincia a

riacquistare la sensibilità necessaria a percepire un fronte sonoro successivo e in corrispondenza di tale punto bisogna partire con il calcolo dell'articolazione

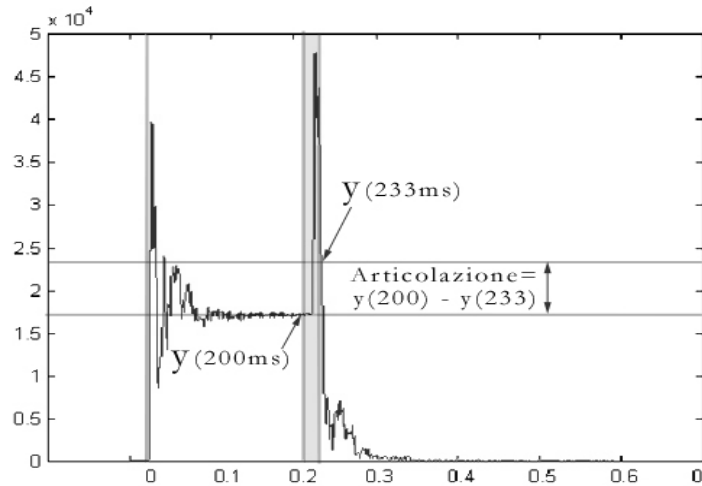


Figura 4.13 : esempio di articolazione negativa.

Infatti il recupero della sensibilità uditiva in presenza di un tono di mascheramento descritto dalla curva di Zwicker, riportata in figura 3.15, parte dal momento in cui effettivamente il segnale di mascheramento comincia ad estinguersi. Nel caso in esame il segnale di masking non è rappresentato dal segnale AQT di origine ma dalla risposta dell'ambiente a tale stimolo.

Il calcolo dell'articolazione allora deve essere presentato nella forma:

$$\text{Art} = y(x_{\max}) - y(x_{\max} + 33 \text{ ms}) \quad \text{con } x_{\max} \geq 200 \text{ ms} \quad (4.6)$$

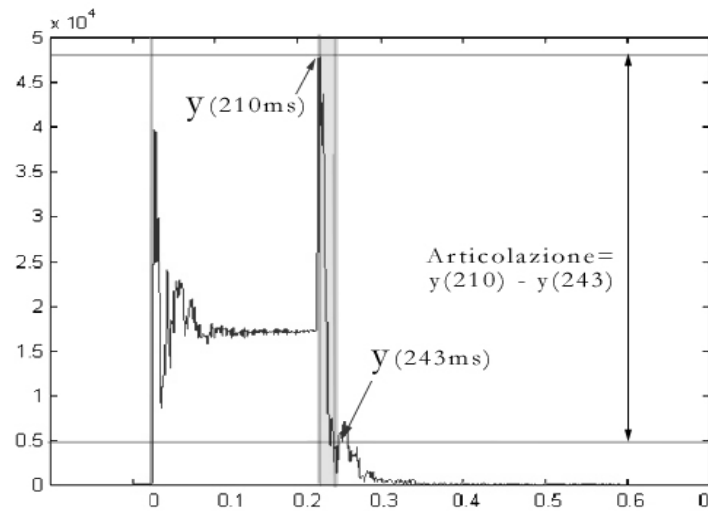


Figura 4.14: esempio di articolazione positiva.

A questo punto abbiamo ricavato l'andamento dell'articolazione all'interno dell'ambiente che vogliamo caratterizzare. In particolare tale procedimento applicato a tutti i burst di risposta del segnale AQT fornisce in uscita una serie di dati corrispondenti ai valori di articolazione nei punti caratteristici in cui si è scelto di generare i burst.

4.6 Andamento armonico

Il bilanciamento esprime lo scostamento della risposta armonica del sistema rispetto ad un target d'ascolto che viene giudicato ottimale. La curva di risposta armonica viene calcolata non attraverso l'analisi classica ma utilizzando le informazioni fornite dal metodo AQT. Viene dunque introdotta una nuova procedura che teoricamente è in grado di fornire la curva di risposta effettivamente percepita dal sistema uditivo umano. La differenza rispetto a uno spettro in frequenza classicamente calcolato con una FFT sta nel fatto che grazie ad AQT è possibile avere un'informazione anche sul comportamento temporale della risposta in frequenza. Tipicamente l'andamento armonico percepito dal nostro sistema uditivo segue l'involuppo dei picchi in quanto a causa dell'effetto

Haas il segnale viene integrato su un determinato intervallo di tempo e quindi il picco risulta contribuire maggiormente alla determinazione della sensazione di livello sonoro restituita. I passaggi che permettono di arrivare alla determinazione del bilanciamento vengono esposti per punti per motivi di chiarezza.

Il calcolo del bilanciamento viene fatto su base psicoacustica partendo dall'involuppo della curva AQT. L'involuppo è stato calcolato prendendo il valore massimo che la risposta AQT assume in corrispondenza di ogni burst nei primi 25 ms della stimolazione; naturalmente la risposta AQT che viene considerata non è quella che esce direttamente dalla convoluzione fra risposta all'impulso e segnale di stimolo AQT, ma quella che è stata involupata e interpolata in corrispondenza di ogni burst durante il calcolo dell'articolazione. Il periodo di 25 ms, come detto, corrisponde alla costante di tempo del sistema uditivo umano, che integra su tale intervallo il segnale di pressione che viene rilevato dal timpano.

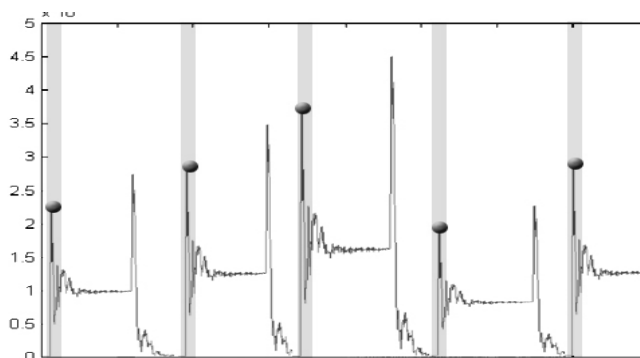


Figura 4.15 : ricerca dei punti di massimo nei primi 25ms della risposta AQT ai vari burst.

A questo punto ricavati i valori di massimi della risposta e i punti in cui tali massimi cadono sulla scala di Bark, è necessario eseguire un'interpolazione su tutto il range di frequenze in modo da ottenere la curva di risposta armonica nel dominio classico delle frequenze lineari. Il fatto di avere a disposizione la curva sul dominio classico dell'analisi di Fourier discreta è essenziale nel momento in cui si decida di implementare l'equalizzazione della stessa tramite filtri FIR. Essi infatti necessitano, per essere generati, della risposta in forma classica. La curva

che si ottiene viene considerata come la risposta in frequenza dell'ambiente effettivamente percepita dal sistema uditivo umano.

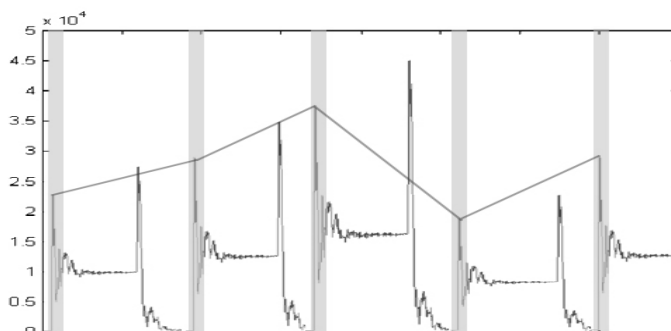


Figura 4.16: interpolazione dei massimi e inviluppo della curva di risposta armonica.

La curva viene poi normalizzata rispetto al suo valor medio e i valori della stessa vengono riportati in dB. Nella pratica questo corrisponde a traslarla intorno all'origine degli assi. Procedendo in tal modo si dà peso unicamente all'andamento relativo della curva ovvero alla sua forma piuttosto che al suo valore assoluto. Questo equivale anche a considerare sull'intero range di frequenze la distribuzione relativa dell'energia privata del suo valor medio. Si giunge allo stesso risultato facendo la media aritmetica dei valori in dB della curva e sottraendola al valore assunto dalla curva in ogni punto. Infatti il logaritmo di una divisione è pari alla sottrazione dei logaritmi di dividendo e divisore.

La curva così ottenuta viene poi smussata adottando una risoluzione ad 1/3 di ottava ed utilizzata per calcolare il bilanciamento. Questa rappresenta la curva complessiva di andamento armonico. La scelta di smussare tale curva è dovuta al fatto che, come è stato esposto nella parte psicoacustica, il segnale tende ad essere processato per bande di rumore equivalente e quindi non assumono tanto importanza difetti locali dello spettro del segnale quanto la distribuzione energetica dello stesso. Sotto questo punto di vista l'analisi può essere considerata prettamente energetica. La curva che si ottiene in seguito a quanto detto è la curva

che deve essere considerata come risposta del sistema e che deve essere utilizzata in sede di equalizzazione.

CAPITOLO 5

Sviluppi nell'analisi AQT

Nel presente capitolo verranno presentati i lavori di sviluppo della tecnica di misura AQT e sarà brevemente illustrato il software che realizza le funzioni di analisi acustica e sintesi dei filtri inversi.

5.1 Burst con inviluppo a dissolvenza

Come noto dal capitolo 4, l'inviluppo di un burst sinusoidale nell'AQT classico è di tipo rettangolare. Ciò implica che i transitori di iniziali e finali siano a larghezza di banda elevata. In figura 5.1 è possibile osservare lo spettrogramma di un burst di questo tipo. Dall'osservazione di questo tipo di analisi, è emersa l'idea di modificare i transitori aggiungendo una dissolvenza (di tipo sinusoidale) d'ingresso ed uscita (fade-in e fade-out) in modo da controllarne la larghezza di banda. La forma della finestra applicata al burst è mostrata in figura 5.2. Il transitorio totale ΔT è stato calcolato come l'inverso della spaziatura tra due frequenze consecutive analizzate sulla banda di Bark. In figura 5.3 e 5.4 sono riportati gli inviluppi di alcuni burst per varie frequenze.

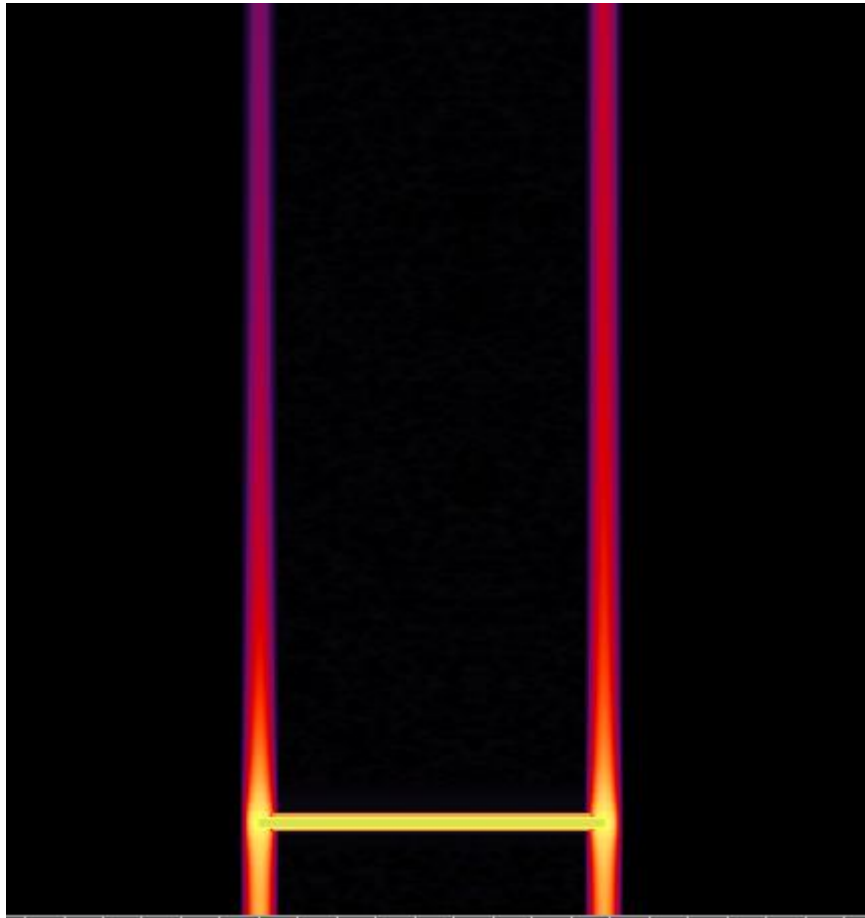


Figura 5.1: particolare del sonogramma di un burst del metodo AQT originale.

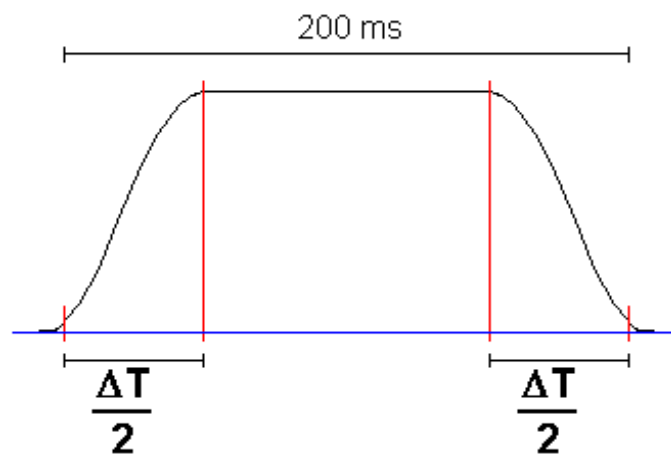


Figura 5.2: esempio di inviluppo con dissolvenze sinusoidali.

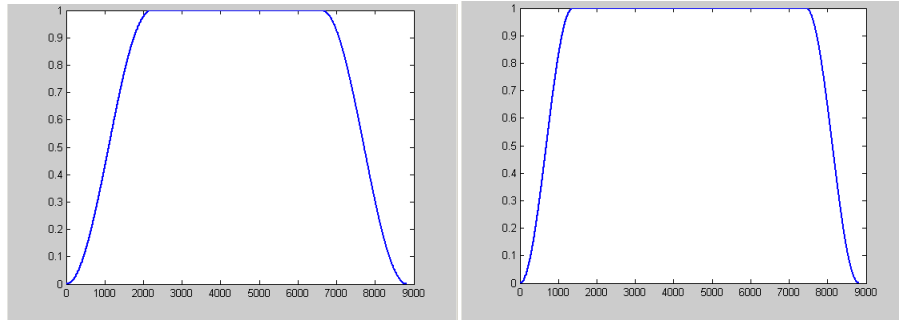


Figura 5.3: esempi di burst con dissolvenza (100 e 1000 Hz).

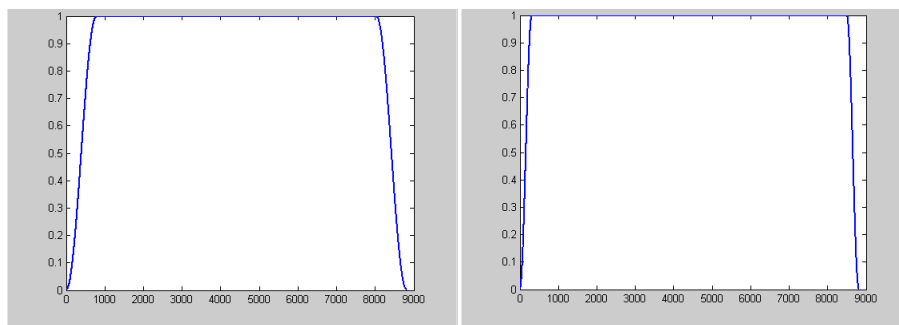


Figura 5.4: esempi di burst con dissolvenza (2 e 4 KHz).

Nelle figure 5.5 e 5.6 sono mostrati gli involuপি dei burst convoluti con una risposta all'impulso misurata.

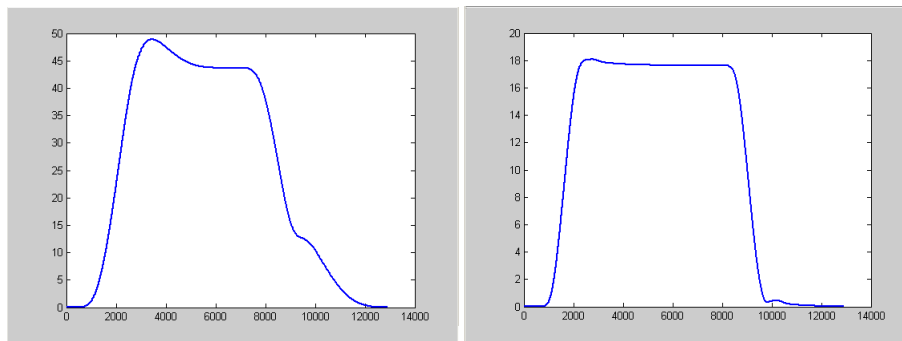


Figura 5.5: esempi di burst con dissolvenza (100 e 1000 Hz) dopo la convoluzione.

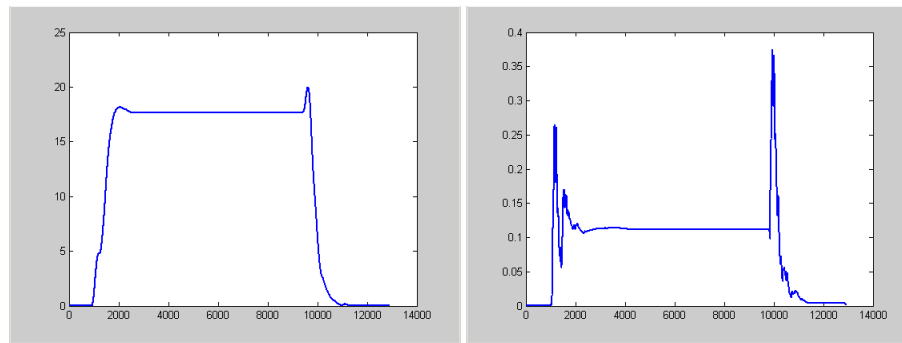


Figura 5.6: esempi di burst con dissolvenza (2 e 4 KHz) dopo la convoluzione.

Si può osservare che la dissolvenza applicata ai transitori, porta ad un addolcimento dei fenomeni di cancellazione o somma delle riflessioni dell'ambiente. Questo fatto riconduce questo tipo di analisi a risultati simili a quelli di una normale analisi mediante DFT. Infatti a causa dello smussamento degli estremi e del fatto che la durata dell'impulso misurato in automobile è normalmente minore di 200 ms, il massimo dell'involuppo si trova quasi sempre nella zona centrale del burst ed è quindi il rispecchiamento di una condizione di regime. Queste affermazioni sono valide in particolar modo per i burst a bassa frequenza, dove la banda di attacco/rilascio del singolo burst è stretta e di conseguenza i transitori temporali sono relativamente lunghi. Si riporta in figura 5.7 un confronto tra risposta armonica DFT e quella con burst con involuppo a dissolvenza.

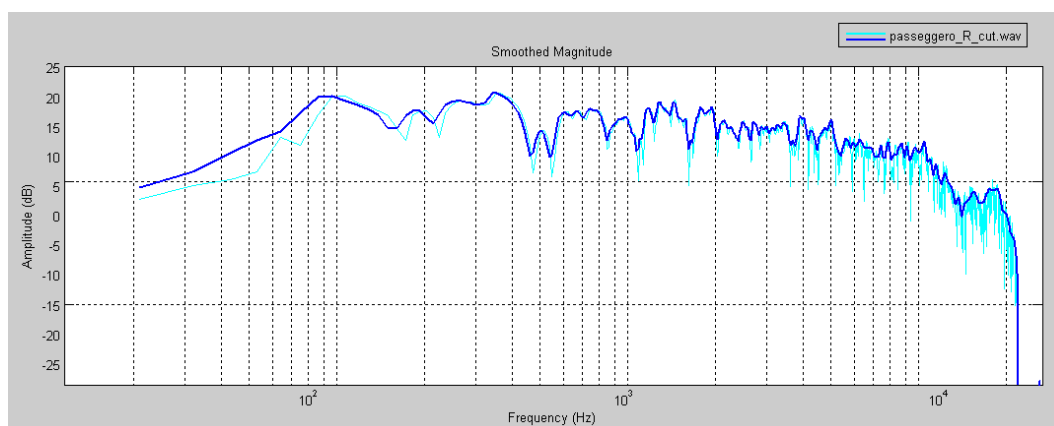


Figura 5.7: confronto tra risposta armonica DFT (azzurro) e AQT con burst a dissolvenza (blu).

5.2 Burst ad inviluppo esponenziale

Proseguendo la strada dello studio dell'inviluppo del burst, è stata apportata la seguente considerazione psicoacustica. L'orecchio umano ha, come ogni sistema, delle costanti di tempo proprie le quali influenzano anche il fenomeno del mascheramento. Col proposito di tener conto di questo fenomeno psicoacustico è stato applicato un inviluppo esponenziale al burst. In questo modo viene simulato il comportamento dell'orecchio tenendo conto anche del mascheramento. In questo contesto non è più necessario limitare la ricerca del massimo dell'inviluppo ai primi 25 ms o applicare finestre di media temporale di larghezze prefissate come visto nel calcolo dell'AQT tradizionale. La scelta della costante di tempo esponenziale è stata impostata a 80 ms (valore considerato anche in alcuni parametri per la misura delle prestazioni delle sale d'ascolto. Ad es. C_{80}) ma non si esclude di proseguire l'indagine con altri valori. Seguono nelle figure alcuni esempi di inviluppi di burst esponenziali.

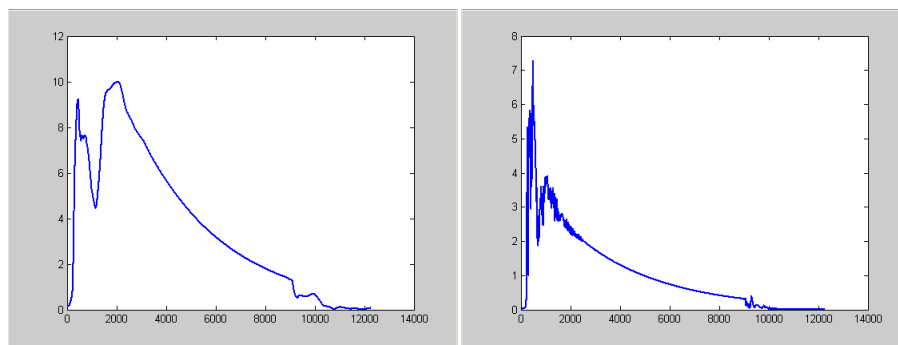


Figura 5.8: esempi di burst esponenziali (100 e 1000 Hz) dopo la convoluzione.

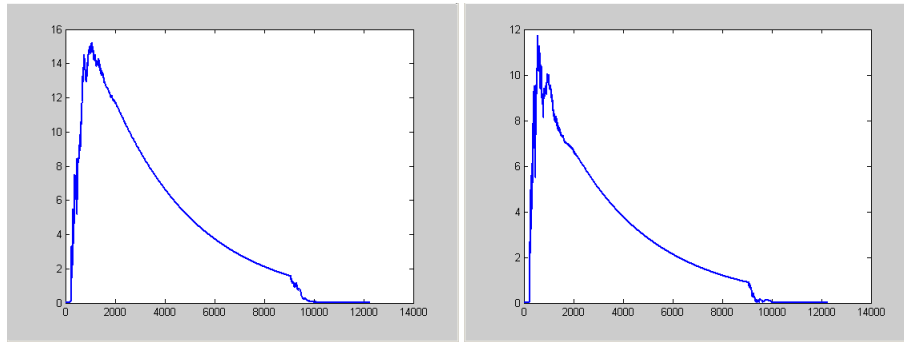


Figura 5.9: esempi di burst esponenziali (2 e 4 KHz) dopo la convoluzione.

In figura 5.10 è riportato uno spettrogramma di un burst con involucro esponenziale, mentre in figura 5.11 è raffigurata la risposta armonica ottenuta da un'analisi AQT con burst esponenziali.

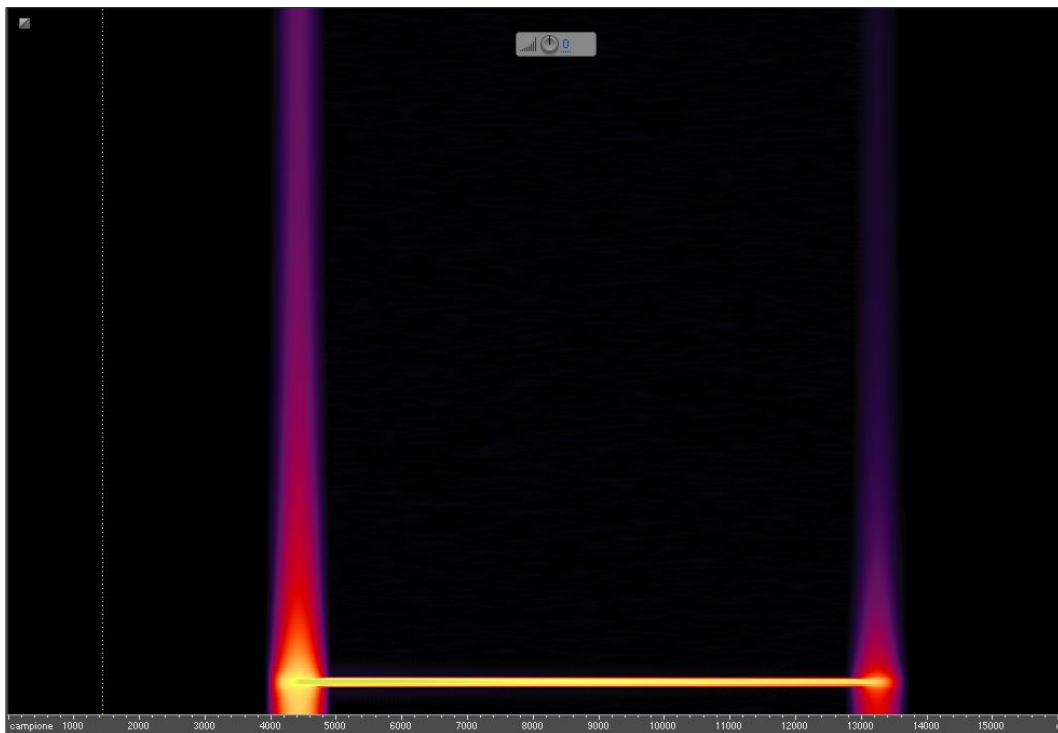


Figura 5.10: particolare del sonogramma di un burst esponenziale.

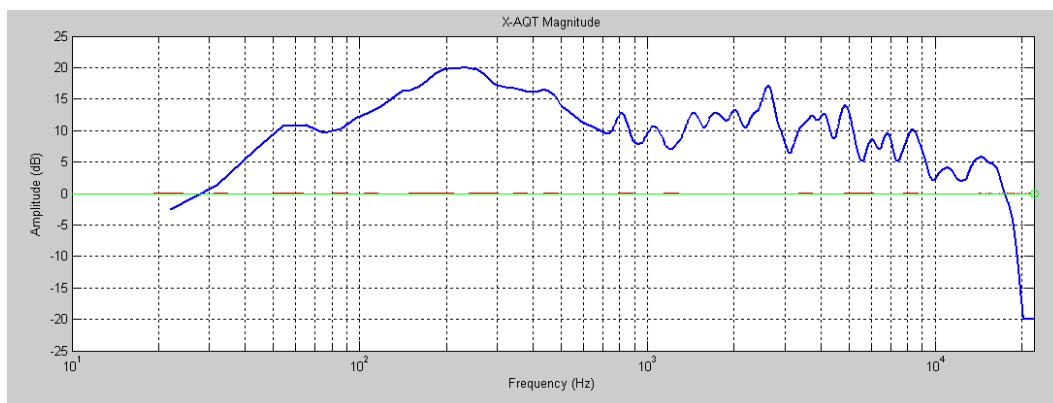


Figura 5.11: grafico di una risposta calcolata con burst esponenziali.

5.3 Confronto tra inviluppo tradizionale, a dissolvenza ed esponenziale

Nelle figure seguenti sono illustrati i confronti tra le risposte in frequenza ricavate dall'analisi DFT, AQT tradizionale, con inviluppo a dissolvenza ed esponenziale per una data risposta impulsiva. In figura 5.12 è riportato il confronto tra modulo DFT ed AQT tradizionale mentre in figura 5.13 è mostrato il confronto tra modulo DFT e AQT a dissolvenza.

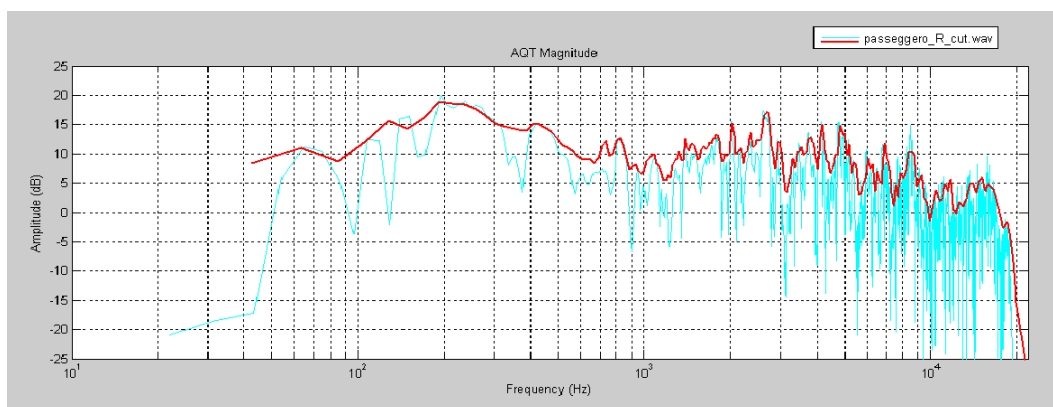


Figura 5.12: confronto tra risposta armonica DFT (azzurro) e AQT tradizionale (rosso).

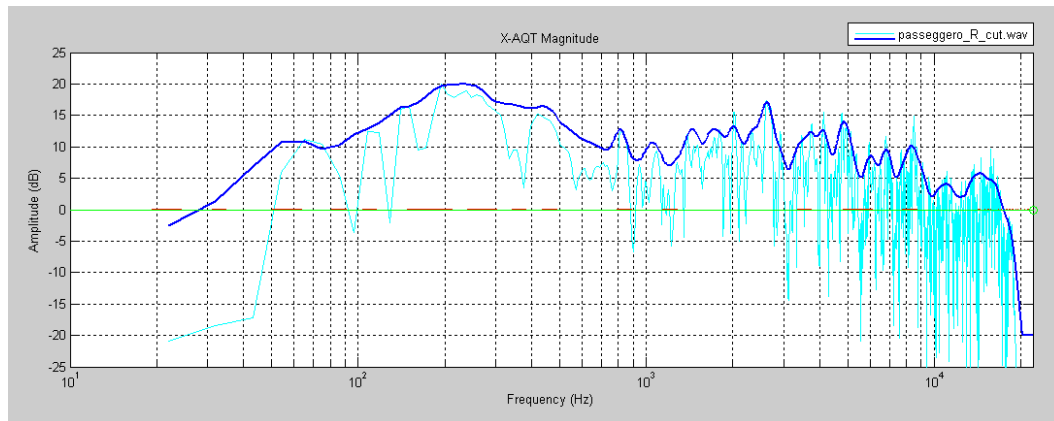


Figura 5.13: confronto tra risposta armonica DFT (azzurro) e AQT ad involuppo esponenziale (blu).

In entrambi i casi si può notare che il modulo AQT ricalca una sorta di involuppo dello spettro DFT. La differenza tra AQT a dissolvenza ed esponenziale è invece evidenziata in figura 5.14.

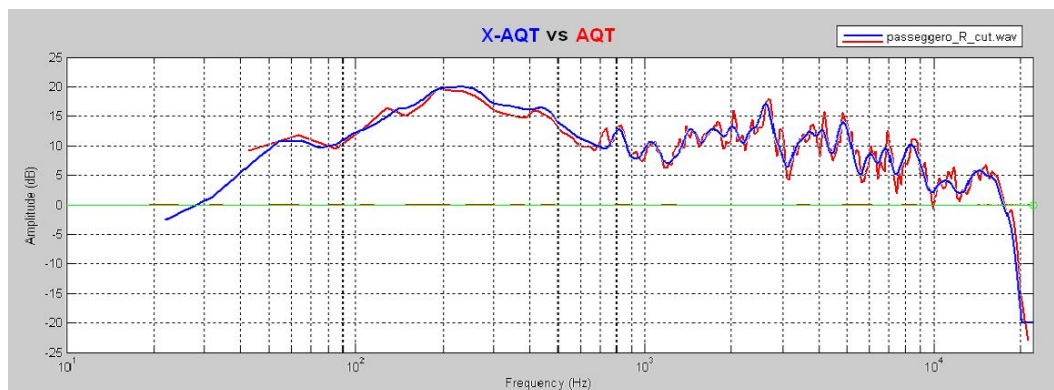


Figura 5.14: confronto tra risposta armonica AQT a dissolvenza (rosso) e AQT esponenziale (blu).

L'AQT esponenziale (X-AQT), oltre alle proprietà di cui già discusso sopra, risulta essere intrinsecamente più levigato e quindi dà maggiori probabilità di ottenere un filtro di equalizzazione piacevole (vedi paragrafo 2.2). Questi fatti portano a ritenere che il sistema di analisi AQT esponenziale sia ancora più rappresentativo della percezione uditiva umana di quanto non sia l'analisi AQT tradizionale.

5.4 Studi sulla fase

Come noto [3] [9] il metodo di analisi AQT riguarda solamente le risposte armoniche in ampiezza, trascurando invece la fase. E' inoltre risaputo [10] [11] che la funzione di trasferimento di un sistema acustico costituito da diffusori ed ambiente non è a fase minima. Ciò implica che non sia possibile un'equalizzazione totale del sistema. Queste considerazioni portano a ritenere di dover indagare nuovi metodi di calcolo della fase, andando di nuovo oltre la DFT.

Il metodo, in fase di sperimentazione, prevede di associare ai filtri di equalizzazione una fase tale da riallineare temporalmente i burst. In figura 5.15 è riportato un grafico dei ritardi associati ai burst calcolati mediante cross-correlazione tra burst filtrato e non.

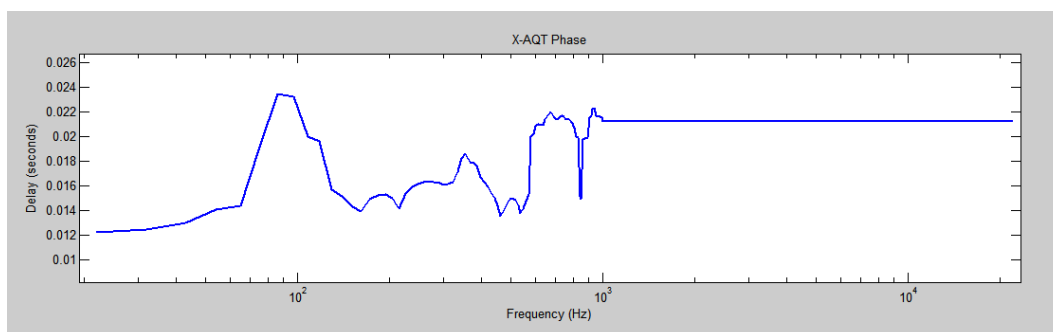


Figura 5.15: ritardo associato ai burst (forzato ad un valore costante sopra 1 KHz).

5.5 Il software di calcolo

Nell'ottica della collaborazione con ASK industries s.p.a. è stato realizzato un software su piattaforma Matlab che permette anche ad un utente ignaro dei dettagli di funzionamento, di analizzare una risposta impulsiva acustica con i metodi precedentemente proposti e sintetizzare filtri di correzione da utilizzare su un generico DSP. Tale software, denominato "Askm", può essere eseguito anche

su calcolatori sprovvisti di software Matlab in quanto è stata realizzata un versione stand alone, eseguibile da qualsiasi PC. Gli aspetti sopra elencati permetteranno un utilizzo immediato da parte degli utilizzatori aziendali.

Askm è dotato di interfaccia grafica che verrà di seguito illustrata. In figura 5.16 è possibile osservare la finestra principale del programma suddivisa in due riquadri. Nel riquadro superiore sono rappresentati i moduli ottenuti dall'analisi (AQT in blu e FFT in azzurro) ed i moduli "target" (verde e rosso). La curva target verde è trascinabile mediante il puntatore, mentre la curva target rossa è la versione spettralmente levigata. Nel riquadro inferiore è riportato il ritardo AQT.

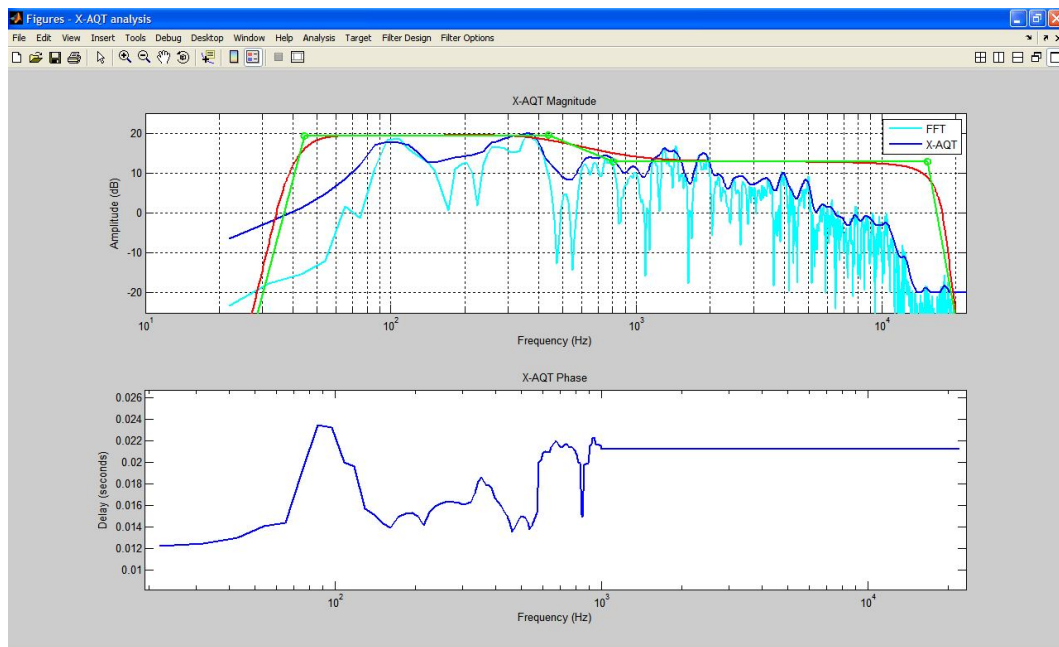


Figura 5.16: finestra principale di Askm.

In figura 5.17 è mostrata una finestra d'analisi in cui sono presenti più moduli. Tale caratteristica è stata sviluppata al fine di poter calcolare un filtro partendo dalla media di più misure. Ciò può servire ad aumentare l'immunità ai disturbi o la robustezza spaziale di un filtro [6].

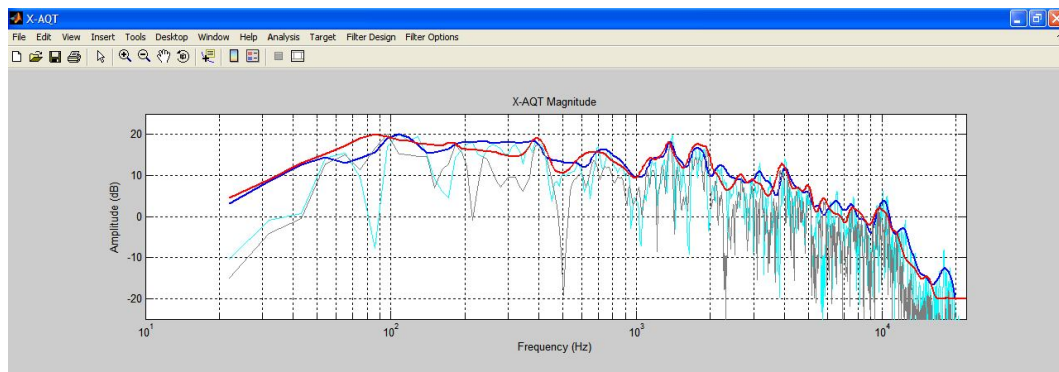


Figura 5.17: finestra di analisi con moduli multipli.

La risoluzione spettrale della curva può essere impostata, in frazioni d'ottava, mediante l'apposito menù riportato in figura 5.18. Lo stesso menù consente di caricare, salvare o reimpostare la curva.

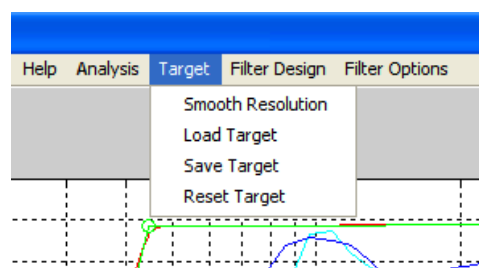


Figura 5.18: menù Target di Askm.

Anche la risoluzione del modulo AQT, la quale andrà poi ad influenzare quella del filtro di equalizzazione, può essere modificata tramite la finestra di dialogo mostrata in figura 5.19. Qui è possibile scegliere il tipo di risoluzione da adottare per la levigatura sia in frazioni d'ottava che in ERB e CB (vedi paragrafo 2.1).

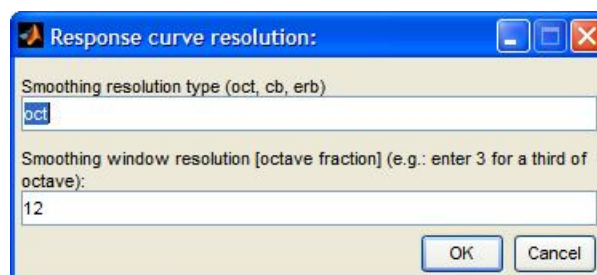


Figura 5.19: finestra di dialogo per l'impostazione della risoluzione di levigatura spettrale.

Nelle figure 5.20 e 5.21, sono evidenziate le finestre di dialogo che permettono di impostare la lunghezza del filtro inverso ed i parametri di regolarizzazione relativi al calcolo di tale filtro secondo il metodo di Kirkeby. Da notare che la lunghezza del filtro è impostabile arbitrariamente (maggiore o minore di quella della risposta impulsiva di partenza) permettendo così l'utilizzo anche su piattaforme a bassa capacità computazionale.

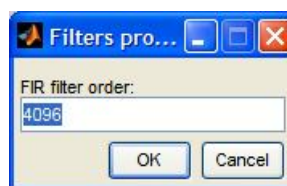


Figura 5.20: finestra di dialogo per l'impostazione della dimensione del filtro di equalizzazione.

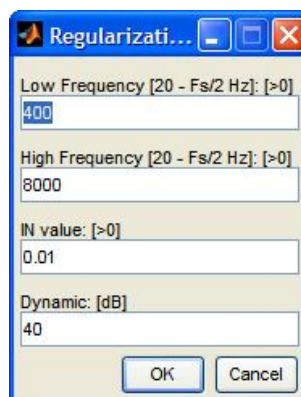


Figura 5.21: finestra di dialogo per l'impostazione dei parametri di regolarizzazione di Kirkeby.

Una volta terminate le operazioni preliminari, è possibile passare al calcolo del filtro. Una volta eseguito il relativo calcolo, il programma crea due finestre di riepilogo esponendo i risultati in termini di moduli e fasi e risposte impulsive. Nella prima finestra, riportata in figura 5.22, si osservano 6 grafici sistemati in 3 righe e 2 colonne. La prima riga contiene il grafico dei moduli relativi alla risposta impulsiva di partenza e la relativa fase; la seconda riga rappresenta modulo e fase del filtro inverso oltre che il valore del parametro di regolarizzazione di Kirkeby (in rosso); la terza riga espone modulo e fase della

convoluzione del filtro con la risposta impulsiva di partenza oltre che la target curve (in verde).

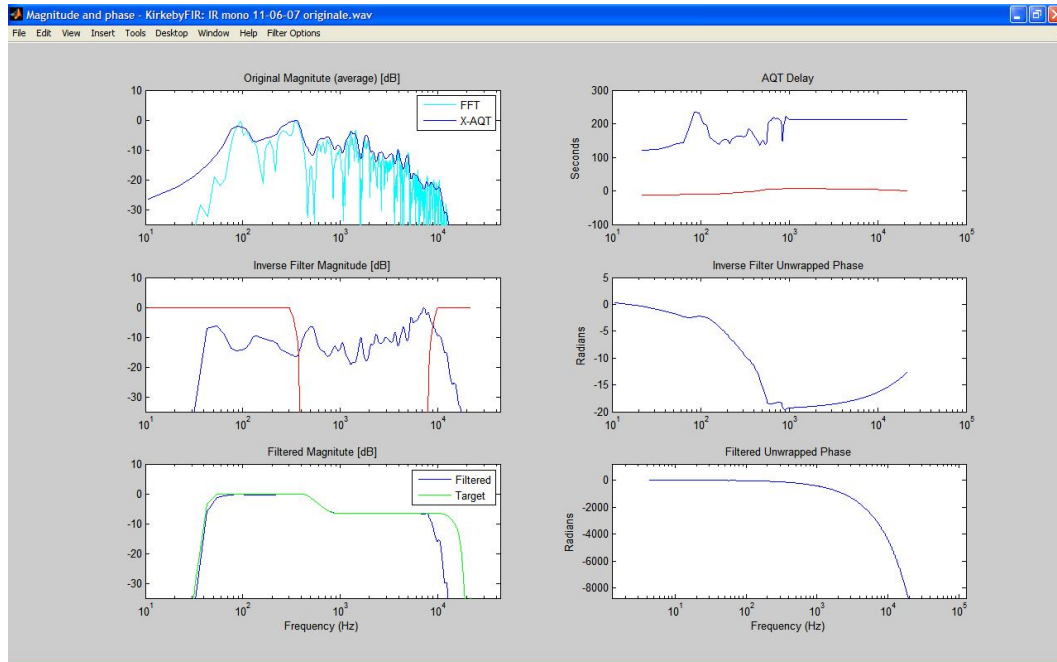


Figura 5.22: finestra di riepilogo finale dei moduli e delle fasi.

Nella seconda finestra, mostrata in figura 5.23, compaiono 3 grafici riportanti le risposte impulsive relative alla misura di partenza, al filtro inverso ed alla convoluzione dei due (in ordine dall'alto verso il basso). Quest'ultima rappresentazione si rivela di particolare interesse nel controllo sulla presenza di eventuali artefatti quali pre o post eco.

Nel caso si abbia l'esigenza di confrontare più di una soluzione di filtraggio cambiando i parametri (lunghezza, regolarizzazione o target), il programma consente di calcolare e confrontare più filtri contemporaneamente creando altre finestre di riepilogo.

Mediante un apposito menù il programma consente il salvataggio dei coefficienti dei filtri generati in formato TXT o WAV.

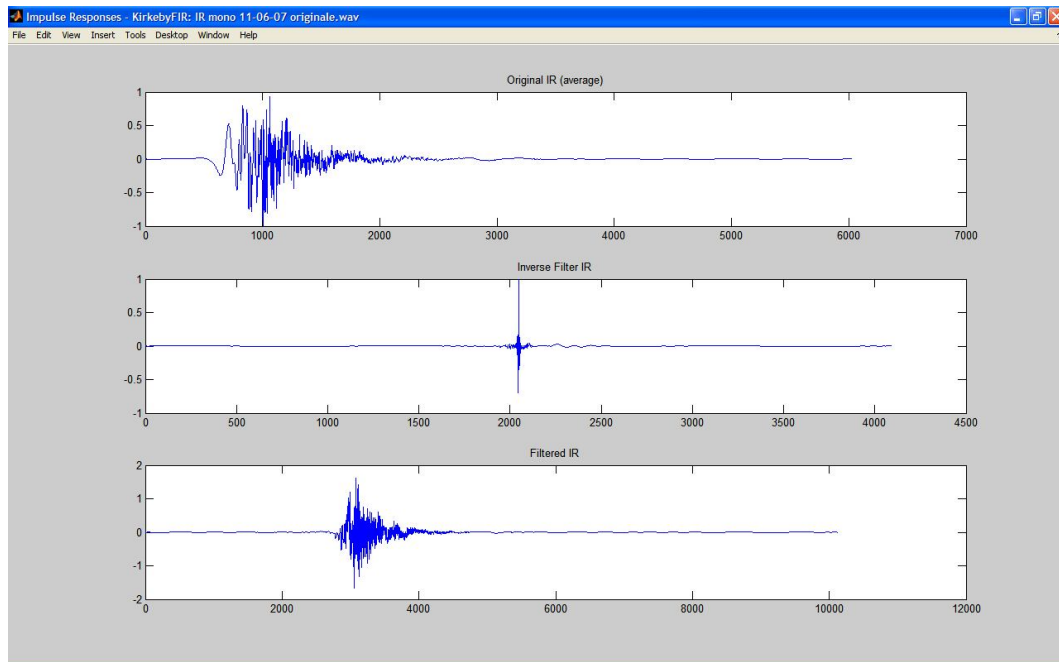


Figura 5.23: finestra di riepilogo finale delle risposte impulsive.

CAPITOLO 6

Conclusioni

L'attività svolta è stata prevalentemente indirizzata allo studio della elaborazione digitale del segnale applicata ai sistemi di riproduzione audio in campo automotive. Sono state studiate, sperimentate e messe a confronto le tecniche tradizionali ed innovative di equalizzazione della risposta del sistema audio, ottenute mediante tecniche quali levigatura e decimazione spettrale, AQT tradizionale, con dissolvenza ed esponenziale, manipolazione della fase.

In particolare è risultato che l'uso di tecniche basate sulla psicoacustica può portare alla realizzazione di sistemi di qualità eccellente, anche in un ambiente particolarmente ostico come l'abitacolo di un'automobile. E' ormai chiaro che è questa la strada da percorrere nelle future indagini.

Considerando il lavoro svolto, è possibile affermare che sono stati compiuti importanti passi verso la realizzazione di un sistema di equalizzazione automatico in quanto l'utilizzo delle nuove tecniche AQT permette di ridurre al minimo l'intervento di valutazione umano. Il passo successivo verso quest'obiettivo sarà la realizzazione di un software in grado di essere eseguito da un sistema di calcolo integrato nel sistema di riproduzione e capace di autocalcolare i filtri di equalizzazione senza intervento dell'operatore.

Bibliografia

- [1] A. Bruce Carlson, “*Communication system*”. McGraw-Hill international editions, 1986.
- [2] A.V. Oppenheim Shafer, “*Elaborazione numerica dei segnali*”. Franco Angeli, 11 edition, 1996.
- [3] A. Azzali, Tesi di Dottorato, “*Elaborazione del metodo dinamico IPM*”, 2006
- [4] E. Ugolotti, G. Gobbi, A. Farina, “*IPA - A Subjective Assessment Method of Sound Quality of Car Sound Systems*”, Proceedings of the 110th AES Convention, 2001 May 12-15 Amsterdam, The Netherlands.
- [5] S. Bharitkar, “Variable-octave complex smoothing”, Proc. of the 123th AES Convention, preprint 7264, (2007).
- [6] P.D. Hatziantoniou and J.N. Mourjopoulos, “Real-time room equalization based on complex smoothing: robustness results”, Proc. of the 116th AES Convention, preprint 6070, (2004).
- [7] E.Zwicker, H.Fastl, “*Psychoacoustics, Facts and Models*”, Springer- Verlag Berlin Heidelberg 1990
- [8] A.Farina, G.Cibelli , A.Bellini, “AQT – A New Objective Measurement Of The Acoustical Quality Of Sound Reproduction In Small Compartments”, AES 110th Convention Paper, Amsterdam 2001
- [9] I. Adami, F. Liberatore, “*La messa a punto del sistema Diffusori-Ambiente*”, Acustica Applicata srl, Via Roma 79, Galliciano – Lucca - Italy
- [10] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response" J. Acoust. Soc. Am. 66 (1) July 1979.
- [11] L. G. Johansen, P. Rubak, “The excess phase in loudspeaker-room transfer functions”, 1996