

**UNIVERSITÀ DEGLI STUDI DI PADOVA**

**DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE**

**TESI DI LAUREA**

**DECOMPOSIZIONE STRUTTURALE DEL  
CONTRIBUTO ACUSTICO DELL'ORECCHIO  
ESTERNO PER IL RENDERING SPAZIALE DEL  
SUONO**

**Laureando:** Fabrizio Granza

**Relatore:** prof. Federico Avanzini

**Correlatore:** ing. Michele Geronazzo

**Corso di Laurea Magistrale in Ingegneria Informatica**

12 Marzo 2013

Anno Accademico 2012 - 2013



# Abstract

I suoni che circondano un ascoltatore giungono al timpano a seguito dell'interazione delle onde sonore con l'ambiente e con le parti del corpo dell'ascoltatore stesso, in particolare la testa, il busto, le spalle e l'orecchio esterno (termine inglese: *pinna*). Gli effetti acustici risultanti per una particolare posizione della sorgente sonora, quali riflessioni e diffrazioni, sono contenuti nella *Head Related Transfer Function* (HRTF); in particolare si definisce la *Pinna Related Transfer Function* (PRTF) considerando singolarmente i contributi dell'orecchio esterno.

L'obiettivo di questo lavoro di tesi è analizzare la PRTF per la localizzazione del suono nello spazio. Considerando le HRTF per diversi individui appartenenti ad un database di dominio pubblico, l'analisi dei dati si sviluppa in una procedura sperimentale di separazione degli effetti dell'orecchio esterno dalla HRTF completa, applicabile a posizioni della sorgente sonora poco studiate in letteratura. I risultati ottenuti forniscono delle decise indicazioni che favoriscono il raffinamento dei modelli strutturali sintetici impiegati per la riproduzione di audio binaurale in cuffia.



# Introduzione

Questo lavoro di tesi mira a studiare gli effetti forniti dall'orecchio esterno per la localizzazione della sorgente sonora nello spazio. La letteratura conferma che la capacità di discriminare la direzione del suono in elevazione dipende in maniera determinante proprio dall'orecchio esterno (Batteau, 1964); tuttavia tali studi si limitano ad analizzare le *Head Related Transfer Function* (HRTF) relative a sorgenti sonore poste nel piano mediano, ovvero equidistanti dalle due orecchie dell'ascoltatore. In questo lavoro di tesi l'analisi viene ampliata anche al di fuori del piano mediano (nei piani sagittali), dove viene a mancare l'accettabile assunzione semplificativa relativa alla trascurabilità degli effetti acustici prodotti dalla testa. Per tale motivo, avvalendosi di tecniche di elaborazione, viene proposta una metodologia che consente la separazione dei soli effetti dell'orecchio esterno a partire da HRTF complete, contenenti i contributi acustici anche delle altre parti del corpo (come busto, spalle e testa).

Lo studio degli effetti dell'orecchio esterno nei piani sagittali ha come scopo principale il miglioramento della resa di modelli strutturali sintetici per la riproduzione individuale di audio binaurale in cuffia. La caratteristica fondamentale di tali modelli risiede nella diretta parametrizzazione della struttura a filtri attraverso grandezze antropometriche estratte dal singolo soggetto.

Il primo capitolo della tesi è dedicato ad una raccolta bibliografica relativa allo stato dell'arte, dalle basi del sistema uditivo e della geometria dello spazio ai contributi acustici per la localizzazione del suono. Nel secondo capitolo viene approfondita la tecnologia legata all'elaborazione binaurale, descrivendo in particolare la procedura di misurazione delle HRTF. Nel terzo capitolo si introduce il formato di standardizzazione *MARL-NYU*, proposto da Andreopoulou e Roginska alla 131<sup>st</sup> *Audio Engineering Society Convention* del 2011, per unificare i database di dominio pubblico di HRTF, fondamentali per l'analisi delle HRTF stesse. Il quarto capitolo è dedicato alle azioni svolte sui database per estendere il formato *MARL-NYU*, aggiungendo quelle informazioni mancanti, fondamentali per l'elaborazione; le considerazioni riportate in questo capitolo sono state formalizzate nella definizione di un nuovo formato di repository di HRTF, descritto nell'articolo *A standardized repository of Head-Related and Headphone Impulse Response data*

accettato alla 134<sup>th</sup> *Audio Engineering Society Convention* del 2013. Nel quinto capitolo sono esposte le metodologie utilizzate per le elaborazioni dei segnali: finestre nel tempo, filtri di sintesi, algoritmi di separazione e di ricerca dell'attacco del suono. Infine il sesto capitolo descrive le procedure operative eseguite sui dati per ottenere la separazione del contributo acustico dell'orecchio esterno e ne analizza i risultati. In appendice si elencano alcuni strumenti statistici d'appoggio all'analisi e vengono riportate le intestazioni delle funzioni che realizzano gli algoritmi sviluppati.

# Indice

<b>Abstract</b>	<b>i</b>
<b>Introduzione</b>	<b>iii</b>
<b>1 Stato dell'Arte</b>	<b>1</b>
1.1 Introduzione alla modalità . . . . .	1
1.1.1 Il suono . . . . .	1
1.1.2 Elaborazione binaurale . . . . .	2
1.1.3 Lateralizzazione . . . . .	2
1.2 Sistemi di riferimento . . . . .	3
1.2.1 Sistemi di riferimento sferici a polo doppio . . . . .	4
1.2.2 Campo vicino e campo lontano . . . . .	7
1.3 Localizzazione del suono . . . . .	7
1.3.1 Head Related Transfer Function (HRTF) . . . . .	7
1.3.2 La teoria Duplex . . . . .	9
1.3.3 Segnali forniti dall'orecchio esterno . . . . .	10
1.4 Decomposizione strutturale dell'HRTF . . . . .	11
1.4.1 La testa . . . . .	11
1.4.2 L'orecchio esterno . . . . .	12
1.4.3 Il busto e le spalle . . . . .	16
<b>2 Elaborazione binaurale</b>	<b>17</b>
2.1 HRTF derivate da modelli sferici della testa . . . . .	17
2.2 HRTF misurate empiricamente . . . . .	17
2.3 Misurazione della HRTF . . . . .	18

2.3.1	Segnali si stimolo . . . . .	18
2.3.2	Post-processing dell'HRTF . . . . .	20
2.4	Sintesi di audio spazializzato utilizzando le HRTF . . . . .	22
2.5	Problemi con la sintesi dell'audio spazializzato basata su HRTF . . . . .	23
<b>3</b>	<b>Dati e strumenti di analisi</b>	<b>25</b>
3.1	Database di dominio pubblico . . . . .	25
3.1.1	Database <i>LISTEN</i> . . . . .	26
3.1.2	Database <i>CIPIC</i> . . . . .	27
3.1.3	Database <i>FIU</i> . . . . .	28
3.1.4	Database <i>KEMAR-MIT</i> . . . . .	29
3.2	Formato <i>MARL-NYU</i> . . . . .	29
3.2.1	Struttura dei dati . . . . .	29
3.2.2	Struttura delle specifiche . . . . .	31
3.2.3	Repository . . . . .	32
<b>4</b>	<b>Azioni sui database</b>	<b>33</b>
4.1	Critica al nuovo formato <i>MARL-NYU</i> . . . . .	33
4.1.1	Modifiche suggerite al formato <i>MARL-NYU</i> . . . . .	34
4.1.2	Algoritmo per aggiungere l'informazione di onset . . . . .	34
4.2	Algoritmo per convertire nuovi soggetti nel formato <i>MARL-NYU</i> . . . . .	35
4.2.1	Il soggetto <i>KEMAR pinnaless</i> . . . . .	35
4.3	Strutture dati di ausilio . . . . .	36
4.3.1	Strutture di raccolta coordinate . . . . .	37
<b>5</b>	<b>Metodologie</b>	<b>41</b>
5.1	Calcolo degli onset . . . . .	41
5.1.1	Onset detection . . . . .	42
5.1.2	Onset detection nel dominio della frequenza . . . . .	42
5.1.3	Onset detection attraverso l'energia locale . . . . .	42
5.2	Finestratura temporale . . . . .	43
5.2.1	La finestra di <i>Hann</i> . . . . .	44
5.3	Peak filter . . . . .	45
5.3.1	Coefficienti del peak filter . . . . .	45
5.4	Partial tracking . . . . .	47



5.5	Algoritmo di separazione . . . . .	49
<b>6</b>	<b>Azioni sui dati</b>	<b>51</b>
6.1	Procedura di calcolo della HRTF . . . . .	51
6.1.1	Caricamento della HRIR . . . . .	53
6.1.2	Allineamento della HRIR all'onset . . . . .	56
6.1.3	Normalizzazione della HRIR . . . . .	57
6.1.4	Finestratura della HRIR con finestra di <i>Hann</i> . . . . .	57
6.1.5	Calcolo della HRTF usando la trasformata di Fourier . . . . .	59
6.1.6	Estrazione della HRTF secondo il range di frequenze desiderato . . . . .	59
6.2	Eliminazione del contributo acustico di busto e spalle . . . . .	61
6.2.1	Analisi della HRTF priva del contributo dell'orecchio esterno al variare della finestra temporale . . . . .	62
6.2.2	Analisi della HRTF completa al variare della finestra temporale . . . . .	62
6.2.3	Risultati sulla HRTF al variare della finestra temporale . . . . .	63
6.3	Eliminazione del contributo acustico della testa . . . . .	65
6.3.1	Eliminazione dei contributi omnidirezionali . . . . .	66
6.3.2	Compensazione del contributo acustico della conca . . . . .	67
6.4	Estrazione e analisi delle feature spettrali dell'orecchio esterno . . . . .	70
6.4.1	Estrazione delle feature spettrali dell'orecchio esterno . . . . .	71
6.4.2	Analisi delle feature spettrali dell'orecchio esterno . . . . .	71
	<b>Conclusion</b>	<b>77</b>
	<b>Sviluppi futuri</b>	<b>79</b>
	<b>Appendici</b>	<b>83</b>
<b>A</b>	<b>Strumenti statistici</b>	<b>83</b>
A.1	Root Mean Square (RMS) . . . . .	83
A.2	Box-plot . . . . .	83
<b>B</b>	<b>Funzioni sviluppate in MATLAB</b>	<b>85</b>
B.1	Funzioni di gestione del repository . . . . .	85
B.2	Funzioni di carattere generale . . . . .	86

B.3	Funzione di elaborazione dei dati . . . . .	92
-----	---	----

# Elenco delle figure

1.1	Sistema di assi cartesiani. . . . .	3
1.2	Piano mediano (a), piano orizzontale (b) e piano frontale (c). . . . .	4
1.3	Sistema interaurale polare (a) e sistema verticale polare (b). . . . .	5
1.4	Anatomia dell'orecchio esterno. . . . .	14
1.5	I sei modi di risonanza di Shaw. . . . .	15
2.1	Diagramma a blocchi di un sistema di misurazione di HRTF. . . . .	19
2.2	Diagramma a blocchi di un semplice sistema di sintesi di audio spazializzato basato su HRTF. . . . .	23
3.1	Struttura del formato MARL-NYU. . . . .	30
4.1	Vettore list_database_by_subject per l'associazione soggetto-database. . . . .	36
4.2	Tabella info_angle1: primo livello della table look-up. . . . .	38
4.3	Tabella info_angle2: secondo livello della table look-up. . . . .	38
4.4	Tabella info_angle2_DBptr (sulla sinistra) relazionata alla tabella info_angle2 (sulla destra). . . . .	39
5.1	Modellizzazione ADSR dell'involuppo spettrale. . . . .	41
5.2	Finestra di Hann. . . . .	44
5.3	Risposta in frequenza di un peak filter con $G = 8$ dB, $f_c = 500$ Hz e $f_b = 400$ Hz. . . . .	46
5.4	Diagramma di flusso dell'algorithmo di separazione. . . . .	50
6.1	Workflow del lavoro di tesi . . . . .	52
6.2	Risposta impulsiva all'orecchio sinistro per $\theta = -45^\circ$ e $\varphi = 0^\circ$ ; le linee verticali in nero evidenziano la lunghezza dell'ITD . . . . .	54
6.3	Risposta impulsiva all'orecchio destro per $\theta = -45^\circ$ e $\varphi = 0^\circ$ ; le linee verticali in nero evidenziano la lunghezza dell'ITD. . . . .	54

6.4	Risposta impulsiva all'orecchio sinistro per $\theta = 0^\circ$ e $\varphi = 0^\circ$ . . . . .	55
6.5	Risposta impulsiva all'orecchio destro per $\theta = 0^\circ$ e $\varphi = 0^\circ$ . . . . .	55
6.6	Risposta impulsiva allineata all'onset. . . . .	56
6.7	Risposta impulsiva normalizzata. . . . .	57
6.8	Risposta impulsiva con sovrapposta la finestra di <i>Hann</i> da 0.6 ms. . . . .	58
6.9	Risposta impulsiva finestrata. . . . .	58
6.10	Risposta in frequenza del segnale finestrato. . . . .	60
6.11	Risposta in frequenza del segnale non finestrato. . . . .	60
6.12	Riflessioni dovute alle spalle. . . . .	61
6.13	HRTF del soggetto KEMAR pinnaless nel piano mediano senza finestratura (a) e con finestratura da 1 ms (b). . . . .	63
6.14	HRTF del soggetto KEMAR pinnaless nel piano mediano con finestratura da 0.6 ms (a) e da 0.3 ms (b). . . . .	64
6.15	HRTF del soggetto KEMAR nel piano mediano senza finestratura (a) e con finestratura da 1 ms (b). . . . .	65
6.16	HRTF del soggetto KEMAR nel piano mediano con finestratura da 0.6 ms (a) e da 0.3 ms (b). . . . .	66
6.17	Confronto tra la HRTF del soggetto KEMAR pinnaless (a) e l'approssimazione del contributo della testa calcolato con RMS (b). . . . .	68
6.18	Analisi del soggetto KEMAR nel piano sagittale di azimuth $\theta = 45^\circ$ all'orecchio sinistro: HRTF (a) e PRTF (b). . . . .	70
6.19	Analisi dei notch per il soggetto KEMAR nel piano sagittale di azimuth $\theta = 45^\circ$ all'orecchio sinistro: PRTF (a) e notch tracking (b). . . . .	72
6.20	Notch tracking no.1 per l'orecchio sinistro. . . . .	73
6.21	Notch tracking no.2 per l'orecchio sinistro. . . . .	73
6.22	Notch tracking no.3 per l'orecchio sinistro. . . . .	74
6.23	Riflessioni generate nell'orecchio al variare dell'azimuth: $\theta = -80^\circ$ (a), $\theta =$ $-40^\circ$ (b) e $\theta = 40^\circ$ (c). . . . .	74
6.24	Peak tracking no.1 per l'orecchio sinistro. . . . .	75
6.25	Peak tracking no.2 per l'orecchio sinistro. . . . .	76
A.1	Diagramma box-plot. . . . .	84

## **Elenco delle tabelle**

2.1	Segnali di stimolo . . . . .	19
4.1	Tabella riassuntiva sulla ridefinizione degli angoli. . . . .	37



# Capitolo 1

## Stato dell'Arte

Questo capitolo è dedicato alla descrizione dello stato dell'arte nell'ambito dell'audio 3D. Si darà una panoramica sulla natura del mondo uditivo con particolare riferimento a tutte quelle informazioni che consentono ad una persona di identificare la provenienza di un suono nello spazio.

### 1.1 Introduzione alla modalità

Un ascoltatore ha una stima immediata dello spazio uditivo, percepisce con rapidità naturalezza e accuratezza la direzione di origine degli eventi acustici. Sebbene l'acutezza spaziale è decisamente inferiore nel dominio uditivo rispetto a quello visivo, il mondo uditivo ha il vantaggio di estendersi in tutte le direzioni attorno all'ascoltatore, mentre quello visivo è ristretto alla sola regione frontale. Questa differente percezione geometrica consente al sistema uditorio di reagire prima agli eventi ambientali cosicché l'ascoltatore possa spostare l'osservazione nella direzione del suono per raffinare l'analisi spaziale attraverso il sistema visivo [6]. Naturalmente, in aggiunta al ruolo di puntamento del sistema visivo, il sistema uditivo è di per sé un sofisticato processore spaziale che permette all'ascoltatore di individuare e monitorare la posizione degli oggetti sonori nelle tre dimensioni, facilitandone l'identificazione.

#### 1.1.1 Il suono

Il suono risulta dalle rapide variazioni di pressione dell'aria causate dalle vibrazioni di un oggetto (ad esempio le vibrazioni di una corda di chitarra) o da un oggetto in movimento. Essendo onde di natura meccanica richiedono un mezzo affinché possano essere propagate (non possono

viaggiare nel vuoto). Le onde sonore possono propagarsi in modo omnidirezionale in cui la propagazione è indipendente dalla direzione; tuttavia è possibile avere anche la propagazione di onde verso una sola direzione nel caso si presentino proprietà direzionali.

La percezione del suono inizia con l'arrivo delle variazioni di pressione sonora ai nostri timpani. Tramite le azioni del timpano queste variazioni oscillanti di pressione dell'aria sono inoltrate attraverso l'orecchio medio, convertite in segnali elettrici nell'orecchio interno e infine codificate in un modello neuronale interpretabile dal cervello. Tuttavia, le variazioni di pressione sonora in arrivo alle nostre orecchie non possono essere identiche alle variazioni di pressione originariamente generate dall'oggetto vibrante. Il mezzo di propagazione interferisce con l'informazione sonora, assorbendo parte delle onde; la quantità di assorbimento è influenzata dalle caratteristiche dell'aria, quali temperatura e umidità [13], e dalla distanza che le onde devono percorrere. Oltre al mezzo anche l'ambiente può interferire, introducendo fenomeni di riflessione e assorbimento dovuti all'interazione delle onde sonore con ostacoli e pareti. Per questo motivo un tipico ambiente di ascolto adatto per le registrazioni è quello anecoico nel quale le riflessioni e le rifrazioni sono ridotte al minimo e il suono registrato giunge all'ascoltatore mediante un percorso diretto (quindi senza percorsi secondari dovuti a riflessioni). Ad alterare il suono infine contribuiscono anche le varie parti del corpo [15]; questi ultimi fenomeni saranno oggetto di studio di questa tesi.

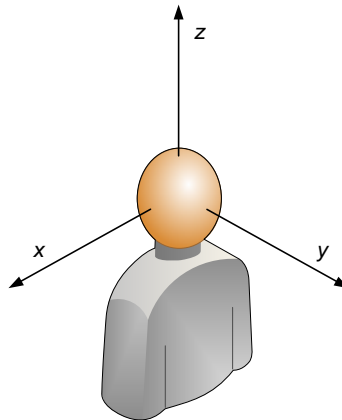
### **1.1.2 Elaborazione binaurale**

L'elaborazione binaurale si riferisce in senso lato alle funzioni sottostanti ogni capacità umana resa possibile o migliorabile mediante lo sfruttamento di due orecchie piuttosto che una sola. Tali capacità includono la localizzazione di sorgenti sonore nelle tre dimensioni, l'identificazione della voce in un ambiente rumoroso, la valutazione della loudness.

### **1.1.3 Lateralizzazione**

Una registrazione stereo ascoltata attraverso delle cuffie generalmente dà l'impressione che il suono provenga dall'interno della testa e che ci sia uno spazio definito, almeno lungo la dimensione che si estende tra le orecchie (l'asse interaurale). Così, il suono di un violino potrebbe apparire all'orecchio sinistro, mentre tra il centro della testa e l'orecchio destro si può udire la voce di un cantante. La percezione dei suoni all'interno della testa è definita *internalizzazione* del suono; mentre con il termine *lateralizzazione* si intende il diverso posizionamento del suono lungo l'asse interaurale. In opposizione si considera il caso in cui il suono è riprodotto da uno





**Figura 1.1:** *Sistema di assi cartesiani.*

o più altoparlanti: tale percezione è definita *esternalizzazione* e la descrizione delle posizioni soggettive dei suoni udibili è una procedura di *localizzazione*.

Un suono presentato identicamente ad entrambe le orecchie mediante l'uso delle cuffie (uno stimolo diotico) è solitamente identificato al centro della testa. Il suono può essere spostato verso l'orecchio destro in due modi: introducendo un ritardo temporale al segnale dell'orecchio sinistro oppure intensificando quello dell'orecchio destro.

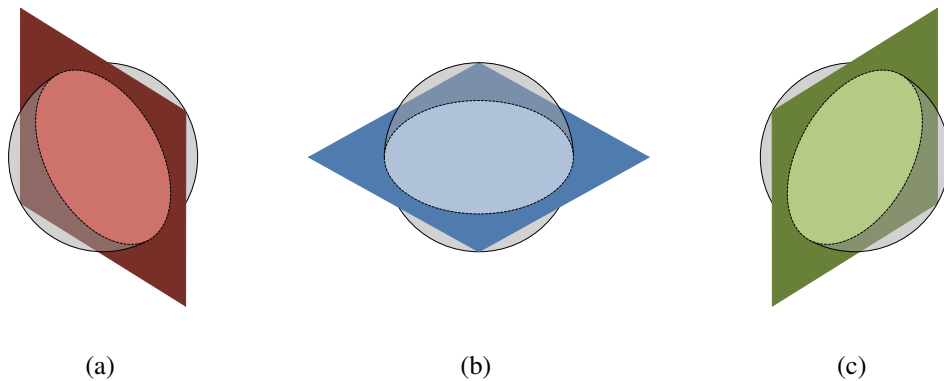
## 1.2 Sistemi di riferimento

In un ambiente di ascolto, la posizione della sorgente sonora è data relativamente ad un punto di riferimento. In un sistema a singolo utente, tipicamente l'ascoltatore è scelto come punto di riferimento e le posizioni delle sorgenti sonore sono date in relazione ad esso. Esistono diversi sistemi di coordinate. Nel *sistema rettangolare a testa centrata* il centro della testa definisce l'origine, con i tre assi definiti come segue (figura 1.1):

- l'asse  $x$  (o asse interaurale) attraversa le orecchie ed assume valori positivi a destra dell'origine;
- l'asse  $y$  esce frontalmente e proprio di fronte alla testa assume valori positivi;
- l'asse  $z$  (o asse verticale) attraversa verticalmente la testa ed assume valori positivi al di sopra dell'origine.

Con tale sistema di coordinate i tre assi formano tre piani:

- il piano  $yz$  è il piano (sagittale) mediano che divide la testa in due metà (sinistra e destra) simmetriche (figura 1.2 (a));
- il piano  $xy$  è il piano orizzontale che è allo stesso livello delle orecchie dell'ascoltatore (figura 1.2 (b));
- il piano  $xz$  è il piano frontale (figura 1.2 (c)).



**Figura 1.2:** Piano mediano (a), piano orizzontale (b) e piano frontale (c).

In generale tutti i piani paralleli al piano mediano sono definiti sagittali <sup>1</sup>. Anziché specificare le singole componenti degli assi  $x$ ,  $y$  e  $z$ , si può utilizzare un sistema di coordinate sferiche, in cui le coordinate sono specificate attraverso due angoli, azimuth ed elevazione, e il raggio.

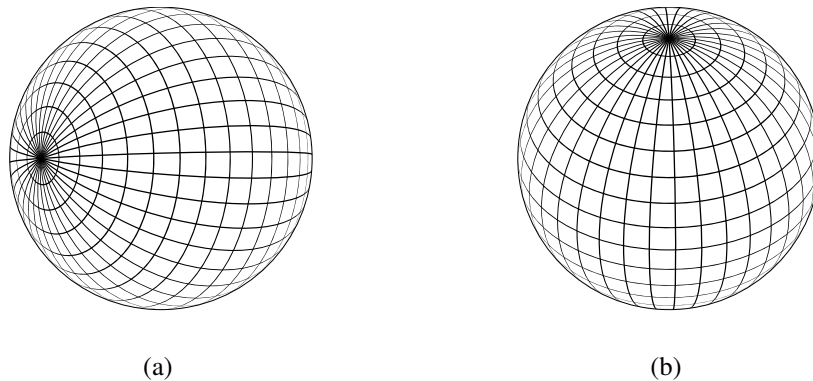
### 1.2.1 Sistemi di riferimento sferici a polo doppio

Assumendo sempre il centro della testa come origine, si possono definire due versioni di un sistema sferico a polo doppio:

- interaurale polare (figura 1.3 (a)),
- verticale polare (figura 1.3 (b)).

Il nome del sistema sferico è legato all'asse di riferimento, ovvero l'asse  $x$  nel primo caso e l'asse  $z$  nel secondo. A seconda del sistema, azimuth ed elevazione assumono definizioni e range di intervalli differenti.

<sup>1</sup>Per il piano mediano si ometterà il termine sagittale



**Figura 1.3:** *Sistema interaurale polare (a) e sistema verticale polare (b).*

### Sistema di riferimento interaurale polare

Nel sistema interaurale polare (si veda la figura 1.3):

- l'elevazione  $\varphi$  è l'angolo tra il piano orizzontale e il semipiano iniziante dall'asse interaurale, contenente la sorgente sonora ( $\varphi \in [-180^\circ, +180^\circ]$ );
- su questo semipiano, l'azimuth  $\theta$  è l'angolo tra la semiretta di intersezione con il piano mediano e la semiretta iniziante dall'origine degli assi, contenente la sorgente sonora ( $\theta \in [-90^\circ, +90^\circ]$ ).

Con tale sistema di coordinate, le superfici ad elevazione costante sono dei semipiani aventi l'asse interaurale come retta delimitatrice<sup>2</sup>, mentre le superfici ad azimuth costante sono coni concentrici sempre attorno all'asse  $x$ .

Di seguito sono elencati alcuni luoghi di punti notevoli per il sistema interaurale polare:

- $\theta = 0^\circ$ : piano mediano,
- $\varphi = 0^\circ$  e  $\varphi = -180^\circ$ : piano orizzontale,
- $\varphi = +90^\circ$  e  $\varphi = -90^\circ$ : piano frontale,
- $\theta = +90^\circ$ : semiasse interaurale positivo,
- $\theta = -90^\circ$ : semiasse interaurale negativo,
- $(\theta = 0^\circ, \varphi = 0^\circ)$ : semiasse positivo delle  $y$ ,

<sup>2</sup>Considerando più superfici ad elevazione costante si ottiene un fascio proprio di piani

- $(\theta = 0^\circ, \varphi = -180^\circ)$ : semiasse negativo delle  $y$ ,
- $(\theta = 0^\circ, \varphi = +90^\circ)$ : semiasse verticale positivo,
- $(\theta = 0^\circ, \varphi = -90^\circ)$ : semiasse verticale negativo.

### **Sistema di riferimento verticale polare**

Nel sistema verticale polare (si veda la figura 1.3):

- l'azimuth  $\theta$  è l'angolo tra il piano mediano e il semipiano iniziante dall'asse verticale, contenente la sorgente sonora ( $\theta \in [-180^\circ, +180^\circ]$ );
- su questo semipiano, l'elevazione  $\varphi$  è l'angolo tra la semiretta di intersezione con il piano orizzontale e la semiretta iniziante dall'origine degli assi, contenente la sorgente sonora ( $\varphi \in [-90^\circ, +90^\circ]$ ).

Con tale sistema di coordinate, le superfici ad azimuth costante sono dei semipiani aventi l'asse verticale come retta delimitatrice<sup>3</sup>, mentre le superfici ad elevazione costante sono coni concentrici sempre attorno all'asse  $z$ .

Di seguito sono elencati alcuni luoghi di punti notevoli per il sistema verticale polare:

- $\theta = 0^\circ$  e  $\theta = -180^\circ$ : piano mediano,
- $\varphi = 0^\circ$ : piano orizzontale,
- $\theta = +90^\circ$  e  $\theta = -90^\circ$ : piano frontale,
- $(\theta = +90^\circ, \varphi = 0^\circ)$ : semiasse interaurale positivo,
- $(\theta = -90^\circ, \varphi = 0^\circ)$ : semiasse interaurale negativo,
- $(\theta = 0^\circ, \varphi = 0^\circ)$ : semiasse positivo delle  $y$ ,
- $(\theta = -180^\circ, \varphi = 0^\circ)$ : semiasse negativo delle  $y$ ,
- $\varphi = +90^\circ$ : semiasse verticale positivo,
- $\varphi = -90^\circ$ : semiasse verticale negativo.

---

<sup>3</sup>Considerando più superfici ad azimuth costante si ottiene un fascio proprio di piani

## 1.2.2 Campo vicino e campo lontano

Uno spazio acustico privo di discontinuità od ostacoli si definisce con il termine di *campo libero* e corrisponde ad una idealizzazione delle reali condizioni in cui il suono solitamente si propaga. Il campo acustico prodotto da una sorgente in condizioni di campo libero può essere schematicamente suddiviso in due regioni:

- *campo vicino (near field)*,
- *campo lontano (far field)*.

La condizione di *campo lontano* si realizza per valori della distanza dalla sorgente approssimativamente superiori ad 1 m.

Nel *campo vicino* l'intensità sonora può avere un andamento complicato, dipendente dal tipo di sorgente e dalle dimensioni, che non necessariamente segue un andamento monotono in funzione della distanza; anche le caratteristiche di direzionalità della sorgente vanno interpretate con molta cautela.

Nella regione del *campo lontano* l'intensità sonora mostra variazioni regolari all'aumentare della distanza lineare; è così possibile definire in modo semplice alcune relazioni di propagazione.

## 1.3 Localizzazione del suono

Di seguito si introducono alcuni degli indicatori più conosciuti che le persone utilizzano per la localizzazione del suono. Si parlerà della teoria *Duplex* dalla quale sono stati definiti due indicatori di primo ordine per la determinazione del suono nel piano orizzontale, mentre si introdurranno le caratteristiche dell'orecchio esterno per la localizzazione nel piano mediano e nella risoluzione del problema della confusione fronte/retro.

### 1.3.1 Head Related Transfer Function (HRTF)

Il suono può raggiungere i timpani dell'ascoltatore mediante più percorsi diversi, diretti oppure indiretti dovuti a riflessioni. Pertanto, nell'ascolto di un suono, le orecchie percepiranno due segnali diversi  $x_L(t)$  e  $x_R(t)$  ottenuti dalla convoluzione delle risposte impulsive  $h_L(t)$  e  $h_R(t)$  sinistra e destra (*Head Related Impulse Response*, HRIR) con il segnale  $x(t)$  emesso dalla sorgente.

$$x_L(t) = \int h_L(t)x(t - \tau)d\tau \quad (1.1)$$

$$x_R(t) = \int h_R(t)x(t - \tau)d\tau \quad (1.2)$$

La funzione di trasferimento che contiene le complesse interazioni delle onde sonore con testa, spalle, busto e soprattutto l'orecchio esterno prima di raggiungere il timpano, è nota come *Head Related Transfer Function* (HRTF). Essenzialmente la HRTF tiene conto delle modifiche dei ritardi subiti dai segnali sonori nel raggiungimento delle orecchie in un modo dipendente dalla posizione che viene riconosciuto dall'ascoltatore e utilizzato come indicatore di localizzazione (la HRTF non tiene in considerazione effetti di riflessione e riverbero dovuti all'ambiente, bensì solo alle componenti del corpo).

Matematicamente le HRTF si ricavano come trasformata di Fourier delle HRIR. Fisicamente le HRTF alle orecchie sinistra e destra,  $H_L$  e  $H_R$  rispettivamente, possono essere definite come il rapporto tra il livello di pressione sonora (*Sound Pressure Level*: SPL) sul timpano di ciascun orecchio  $\Phi_L(\omega, \theta, \varphi, d)$  e  $\Phi_R(\omega, \theta, \varphi, d)$  e quello nel campo libero nella posizione corrispondente al centro della testa, ma senza testa,  $\Phi_f(\omega)$ :

$$H_L(t) = \frac{\Phi_L(\omega, \theta, \varphi, d)}{\Phi_f(\omega)} \quad (1.3)$$

$$H_R(t) = \frac{\Phi_R(\omega, \theta, \varphi, d)}{\Phi_f(\omega)} \quad (1.4)$$

In queste formulazioni  $\omega$  è la frequenza angolare,  $\theta$  e  $\varphi$  sono rispettivamente gli angoli di azimuth e di elevazione,  $d$  è la distanza tra l'ascoltatore e la sorgente sonora (misurata dal centro della testa dell'ascoltatore) [2].

Sono quindi molti i fattori che definiscono una HRTF, dalla posizione della sorgente ai contributi acustici delle varie parti del corpo menzionate. I modelli strutturali sono un valido modello per approssimare la HRTF e sono rappresentati dalla cascata dei filtri che tengono in considerazione separatamente gli effetti rispettivamente di busto e spalle, della testa e dell'orecchio esterno.

Utilizzando tale decomposizione si può comprendere il significato di una struttura anatomica nella HRTF sottraendone il contributo e osservandone la differenza [20]; per questo motivo saranno utilizzati in questo lavoro di tesi.

### 1.3.2 La teoria Duplex

La teoria Duplex formulata da Lord Rayleigh nel 1907 [19] è una teoria sulla localizzazione del suono basata su due indicatori binaurali:

- *Interaural Time Difference* (ITD)
- *Interaural Level Difference* (ILD)

Data la posizione delle orecchie, la distanza percorsa dalle onde sonore provenienti dalla sorgente sonora alle orecchie destra e sinistra dell'ascoltatore sarà diversa (a meno che la sorgente non si trovi sul piano mediano). Ciò farà sì che il suono raggiunga l'orecchio ipsilaterale prima che l'orecchio controlaterale<sup>4</sup>. La differenza nel tempo di arrivo del fronte d'onda dei segnali di pressione acustica ai timpani sinistro e destro è proprio l'ITD.

In virtù della presenza della testa, l'orecchio controlaterale tenderà ad essere in "ombra" rispetto alla sorgente sonora (a meno che questa non si trovi sul piano mediano) e quindi il livello di pressione acustica giunto ai due timpani sarà diverso. La differenza di ampiezza tra i due segnali è quindi definita come ILD, un indicatore che però dipende dalla frequenza.

Quando la sorgente sonora si trova sul piano mediano la distanza alle orecchie sinistra e destra sarà la stessa, pertanto il suono arriverà ai due timpani allo stesso istante con il medesimo livello di pressione acustica. Come risultato, sia l'ITD che l'ILD saranno prossimi allo zero. Non appena la sorgente si sposta verso destra o verso sinistra [15], l'ITD e l'ILD aumenteranno fino a raggiungere i valori massimi nel momento in cui la fonte è posta direttamente sulla destra o sulla sinistra relativamente all'ascoltatore.

#### Caratterizzazione dell'ITD

L'ITD assume un ruolo prevalente soprattutto per le frequenze basse, approssimativamente inferiori a 1500 Hz, in cui le lunghezze d'onda del suono in arrivo sono lunghe rispetto al diametro della testa e quindi la fase dei segnali che raggiungono le orecchie possono essere determinate senza ambiguità perché inalterate lungo il percorso interaurale. Per lunghezze d'onda più piccole del diametro della testa, la differenza nella distanza potrebbe essere maggiore di una singola lunghezza d'onda, portando ad una situazione ambigua (causata appunto da un problema di aliasing) dove tale differenza non corrisponde ad una posizione unica e quindi l'ITD non risulta significativo.

---

<sup>4</sup>L'aggettivo ipsilaterale si riferisce a ciò che si trova nello stesso emisfero della sorgente sonora; l'aggettivo controlaterale invece si riferisce a ciò che si trova nell'emisfero opposto

## Caratterizzazione dell'ILD

Per i suoni a bassa frequenza, in cui la lunghezza d'onda è superiore al diametro della testa, le onde non sono bloccate dalla testa, ma piuttosto le si "incurvano" attorno per raggiungere l'orecchio controlaterale (la testa non è un ostacolo acustico significativo), mantenendo praticamente inalterata l'intensità sonora. Ne risulta un ILD per questi suoni a bassa frequenza molto piccolo. Tuttavia, per i suoni ad alta frequenza (superiore a circa 1500 Hz), in cui le lunghezze d'onda sono più piccole della testa, le onde non riescono a piegare intorno ad essa per raggiungere l'orecchio controlaterale e sono quindi bloccate dalla stessa che provoca un vero e proprio effetto di "ombra acustica". Come risultato si ha una diminuzione dell'energia del suono all'orecchio più distante con conseguente aumento di valore dell'ILD che così assume maggiore significatività per i suoni ad alta frequenza.

Nonostante la semplicità e il successo della teoria Duplex siano attraenti, questa spiega solo la percezione al variare dell'azimuth, quindi su spostamenti orizzontali [4]. Se si cerca di applicare la teoria Duplex per la stima della posizione di un suono nello spazio libero, dove è possibile che il suono varii in altezza e distanza, ITD e ILD non specificano una posizione spaziale unica in quanto vi sono un numero infinito di posizioni lungo le curve di uguale distanza dalla testa dell'osservatore che hanno stesso ITD e/o stesso ILD associati. Questa ambiguità definisce il "cono di confusione" come il luogo di tutti i punti che condividono gli stessi ITD e/o ILD; esso ha come vertice il centro della testa ed è centrato nell'asse interaurale.

### 1.3.3 Segnali forniti dall'orecchio esterno

La teoria Duplex pertanto da sola non è in grado di spiegare tutti i fenomeni concernenti la localizzazione del suono; ad essa si sono aggiunte quindi le teorie di Batteau [3] che hanno posto l'attenzione sul contributo dell'orecchio esterno<sup>5</sup> per il raffinamento della percezione uditiva, laddove ITD e ILD risultavano insufficienti.

L'importanza dell'orecchio esterno è apprezzabile osservando la differenza nella HRTF in sua presenza e in sua assenza. Utilizzando il manichino KEMAR, un modello artificiale dotato di busto e testa utilizzato nella ricerca acustica da cui è possibile rimuovere con facilità la pinna<sup>6</sup>, si possono evidenziare le caratteristiche della pinna.

L'orecchio esterno gioca un ruolo significativo nella capacità di stimare l'elevazione dei suoni [26]; inoltre esso si rivela importante anche nella risoluzione della confusione fronte/retro perché

---

<sup>5</sup>Il corrispettivo in inglese per l'orecchio esterno è *pinna*

<sup>6</sup>Il soggetto KEMAR senza pinnae è denominato *pinnaless*



a causa della sua svasatura causa delle attenuazioni ai suoni che provengono da dietro.

## 1.4 Decomposizione strutturale dell'HRTF

L'HRTF può essere decomposta nella cascata di più filtri, corrispondenti al busto, alla testa e alla pinna. Di seguito saranno descritti gli effetti delle singole parti del corpo in accordo con i principi di localizzazione del suono precedentemente introdotti [20].

### 1.4.1 La testa

La testa in acustica è intesa come un corpo rigido sul cui asse interaurale sono ancorate alla stessa altezza le due orecchie. La testa, nel caso in cui la sorgente sonora non sia nel piano mediano, agisce come un ostacolo alla propagazione del suono poiché si frappone tra la sorgente stessa e l'orecchio più lontano. In accordo con la teoria Duplex, la testa condiziona direttamente i due indicatori binaurali introdotti da Rayleigh:

- ITD: il ritardo nel tempo di arrivo all'orecchio più distante è dovuto alla maggiore distanza che l'onda deve percorrere, causata proprio dalla presenza della testa;
- ILD: l'intensità sonora all'orecchio controlaterale è attenuata dalla testa che realizza un effetto di shadowing.

Attraverso la formula di Woodworth è possibile descrivere in termini matematici l'ITD nell'ipotesi di testa sferica e sorgente sonora sufficientemente distante affinché le onde siano planari. La formula applica semplici principi di cinematica, dividendo la distanza supplementare che il suono deve percorrere per raggiungere l'orecchio più distante con la velocità del suono <sup>7</sup>:

$$ITD = \frac{a(\sin \theta) + \theta}{c} \quad (1.5)$$

dove:

- $a$  è il raggio della testa in m,
- $c$  è la velocità del suono in m/s,
- $\theta$  è l'angolo di azimuth in radianti.

---

<sup>7</sup>La velocità del suono nell'aria è di 343.2 m/s

Quando la sorgente si trova sul piano mediano ( $\theta = 0$  rad) l'ITD è nullo assumendo così il suo valore minimo, mentre se si trova sul piano frontale ( $\theta = \pi/2$  rad) l'ITD assume il suo valore massimo. Il calcolo dell'ITD risulta indipendente dalla frequenza; esso si dimostra però dipendente dalla frequenza nel momento in cui viene utilizzato come indicatore binaurale (1.3.2). Al contrario dell'ITD, il valore dell'ILD è fortemente influenzato dalla frequenza delle onde sonore [2]. Questa proprietà può essere osservata studiando la funzione di trasferimento di una testa ideale sferica di raggio  $a$  sottoposta ad una sorgente sonora posta ad una distanza  $r > a$  dal centro della sfera:

$$H_{sphere}(\rho, \theta_{inc}, \mu) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{+\infty} (2m+1) P_m(\cos \theta_{inc}) \frac{h_m(\mu\rho)}{h'_m(\mu)} \quad (1.6)$$

dove:

- $\mu = \omega a/c$  è la frequenza normalizzata,
- $\rho = r/a$  è la distanza normalizzata,
- $\theta_{inc}$  è l'angolo di incidenza
- $P_m$  è il polinomio di Legendre di ordine  $m$ -esimo
- $h_m$  è il funzione sferica di Hankel di ordine  $m$ -esimo

A basse frequenze la funzione di trasferimento non è dipendente dalla direzione del suono e l'ampiezza è in prossimità dell'unità per ogni angolo di incidenza. Ad alte frequenze ( $\mu > 1$ ) si fa sentire la dipendenza direzionale; nel caso di incidenza normale ( $\theta_{inc} = 0$ ) si osserva un guadagno di 6 dB, in quanto questa situazione è paragonabile ad un'onda piana che incide perpendicolarmente con una superficie piana rigida. All'aumentare dell'angolo di incidenza il guadagno diminuisce; spostando la sorgente dalla parte opposta della testa si ha un'attenuazione della pressione sonora, con la presenza di varie oscillazioni dovute alla propagazione delle onde in diverse direzioni attorno alla sfera. La risposta minima non corrisponde ad una sorgente situata esattamente dalla parte opposta ( $\theta_{inc} = \pi$ ); in un particolare punto si nota l'effetto bright spot dove tutte le onde che si propagano attorno alla sfera arrivano in fase.

### 1.4.2 L'orecchio esterno

L'orecchio esterno è la parte periferica del sistema uditivo, costituito dal padiglione auricolare. A seguire si trova il canale uditivo, approssimato come un semplice "tubo" che si estende dal

padiglione auricolare fino al timpano agendo principalmente come risonatore. La struttura fisica della pinna (la parte effettivamente visibile dell'orecchio) è costituita da una serie di sporgenze e depressioni che variano notevolmente tra gli individui. Queste tendono ad accentuare oppure a sopprimere il contenuto energetico a media e alta frequenza dello spettro acustico a determinate posizioni della sorgente sonora.

Acusticamente l'orecchio esterno agisce come un filtro la cui funzione di trasferimento dipende in generale dalla distanza e dalla direzione della sorgente sonora relativamente all'ascoltatore.

Nel "viaggio" dalla sorgente al canale uditivo, il suono può arrivare mediante due percorsi: uno diretto e uno più lungo dovuto a riflessioni della pinna. Quando i due segnali giungono in fase il risultato è un aumento dell'energia sonora (interferenza costruttiva). Invece quando il segnale ritardato non è in fase con quello diretto si generano delle interferenze distruttive che attenuano l'energia del suono: in particolare quando i due segnali sono in opposizione di fase si verifica la massima interferenza, ovvero un *notch*.

I notch spettrali generati dall'azione filtrante dell'orecchio esterno sono indicatori importanti per la localizzazione del suono. Alcuni di questi notch sono utilizzati come indicatori per valutare la posizione verticale delle sorgenti sonore [5].

Al variare dell'elevazione cambia l'effetto dell'orecchio esterno sulle riflessioni. I notch all'aumentare dell'elevazione crescono in frequenza e diventano sempre meno pronunciati. Infatti l'orecchio esterno produce meno riflessioni per i suoni che giungono dall'alto.

### **Anatomia dell'orecchio esterno**

L'orecchio esterno, a livello anatomico, è composto delle seguenti parti [26] (figura 1.4):

**Conca** (*concha*) è una depressione che si trova al centro e circonda il condotto uditivo esterno;

**elice** (*helix*) rappresenta il più esterno dei rilievi ed è formato da una piega che segue il margine della pinna;

**antelice** (*antihelix*) è una plica cartilaginea concentrica che decorre dall'elice attraverso due branche che poi si uniscono, delimitando una depressione detta fossa triangolare; la depressione tra l'antielice e l'elice si chiama scapha;

**trago** (*tragus*) è un'eminanza rettangolare posta al di sotto dell'origine dell'elice, davanti alla conca e al condotto uditivo esterno (che nasconde e protegge);

**antitrago** (*antitragus*) è posto dietro il trago;



**Figura 1.4:** *Anatomia dell'orecchio esterno.*

**lobulo** (*lobule*) la parte inferiore della pinna.

### **Modi di risonanza di Shaw**

Edgar Shaw [21] ha introdotto un modello per la pinna, sviluppato a partire da una semplice struttura geometrica raffinata sistematicamente con l'aggiunta di varie complessità. Il modello di Shaw si presenta come un buon punto di partenza per gli studi di questa tesi.

Il modello per la pinna di Shaw si può definire come un sistema acustico leggermente smorzato che pertanto è esprimibile attraverso i suoi modi normali. Attraverso la sperimentazione acustica, Shaw ha individuato una serie di modi di risonanza longitudinali e trasversali della pinna, e le direzioni preferenziali per l'eccitazione di questi modi (figura 1.5).

Il modo 1 è il modo dominante dell'orecchio e dipende soprattutto dalla profondità della conca. I modi 2 e 3 sono risonanze trasversali e possono essere meglio descritte come una coppia verticale poiché sono maggiormente eccitati ad alti angoli di elevazione. I modi 4, 5 e 6 sono anch'essi risonanze trasversali che possono essere descritte però come una tripletta orizzontale poiché sono eccitati ad elevazioni prossime allo zero.

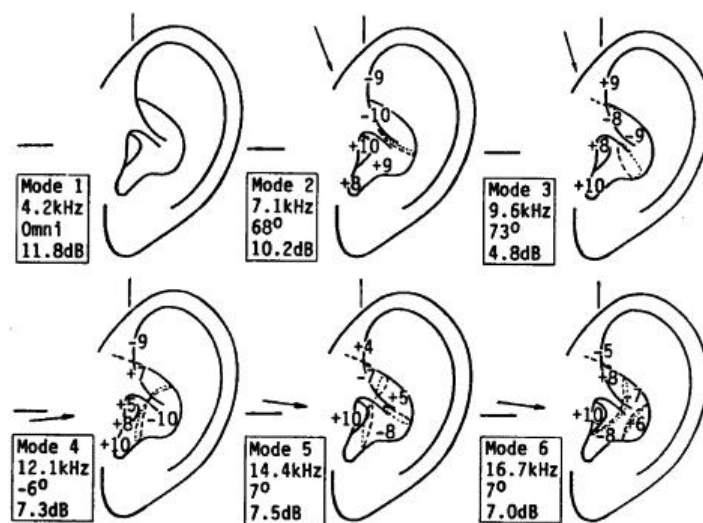


Figura 1.5: I sei modi di risonanza di Shaw.

- Il modo 1 ha un carattere omnidirezionale e si verifica attorno alla frequenza di 4,2 kHz; può essere descritto come un modo di risonanza approssimato ad un quarto di lunghezza d'onda con pressione sonora uniforme attraverso la base della conca.
- Il modo 2 si presenta a 7,1 kHz ed è maggiormente eccitato quando il suono giunge con un angolo di 68° nel piano mediano; qui si osserva una separazione tra due zone di pressione che si verificano vicino alla cymba.
- Il modo 3 appare a 9,6 kHz e, soprattutto quando è eccitato con un angolo di elevazione di 73° nel piano mediano, delinea due transizioni tra zone di pressione, c'è una zona a pressione positiva vicino alla fossa, seguita da una negativa attorno alla cymba o all'elice e nuovamente una positiva nella conca.
- Il modo 4 si verifica a 12,1 kHz, eccitato con un angolo di elevazione di -6°, e presenta tre transizioni: anche in questo caso si osserva una zona di pressione negativa vicino alla fossa che diventa positiva quando si sposta nella cymba.
- Il modo 5 appare a 14,4 kHz con un angolo di eccitazione di 7°.
- Il modo 6 si manifesta a 16,7 kHz con lo stesso angolo di eccitazione del modo 5; entrambi questi ultimi due modi sono caratterizzati dalle stesse zone di pressione e quindi da transizioni simili.

### 1.4.3 Il busto e le spalle

Il busto e le spalle influenzano le onde sonore incidenti in due aspetti principali:

- forniscono riflessioni aggiuntive che si sommano al suono diretto;
- forniscono un effetto di shadowing ai suoni provenienti dal basso.

Considerando la risposta impulsiva ad un orecchio, si osserverebbe un impulso iniziale seguito da una serie di impulsi successivi causati dalle riflessioni sulle spalle. Il ritardo tra il suono diretto e i raggi riflessi non varia molto se la posizione della sorgente sonora varia su una circonferenza nel piano orizzontale (soprattutto se il raggio è grande rispetto a quello della testa); il ritardo invece varia notevolmente se la sorgente sonora si muove verticalmente (in particolare il massimo ritardo per gli impulsi riflessi si ottiene quando la sorgente è posta esattamente sopra l'ascoltatore).

Nel dominio della frequenza le riflessioni del busto agiscono come un comb filter, introducendo dei notch periodici nello spettro. Le frequenze in cui si verificano i notch sono inversamente proporzionali ai ritardi e così si produce un pattern che varia con l'elevazione della sorgente. La frequenza del notch più basso corrisponde al ritardo più lungo. I ritardi più lunghi di un sesto di un millisecondo produrranno uno o più notch al di sotto dei 3 kHz, che è approssimativamente la frequenza più bassa in cui gli effetti dell'orecchio esterno iniziano ad essere evidenti.

Al calare dell'elevazione si raggiunge un punto di incidenza al di sotto del quale le riflessioni del busto scompaiono mentre emergono dei fenomeni di shadowing dovuti ad un vero e proprio cono d'ombra che avvolge l'orecchio più distante. Come discusso precedentemente per la testa, un ostacolo incide maggiormente per i suoni ad alta frequenza (approssimativamente superiori a 3 kHz nel caso di busto e spalle), mentre per quelli a bassa frequenza la diffrazione e lo scattering garantiscono comunque il raggiungimento dell'orecchio controlaterale.

Sebbene gli effetti acustici di busto e spalle non siano intensi come quelli introdotti dall'orecchio esterno, sono comunque importanti perché appaiono a frequenze più basse, dove i segnali sonori tipici hanno la maggior parte della loro energia e dove la risposta della pinna è essenzialmente piatta. In termini di intervalli di frequenza gli effetti previsti dal busto sono quindi complementari a quelli forniti dall'orecchio esterno.

# Capitolo 2

## Elaborazione binaurale

Al fine di valutare la struttura delle HRTF al variare dei parametri spaziali, si confrontano data set di HRTF misurati empiricamente su soggetti umani e/o artificiali in ambiente adatto alla registrazione. In questo capitolo si descriveranno la procedura di misurazione delle HRTF e la sintesi dell'audio spazializzato basata su di esse.

### 2.1 HRTF derivate da modelli sferici della testa

Il più semplice modello analitico di HRTF è derivato da un modello sferico della testa. Risolvendo l'equazione d'onda acustica corrispondente ad un'onda piana monotona incidente su una sfera rigida, è possibile calcolare la pressione risultante prodotta sulla superficie della sfera. In particolare, è possibile calcolare la pressione nei due punti sulla superficie della sfera che corrispondono all'ingresso del canale uditivo per le orecchie sinistra e destra. Valutando la pressione in questi due punti per diverse frequenze e differenti angoli di incidenza dell'onda piana, si possono calcolare sistematicamente le HRTF sinistra e destra.

### 2.2 HRTF misurate empiricamente

Le misurazioni empiriche si svolgono di solito in una camera anecoica<sup>1</sup>; qui il soggetto si colloca in un punto preciso della stanza dove udirà un segnale di stimolo prodotto da un altoparlante che viene via via posizionato a diverse angolazioni di azimuth ed elevazione. Per questi esperimenti solitamente si predispone una struttura sulla quale è fissato un altoparlante; tale struttura, avendo

---

<sup>1</sup>La camera anecoica è un ambiente di laboratorio strutturato in modo da ridurre il più possibile la riflessione dei segnali sulle pareti.

l'asse coincidente con l'asse interaurale del soggetto, può ruotare per collocarsi a diversi angoli di elevazione. Per quanto riguarda la variazione di azimuth, si può far scorrere l'altoparlante lungo il supporto o, in alternativa, si ruota direttamente il soggetto. Nel primo caso si ottiene una *spatial grid* secondo il sistema di riferimento interaurale polare, mentre nel secondo caso secondo quello verticale polare.

Quando si eseguono delle misurazioni utilizzando un manichino KEMAR, per registrare le risposte allo stimolo, si sfruttano i microfoni incapsulati all'imbocco degli ipotetici canali uditivi. Nel caso in cui il soggetto sia un individuo umano, questo viene attrezzato con dei piccoli microfoni sonda installati all'imbocco del canale uditivo di entrambe le orecchie, così da poter captare le risposte agli stimoli sonori.

## 2.3 Misurazione della HRTF

Al fine di eseguire una misurazione empirica della HRTF, si colloca il soggetto (umano o manichino) in una posizione fissa, munito dei microfoni per le registrazioni; viene quindi predisposto un semplice sistema di identificazione eseguendo uno stimolo  $s(n)$  a spettro noto attraverso un altoparlante posto a specifici azimuth, elevazione e distanza dalla testa del soggetto. Alla registrazione segue una procedura per identificare le risposte: ci sono molti metodi diversi che possono essere impiegati per eseguire il sistema di identificazione richiesto per misurare le HRTF. Poiché generalmente si assume che la HRTF sia ben modellata da sistemi lineari e tempo-invarianti, in molte elaborazioni si utilizzano appunto i sistemi LTI (*Linear and Time-Invariant*) per l'analisi di correlazione mutua <sup>2</sup>, per giungere poi alla misurazione effettiva delle HRTF. Nella figura 2.1 viene data una schematizzazione di un sistema di misurazione di HRTF.

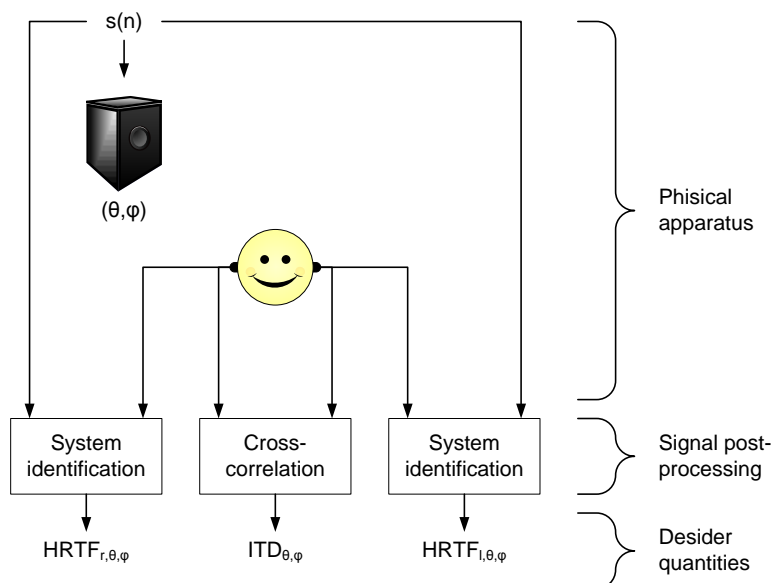
### 2.3.1 Segnali di stimolo

Alcuni dei primi studi hanno cercato di misurare direttamente le HRTF riproducendo degli impulsi attraverso gli altoparlanti e registrando la risposta alle orecchie. Altri studi hanno tentato di aumentare il rapporto segnale-rumore (*Signal to Noise Ratio*, SNR) durante la misurazione della HRTF utilizzando come stimolo un rumore gaussiano bianco avente una banda larga con energia di segnale distribuita più equamente su un periodo di tempo più lungo. Più recentemente si è passati all'utilizzo del metodo *Maximal Length Shift* (MLS) per la misurazione della HRTF; questa tecnica mantiene il vantaggio di utilizzare uno stimolo con energia distribuita più

---

<sup>2</sup>In teoria dei segnali la correlazione mutua rappresenta la misura di similitudine di due segnali come funzione di uno spostamento o traslazione temporale applicata ad uno di essi [24].





**Figura 2.1:** Diagramma a blocchi di un sistema di misurazione di HRTF.

uniformemente con un elevato SNR mentre allevia i problemi causati da uno stimolo casuale. I *Golay codes*, o codici complementari, sono stati introdotti per misurare le HRTF nel tentativo di superare i problemi che ha la tecnica di misurazione MLS a basse frequenze, pur mantenendo buoni SNR. Un'ultima categoria di segnali di stimolo è rappresentata dal *sine sweep*; consente misure più precise negli ambienti silenziosi, anche se in presenza di rumore di fondo sono preferibili MLS e Golay codes.

Segnale di stimolo	Descrizione
<i>Maximal Length Shift</i>	è un tipo di sequenza binaria pseudocasuale composta di anelli polinomiali generati usando registri a scorrimento a retroazione lineare [25]
<i>Golay codes</i>	una coppia di codici binari con la proprietà che le loro funzioni di autocorrelazione hanno lobi laterali complementari, cioè la somma delle sequenze di autocorrelazione è esattamente zero ovunque tranne che all'origine [9]
<i>Sine sweep</i>	una funzione seno nella quale la frequenza varia linearmente nel tempo, crescendo o decrescendo [23]

**Tabella 2.1:** Segnali di stimolo

### 2.3.2 Post-processing dell'HRTF

Indipendentemente dal procedimento del sistema di identificazione utilizzato per misurare le HRTF[4], le funzioni di trasferimento dell'apparecchiatura di misurazione (come il microfono e l'altoparlante) contribuiscono a caratterizzare le funzioni di trasferimento misurate e pertanto devono essere rimosse durante la procedura di post-processing dell'HRTF. Queste funzioni di trasferimento possono essere misurate con un'apparecchiatura di calibrazione e le inverse di queste funzioni possono essere utilizzate per equalizzare le misurazioni grezze dell'HRTF. Inoltre, le caratteristiche spettrali di queste misurazioni grezze, che sono uguali per tutte le posizioni, sono considerate prive di contenuto informativo significativo per l'indicazione direzionale del suono e, quindi, possono anche essere rimosse dalle HRTF originali. Queste caratteristiche spettrali comprendono la funzione di trasferimento diffusa che può essere calcolata eseguendo la media delle HRTF equalizzate su tutte le posizioni spaziali.

Seguendo una convenzione di Middlebrooks si definisce una funzione di trasferimento comune (*Common Transfer Function*, CTF) che durante la fase di post-processing viene rimossa dalle HRTF grezze per produrre la funzione di trasferimento direzionale (*Directional Transfer Function*, DTF). Quest'ultima è una funzione dell'azimuth e dell'elevazione, contenente gli indicatori spettrali responsabili dell'audio spazializzato.

Matematicamente si può descrivere il post-processing come segue. Sia  $s(n)$  il segnale di stimolo noto presentato ad azimuth  $\theta$  ed elevazione  $\phi$ . Sia poi  $c(n)$  la CTF nota mentre si definiscono  $d_{l,\theta,\phi}(n)$  e  $d_{r,\theta,\phi}(n)$  rispettivamente le DTF sconosciute alle orecchie sinistra e destra; siano  $\hat{h}_{l,\theta,\phi}(n)$  e  $\hat{h}_{r,\theta,\phi}(n)$  le HRTF grezze stimate, sinistra e destra, misurate con uno dei metodi di identificazione del sistema di cui sopra.

Così:

$$\hat{h}_{l,\theta,\phi}(n) = s(n) * c(n) * d_{l,\theta,\phi}(n) \quad (2.1)$$

$$\hat{h}_{r,\theta,\phi}(n) = s(n) * c(n) * d_{r,\theta,\phi}(n) \quad (2.2)$$

o equivalentemente, espresso nel dominio della frequenza:

$$\hat{H}_{l,\theta,\phi}(k) = S(k)C(k)D_{l,\theta,\phi}(k) \quad (2.3)$$

$$\hat{H}_{r,\theta,\phi}(k) = S(k)C(k)D_{r,\theta,\phi}(k) \quad (2.4)$$

Qui si assume che  $c(n)$  sia spazialmente invariante e possa essere calcolato dalle misure note dell'apparecchio di registrazione e dai valori di  $\hat{h}_{l,\theta,\phi}(n)$  e  $\hat{h}_{r,\theta,\phi}(n)$  per le diverse posizioni.

Quindi si calcolano le DTF come segue:

$$|D_{l,\theta,\phi}(k)| = \frac{|\hat{H}_{l,\theta,\phi}(k)|}{|S(k)||C(k)|} \quad (2.5)$$

$$|D_{r,\theta,\phi}(k)| = \frac{|\hat{H}_{r,\theta,\phi}(k)|}{|S(k)||C(k)|} \quad (2.6)$$

$$\angle D_{l,\theta,\phi}(k) = \angle \hat{H}_{l,\theta,\phi}(k) - \angle S(k) - \angle C(k) \quad (2.7)$$

$$\angle D_{r,\theta,\phi}(k) = \angle \hat{H}_{r,\theta,\phi}(k) - \angle S(k) - \angle C(k) \quad (2.8)$$

$$D_{l,\theta,\phi}(k) = |D_{l,\theta,\phi}(k)| \exp[j\angle D_{l,\theta,\phi}(k)] \quad (2.9)$$

$$D_{r,\theta,\phi}(k) = |D_{r,\theta,\phi}(k)| \exp[j\angle D_{r,\theta,\phi}(k)] \quad (2.10)$$

$$D_{l,\theta,\phi}(n) = \mathcal{F}[d_{l,\theta,\phi}(k)] \quad (2.11)$$

$$D_{r,\theta,\phi}(n) = \mathcal{F}[d_{r,\theta,\phi}(k)] \quad (2.12)$$

$$d_{l,\theta,\phi}(n) = \mathcal{F}^{-1}[D_{l,\theta,\phi}(k)] \quad (2.13)$$

$$d_{r,\theta,\phi}(n) = \mathcal{F}^{-1}[D_{r,\theta,\phi}(k)] \quad (2.14)$$

L'informazione di fase ottenuta dalle DTF calcolate nel dominio del tempo può essere utilizzata per ricavare l'ITD associato, con azimuth  $\theta$  ed elevazione  $\phi$ . Nello specifico, il valore di ITD,  $n_{ITD,\theta,\phi}$ , si ottiene come il ritardo per cui la funzione di cross-correlazione tra  $d_{l,\theta,\phi}(n)$  e  $d_{r,\theta,\phi}(n)$  è massimizzata:

$$n_{ITD,\theta,\phi} = \arg \max_{\tau} \sum_n d_{l,\theta,\phi}(n) d_{r,\theta,\phi}(n + \tau) \quad (2.15)$$

La versione a fase minima delle DTF può essere calcolata applicando una finestra al real cepstrum di  $d_{l,\theta,\phi}(n)$  e  $d_{r,\theta,\phi}(n)$ . Si definisce la finestra

$$w(n) = \begin{cases} 1 & n = 0 \text{ or } n = L/2 \\ 0 & n = 1, \dots, L/2 - 1 \\ 2 & n = L/2 + 1, \dots, L - 1 \end{cases} \quad (2.16)$$

dove  $L$  è un valore pari che rappresenta la lunghezza in numero di campioni dei segnali. Il calcolo del cepstrum di  $d_{l,\theta,\phi}(n)$  e  $d_{r,\theta,\phi}(n)$  si compone come segue:

$$c_{l,\theta,\phi}(n) = \mathcal{F}^{-1}[\log |\mathcal{F}[d_{l,\theta,\phi}(n)]|] \quad (2.17)$$

$$c_{r,\theta,\phi}(n) = \mathcal{F}^{-1}[\log |\mathcal{F}[d_{r,\theta,\phi}(n)]|] \quad (2.18)$$

$$\hat{c}_{l,\theta,\phi}(n) = c_{l,\theta,\phi}(n)w(n) \quad (2.19)$$

$$\hat{c}_{r,\theta,\phi}(n) = c_{r,\theta,\phi}(n)w(n) \quad (2.20)$$

Infine si calcola la versione nel dominio del tempo a fase minima di  $d_{l,\theta,\phi}(n)$  e  $d_{r,\theta,\phi}(n)$ :

$$d_{\min_{l,\theta,\phi}}(n) := \mathcal{F}^{-1}[\exp(\mathcal{F}[\hat{c}_{l,\theta,\phi}(n)])] \quad (2.21)$$

$$d_{\min_{r,\theta,\phi}}(n) := \mathcal{F}^{-1}[\exp(\mathcal{F}[\hat{c}_{r,\theta,\phi}(n)])] \quad (2.22)$$

## 2.4 Sintesi di audio spazializzato utilizzando le HRTF

Anche se non tutte le strutture percettivamente salienti delle HRTF misurate empiricamente sono pienamente comprese, le HRTF grezze sono già state ampiamente utilizzate per sintetizzare suoni spazializzati in cuffia. Presumibilmente le HRTF sinistra e destra e l'ITD associati a una posizione specifica caratterizzano il filtraggio acustico di un suono proveniente completamente da quella posizione. Quindi supponendo che il sistema uditivo associ queste quantità ad una particolare posizione spaziale, le HRTF e l'ITD possono essere utilizzati per filtrare un suono monoaurale in uno binaurale, che suonerà come se fosse stato originato da tale posizione.

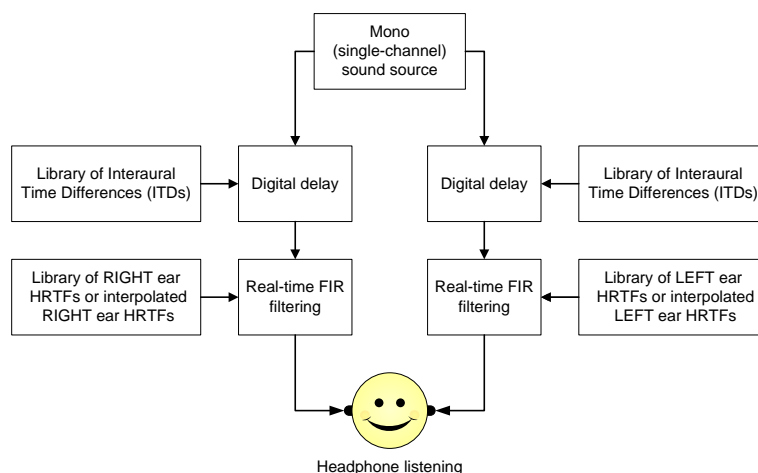
Formalmente si supponga che si voglia elaborare un segnale monofonico  $x(n)$  tale da sembrare situato ad azimuth  $\theta$  ed elevazione  $\phi$ . Nello specifico, siano  $d_{\min_{l,\theta,\phi}}(n)$  e  $d_{\min_{r,\theta,\phi}}(n)$  le risposte impulsive a fase minima misurate ad azimuth  $\theta$  ed elevazione  $\phi$ . Si costruiscono due suoni  $x_l(n)$  e  $x_r(n)$  che saranno riprodotti rispettivamente alle orecchie sinistra e destra contemporaneamente in cuffia:

$$x_l(n) = x(n - n_{ITD,\theta,\phi}) * d_{\min_{l,\theta,\phi}}(n) \quad (2.23)$$

$$x_r(n) = x(n) * d_{\min_{r,\theta,\phi}}(n) \quad (2.24)$$

dove  $n_{ITD,\theta,\phi}$  è definita negativa per i suoni che giungono prima all'orecchio sinistro.

Diverse applicazioni coinvolgono la sintesi in tempo reale di audio spazializzato in cui la sorgente sonora si sposta nel tempo. Pertanto, in pratica, dell'hardware ad alta velocità viene utilizzato per



**Figura 2.2:** Diagramma a blocchi di un semplice sistema di sintesi di audio spazializzato basato su HRTF.

implementare le circonvoluzioni che porteranno ad ottenere  $x_l(n)$  e  $x_r(n)$ . Per sintetizzare i suoni in movimento, le HRTF e l'ITD sono aggiornati dinamicamente nel tempo per corrispondere alle nuove posizioni spaziali. La figura 2.2 mostra uno schema a blocchi di un semplice sistema spazializzato in tempo reale basato su cuffia.

## 2.5 Problemi con la sintesi dell'audio spazializzato basata su HRTF

Sebbene la teoria di utilizzare le HRTF per sintetizzare l'audio spazializzato sia semplice, ci sono ancora diversi problemi che si verificano nella pratica. Per esempio, un semplice algoritmo di spazializzazione basato su HRTF, come quello mostrato in figura 2.2, non sempre produce suoni con gli effetti previsti: alcuni suoni sembrano provenire dall'interno della testa anziché da fuori, mentre i suoni trattati in modo tale da apparire frontali rispetto all'ascoltatore, in realtà sembrano provenire da dietro (la cosiddetta confusione fronte/retro). La sintesi di suoni con elevazione al di sopra o al di sotto del piano orizzontale è più difficile. Inoltre, poiché ogni individuo ha un unico insieme di HRTF, un soggetto che ascolta una riproduzione di un suono spazializzato prodotto da un insieme "generalizzato" di HRTF non può percepire il suono nella posizione spaziale prevista.

Oltre ai suoi problemi di qualità del suono, la sintesi basata su HRTF deve affrontare diversi problemi di tipo computazionale. Per esempio è spesso necessario un hardware real-time dedi-

cato per implementare anche gli algoritmi di spazializzazione più semplici. Poiché le HRTF sono tipicamente misurate su centinaia di posizioni spaziali diverse, vi è una notevole quantità di dati che devono essere immagazzinati ed elaborati velocemente. Nella sintesi sonora con sorgente in movimento (come rappresentato in Fig. xxx), sono richiesti diversi HRTF e ITD per produrre un suono in movimento di qualità. La questione si concentra quindi su come si debbano calcolare queste interpolazioni a partire da un insieme finito di HRTF e ITD esistenti.

Molti ricercatori ritengono che la soluzione a questi problemi comporti una comprensione più profonda della struttura percettiva dei dati delle HRTF. Attraverso l'analisi, la modellazione e la parametrizzazione della struttura delle HRTF, ci si aspetta di poter collegare le caratteristiche salienti delle HRTF, come i picchi e i notch, ma anche le caratteristiche temporali, a specifici parametri spaziali, come l'azimuth, l'elevazione e la distanza.

# Capitolo 3

## Dati e strumenti di analisi

Questo capitolo è dedicato alla descrizione dei dati di dominio pubblico utili per la fase di analisi delle HRTF. Si introdurranno i diversi progetti di ricerca che hanno registrato risposte impulsive su soggetti umani e manichini, sui quali sono stati portati avanti numerosi studi . A seguire si darà un approfondimento su un nuovo formato di archivio dei dati, chiamato *MARL-NYU*, atto all'unificazione e alla standardizzazione (anche se non ufficiale) dei dati raccolti dai principali enti di ricerca.

MATLAB è un linguaggio di alto livello e un ambiente interattivo per il calcolo numerico, l'analisi e la visualizzazione dei dati e la programmazione [14]. È possibile usare MATLAB in un'ampia gamma di applicazioni, tra cui l'elaborazione di segnali.

I dati relativi ad ogni singolo soggetto nel repository del MARL-NYU sono forniti nel formato .MAT, orientato all'utilizzo nell'ambiente MATLAB; assieme a questi, alcune utili funzioni ausiliarie di ricerca sono state messe a disposizione nel formato .M, tipico degli script di MATLAB. L'ambiente MATLAB si è rivelato il più adatto per quanto riguarda il caricamento di dati dal repository, le elaborazioni sui segnali e la visualizzazione dei risultati. Tutti gli script realizzati sono stati implementati in questo ambiente; i risultati sono stati forniti in formati grafici per un'analisi più intuitiva relativamente all'osservazione di soli valori numerici.

### 3.1 Database di dominio pubblico

Ogni database di HRTF di dominio pubblico è stato realizzato secondo standard diversi (per ciò che riguarda gli incrementi di azimuth ed elevazione, le aree spaziali coperte, la durata e la frequenza di campionamento, i segnali di stimolo). Il *Music and Audio Research Laboratory* della *New York University* ha definito un formato di file per la memorizzazione di HRIR, verso

il quale ha già convertito i dati relativi ai progetti di alcuni istituti di ricerca: questi progetti sono chiamati:

- *LISTEN*,
- *CIPIC*,
- *FIU*
- *KEMAR-MIT*

Di seguito saranno descritte le specifiche dei vari progetti, dalle quali si evinceranno le differenze che richiamano l'utilità e la necessità di un formato standard. Infine sarà presentato il formato di unificazione *MARL-NYU*.

### 3.1.1 Database *LISTEN*

Il progetto *LISTEN* svolto all'*Institute de Recherche et Coordination Acoustique/Musique (IR-CAM)* di Parigi [7] ha raccolto le risposte impulsive di 51 soggetti umani. Per ciascun soggetto sono state eseguite delle misurazioni su un totale di 187 posizioni. Il sistema di coordinate considerato è quello verticale polare<sup>1</sup>. L'elevazione è stata variata su 10 valori, a partire da  $-45^\circ$  fino a  $90^\circ$  con un incremento costante di  $15^\circ$ . L'azimuth invece non è stato preso con lo stesso passo per ogni angolo di elevazione; a partire dalla porzione anteriore del piano mediano (azimuth pari a  $0^\circ$ ) i valori di azimuth considerati sono i seguenti:

- ogni  $15^\circ$  per le elevazioni comprese tra  $-45^\circ$  e  $45^\circ$  (per un totale di 24 diversi valori di azimuth);
- ogni  $30^\circ$  per l'elevazione di  $60^\circ$  (per un totale di 12 diversi valori di azimuth);
- ogni  $60^\circ$  per l'elevazione di  $75^\circ$  (per un totale di 6 diversi valori di azimuth).

All'elevazione di  $90^\circ$  è stato considerato solo un valore di azimuth.

Il segnale di stimolo riprodotto per le misurazioni è il sine sweep. La frequenza di campionamento è stata fissata a 44,1 kHz; in totale sono stati registrati 512 campioni per ogni misurazione, per un tempo complessivo di circa 11.6 ms.

I dati forniti dall'istituto di ricerca sono disponibili sia nella modalità grezza che in quella compressa, in due formati diversi per entrambe: il formato WAV e quello MAT.

---

<sup>1</sup>Diversamente dalla convenzione, nel database *LISTEN* l'azimuth assume valori nell'intervallo  $[0^\circ, 360^\circ[$ ; con il valore di  $0^\circ$  si individua la porzione anteriore del piano mediano



Per ogni soggetto ci sono 187 file WAV stereo, uno per ogni posizione della sorgente. Questi file riportano le HRIR, quindi una descrizione nel dominio del tempo. Ad ogni soggetto è anche associato un file MAT contenente una struttura, capace di raccogliere insieme sia le informazioni di posizione che i dati del segnale. La struttura riporta le seguenti informazioni:

- le elevazioni in gradi (nel vettore colonna `elev_v`),
- gli azimuth in gradi (nel vettore colonna `azim_v`),
- i dati campionati (nella matrice `content_m`),
- la tipologia dei dati (nella variabile `type_s`),
- la frequenza di campionamento (nella variabile `sampling_hz`).

I vettori colonna dedicati alle coordinate spaziali hanno lunghezza pari a 187. Tale valore corrisponde anche al numero di righe della matrice contenente i campioni delle risposte impulsive; le colonne della matrice invece sono 512, ovvero la quantità di campioni registrati. Infine, le ultime due informazioni riportano un valore singolo: la frequenza di campionamento (come accennato precedentemente) è pari a 44100 Hz, mentre FIR (*Finite Impulse Response*<sup>2</sup>) è la sigla legata alla tipologia di risposta al segnale.

### 3.1.2 Database CIPIC

Il *CIPIC Interface Laboratory* della *University of California Davis* [22] ha misurato le HRTF ad alta risoluzione spaziale per 45 soggetti, rendendone i dati di pubblico dominio. Tra i soggetti compare, oltre agli individui umani, anche il manichino KEMAR in due versioni diverse:

- con pinnae grandi,
- con pinnae piccole.

Ogni soggetto è stato sottoposto a 1250 misurazioni, ciascuna delle quali corrispondente ad una diversa posizione della sorgente sonora. Le posizioni sono specificate dagli angoli di azimuth ed elevazione nel sistema di coordinate interaurale polare<sup>3</sup>, mentre la distanza è mantenuta fissa al valore di 1 m. L'elevazione è campionata uniformemente con un passo costante di 5,625°

<sup>2</sup>In teoria dei segnali FIR identifica una risposta finita all'impulso

<sup>3</sup>Diversamente dalla convenzione, nel database *CIPIC* l'elevazione assume valori nell'intervallo  $[-90^\circ, 270^\circ]$ ; con il valore di  $0^\circ$  si individua la porzione anteriore del piano orizzontale

(=  $360^\circ/64$ ) a partire da  $-45^\circ$  fino a  $230,625^\circ$  (per un totale di 50 valori); per ottenere una densità approssimativamente uniforme attorno al soggetto, l'azimuth è stato campionato a  $-80^\circ$ ,  $-65^\circ$ ,  $-55^\circ$ , da  $-45^\circ$  a  $45^\circ$  con passo costante di  $5^\circ$ , a  $55^\circ$ ,  $65^\circ$  e  $80^\circ$  (per un totale di 25 valori). I microfoni dei vari soggetti hanno registrato le risposte campionandoli alla frequenza di 44,1 kHz; come segnali di stimolo sono stati riprodotti dei Golay codes. La HRIR grezza "uscita" dai microfoni è stata modificata attraverso l'applicazione di una finestra di *Hann*<sup>4</sup> al fine di eliminare le riflessioni della stanza. La HRIR risultante è composta di 200 campioni, per una durata complessiva di circa 4,5 ms.

I dati risultanti sono stati suddivisi per soggetto; ad ognuno è dedicato un file nel formato MAT di MATLAB contenente:

- le risposte impulsive sinistra e destra (nelle matrici HRIR\_l e HRIR\_r),
- gli onset sinistro e destro<sup>5</sup> (nelle matrici OnL e OnR),
- l'ITD (nella matrice ITD).

Tutte le matrici associano la prima dimensione all'angolo di azimuth e la seconda dimensione all'angolo di elevazione. Dovendo registrare un solo valore per ogni posizione, le matrici dedicate agli onset e all'ITD si presentano bidimensionali ( $25 \times 50$ ). Alle matrici delle risposte impulsive serve una terza dimensione poiché per ogni posizione ci sono 200 valori da memorizzare: quindi queste matrici sono tridimensionali ( $25 \times 50 \times 200$ ).

### 3.1.3 Database FIU

Il *Digital Signal Processing Lab* della *Florida International University* [8] ha realizzato il proprio database raccogliendo le risposte impulsive di 15 soggetti. Le posizioni considerate della sorgente sonora sono 72, descritte attraverso il sistema di coordinate verticale polare<sup>6</sup>. Gli angoli di elevazione sono 6 (ogni  $18^\circ$  a partire da  $-36^\circ$  fino a  $54^\circ$ ) mentre i diversi valori di azimuth sono 12 (ogni  $30^\circ$  a partire da  $0^\circ$ ).

Il segnale di stimolo (dei Golay codes) captato dai microfoni è stato campionato alla frequenza di 96 kHz, collezionando 256 campioni per ogni risposta impulsiva.

---

<sup>4</sup>Nel capitolo 5 saranno descritti la finestra di *Hann* e in generale gli effetti della finestratura di un segnale audio

<sup>5</sup>L'onset corrisponde all'istante o al campione in cui inizia un suono

<sup>6</sup>Le convenzioni circa gli angoli sono le stesse del *LISTEN*

### 3.1.4 Database *KEMAR-MIT*

Il *Media Lab* del *Massachusetts Institute of Technology* [10] ha lavorato utilizzando il segnale MLS come stimolo su un manichino KEMAR nelle varianti a pinna normale e pinna grande (quindi su 2 soggetti). Il sistema di riferimento è ancora una volta quello verticale polare, con l'altoparlante collocato a 1,4 m di distanza dal manichino. Nel complesso sono state effettuate 710 registrazioni per soggetto; 14 sono le elevazioni differenti, da  $-40^\circ$  a  $90^\circ$  con un incremento verticale di  $10^\circ$ , mentre il numero di posizioni azimuthali corrispondenti varia da 72 (incrementi orizzontali di  $5^\circ$ ) a solo 1 (per l'elevazione di  $90^\circ$ ). Le risposte all'impulso risultanti sono composte di 512 campioni, registrati ad una frequenza di campionamento di 44,1 Hz.

## 3.2 Formato *MARL-NYU*

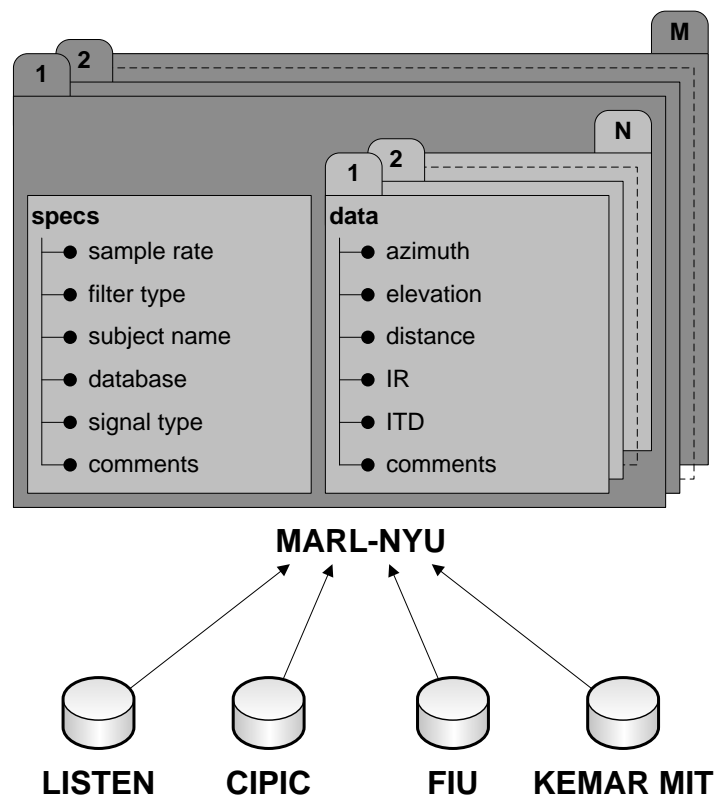
Nel formato MARL-NYU [1] tutte le misurazioni effettuate su un soggetto sono organizzate in due diverse componenti relativamente al formato MAT di MATLAB: il vettore (di strutture) dati, *data*, e la struttura delle specifiche, *specs*. *Data* contiene tutte le informazioni relative alle misurazioni, mentre *specs* contiene tutte le informazioni generali sul data set di HRTEF. Nella figura 3.1 viene schematizzata la struttura del formato *MARL-NYU*.

### 3.2.1 Struttura dei dati

Ogni elemento del vettore *data* è una struttura che racchiude in sé tutte le informazioni specifiche di una misurazione, vale a dire: gli angoli di azimuth ed elevazione, la distanza della sorgente sonora dal soggetto, le HRIR oppure le HRTF alle orecchie sinistra e destra ed il corrispondente valore di ITD. Ogni misurazione consecutiva è memorizzato in una struttura separata e viene aggiunta al vettore originale.

#### Coordinate spaziali

I campi della struttura che identificano le coordinate spaziali sono: *azimuth*, *elevation* e *distance*. L'azimuth è definito nell'intervallo da  $[-180^\circ, 180^\circ]$ ; con il valore di  $0^\circ$  si identifica la parte anteriore del piano mediano. L'elevazione è definita nell'intervallo da  $-90^\circ$  a  $90^\circ$  (estremi inclusi); con il valore di  $0^\circ$  si identifica il piano orizzontale. La terza coordinata è la distanza tra il soggetto e la sorgente sonora ed è espressa in metri. In ogni struttura del vettore *data* le tre coordinate spaziali identificano univocamente la posizione della sorgente sonora rispetto all'ascoltatore.



**Figura 3.1:** *Struttura del formato MARL-NYU.*

## HRIR e HRTF

I campioni sono contenuti nel campo *IR* della struttura che è organizzato come una matrice  $N \times 2$  dove  $N$  è la lunghezza del filtro ovvero il numero di campioni. La prima colonna è dedicata alla risposta dell'orecchio sinistro, mentre la seconda a quello destro. Nel campo *IR*, anziché l'HRIR, può anche essere memorizzata l'equivalente nel dominio della frequenza, cioè l'HRTF.

## ITD

Il valore dell'ITD è espresso in numero di campioni (nell'omonimo campo della struttura). Il segno rappresenta una discriminante fondamentale, in quanto:

- se  $ITD < 0$  allora il suono proviene dalla sinistra;
- se  $ITD > 0$  allora il suono proviene dalla destra.

Per convenzione, se l'ITD è già stato precedentemente incorporato nel segnale nel campo *IR* il campo *ITD* viene impostato di default a 0.

## Commenti

È presente anche un campo *comments* per eventuali note aggiuntive riguardanti la misurazione in questione.

### 3.2.2 Struttura delle specifiche

Tutte le informazioni che sono proprie di un data set di misurazioni vengono memorizzate in una struttura separata chiamata *specs*. È molto probabile che queste non cambino durante il processo di misurazione. La struttura contiene i seguenti campi:

*sample rate* è la frequenza di campionamento espressa in Hz;

*filter type* indica se il segnale campionato è nella sua forma originale (*Fixed Filters*) oppure se sono state eliminate tutte le informazioni di fase (*Minimum Phase*);

*subject name* è il nome del soggetto relativo al dataset originale;

*database* rappresenta il data set di origine (può essere uno tra *LISTEN*, *CIPIC*, *FIU* e *KEMAR-MIT*);

*signal type* è il segnale di stimolo utilizzato nelle misurazioni (può essere *Sine Sweep*, *MLS* o *Golay Codes*);

*comments* è un campo per eventuali note aggiuntive riguardanti il soggetto [1].

### 3.2.3 Repository

Il repository colleziona nel complesso 113 soggetti, provenienti dai 4 database di dominio pubblico. Ad ciascun soggetto è associato un file MAT contenente il vettore *data* e la struttura *specs*. Nel passaggio di conversione dai database originali al formato *MARL-NYU*, è stata rielaborata l'indicizzazione dei soggetti, i quali riportano un numero identificativo secondo lo schema seguente:

- da 001 a 051 per i soggetti originati dal database *LISTEN*;
- da 052 a 096 per i soggetti originati dal database *CIPIC*;
- da 097 a 111 per i soggetti originati dal database *FIU*;
- da 112 a 113 per i soggetti originati dal database *KEMAR-MIT* [1].

# Capitolo 4

## Azioni sui database

La prima parte del lavoro di tesi è stata dedicata allo studio di un nuovo repository. Questa fase si è resa necessaria dopo un'analisi delle strutture che contengono i dati dei vari soggetti che ha evidenziato alcune lacune nella reperibilità di alcune informazioni determinanti. Nel capitolo sono riportate le azioni svolte sul repository; dalla modifica della struttura del formato *MARL-NYU* all'aggiunta di nuovi soggetti.

### 4.1 Critica al nuovo formato *MARL-NYU*

Nel corso del lavoro di tesi l'analisi è stata spesso ostacolata dall'assenza di alcune informazioni essenziali omesse nel formato *MARL-NYU*. In particolare ciò riguarda:

- l'informazione sugli onset,
- la tipologia di coordinate utilizzate nel database originale.

L'onset risulta importante per l'allineamento delle risposte impulsive all'istante di attacco <sup>1</sup>: benchè sia comunque possibile ricavare gli onset a partire dai campioni delle HRIR, l'assenza di un campo ad essi riservato nella struttura dei dati ne impedisce la memorizzazione; nel caso dei soggetti del database CIPIC, tale informazione viene già fornita nel repository originale, ma l'organizzazione attuale del formato *MARL-NYU* non ne consente l'utilizzo.

Le risposte dei soggetti nei piani mediano e orizzontale sono state maggiormente analizzate. Tuttavia, uno degli obiettivi di questa tesi è quello di estendere le conoscenze note nei piani

---

<sup>1</sup>L'attacco è la prima fase nella modellazione ADSR (*Attack, Delay, Sustain, Release*) dell'evoluzione nel tempo dell'ampiezza del suono

mediano e orizzontale agli altri piani sagittali: tale analisi è supportata dal sistema di coordinate interaurale polare. Ecco che la conoscenza del sistema di riferimento impiegato si rivela fondamentale per avviare l'analisi nella direzione desiderata. L'indicazione sulla tipologia del sistema di riferimento adottata è un'altra importante informazione assente nel formato MARL-NYU. Dalla descrizione degli intervalli di azimuth ed elevazione non emerge come interpretare le due coordinate angolari, dal momento che coesistono nel repository soggetti misurati sia nel sistema interaurale polare che in quello verticale polare.

#### **4.1.1 Modifiche suggerite al formato MARL-NYU**

Per rendere più esaustivo il formato MARL-NYU sono state apportate delle modifiche sia alla struttura *data* che a quella *specs*.

Nella struttura *data* è stato aggiunto un campo *onset* implementato come un vettore  $2 \times 1$  (affinché sia concorde con il campo *IR*), dove il primo valore è dedicato all'orecchio sinistro mentre il secondo a quello destro. L'unità di misura deve naturalmente essere concorde con l'ITD (istante di tempo o numero di campioni). In caso di dato non disponibile si può ripiegare su un numero negativo di default (in quanto l'onset è un valore sempre positivo).

Nella struttura *specs* è stato aggiunto un campo *coordinate system* nel quale inserire il sistema di riferimento impiegato per le registrazioni sul soggetto in questione.

Tali modifiche sono state apportate in locale ai file nel formato MARL-NYU, riscontrando piena compatibilità con i dati preesistenti e un effettivo vantaggio nell'avere a disposizione informazioni supplementari (senza doverle recuperare da altre fonti).

Considerando anche altri database di HRIR di dominio pubblico, si osserva che alcuni di essi forniscono le risposte impulsive senza compensazione, quindi in un formato "grezzo" (è il caso del database LISTEN). Per tale motivo si vuole ampliare il formato MARL-NYU affinché possa contenere sia dati di natura grezza (*raw*) che compensati (*compensated*). A livello implementativo tale adattamento si realizza replicando il vettore *data* e la struttura *specs* per la sezione *raw* con l'eccezione del campo dedicato al tipo di filtro utilizzato per la compensazione.

#### **4.1.2 Algoritmo per aggiungere l'informazione di onset**

Una volta predisposto il campo nella struttura *data* per contenere gli onset sinistro e destro, bisogna provvedere a riempirlo; purtroppo non tutti i database forniscono tale informazione insieme alle risposte impulsive. Quindi l'algoritmo deve sviluppare due diverse procedure:



- se l'onset è fornito nel database originale, copiarlo nell'omonimo campo della struttura *data*,
- altrimenti eseguire un algoritmo di *onset detection*<sup>2</sup> per ricavare gli onset direttamente dalle HRIR.

Tra i database attualmente presenti nel formato *MARL-NYU* solo il *CIPIC* mette a disposizione gli onset.

## 4.2 Algoritmo per convertire nuovi soggetti nel formato *MARL-NYU*

Il repository attualmente contiene dati relativi a soggetti appartenenti a quattro database differenti; l'idea di base del *MARL* è di unificare sotto un unico formato i vari database di dominio pubblico. Pertanto il repository allo stato attuale è destinato ad essere ampliato attraverso l'aggiunta di nuovi soggetti.

A tal proposito è stato implementato uno script MATLAB che, date le risposte impulsive di un soggetto e le relative specifiche (come angoli di azimuth ed elevazione, frequenza di campionamento etc), restituisca un file .MAT nel formato *MARL-NYU* relativo al soggetto stesso. L'algoritmo deve conoscere il database di origine poiché ogni database ha adottato un proprio standard di memorizzazione dei dati; in base al nome del database di origine, l'algoritmo eseguirà la sezione di codice corrispondente al database stesso. Una volta nota la politica di archiviazione, la conversione dei dati nel formato *MARL-NYU* da parte dell'algoritmo si traduce in una copia dei dati contenuti nei file del database originale in un nuovo file realizzato secondo le direttive del formato unificato. Concorde con le modifiche progettuali alla struttura del formato *MARL-NYU*, l'algoritmo provvede anche a memorizzare l'onset e il tipo di sistema di riferimento.

### 4.2.1 Il soggetto KEMAR *pinnaless*

Il *CIPIC Interface Laboratory* oltre ai 45 soggetti di dominio pubblico, che sono stati inseriti nel database *MARL-NYU*, ne ha registrati altri, tra i quali compare un manichino KEMAR nella versione senza pinnae (termine inglese: *pinnaless*). Questo soggetto, i cui dati sono memorizzati sempre secondo le specifiche del database del *CIPIC*, è risultato utile al lavoro di analisi e pertanto è stato aggiunto al repository locale, tradotto nel nuovo formato.

---

<sup>2</sup>Nella sezione 5.1.1 sono descritte alcune tecniche per la ricerca degli onset

	Database
1	LISTEN
2	LISTEN
...	...
51	LISTEN
52	CIPIC
...	...
97	FIU
...	...
112	KEMAR-MIT
113	KEMAR-MIT
114	CIPIC

} Soggetti di dominio pubblico

**Figura 4.1:** Vettore *list\_database\_by\_subject* per l'associazione soggetto-database.

Seguendo la numerazione dei soggetti preesistenti, il KEMAR *pinnaless* è stato registrato, in locale, come soggetto 114.

### 4.3 Strutture dati di ausilio

Ogni file di dati si riferisce ad un singolo soggetto. Le ricerche più comuni (ad esempio l'elenco di tutti i soggetti originati da uno stesso database) richiedono pertanto il caricamento di tutti i file (anche quelli non richiesti) a cui segue l'interrogazione di uno specifico campo (il database di origine secondo l'esempio precedente); tale procedura naturalmente si presenta poco efficiente perché si sprecano tempo e memoria per caricare soggetti che non sono richiesti. Per ovviare a questa e ad altre problematiche circa l'ottimizzazione nell'uso delle risorse, è stato predisposto un file di configurazione (chiamato *config.mat*) contenente tutte quelle informazioni di utilizzo comune atte a semplificare la scrittura degli algoritmi e soprattutto ad ottimizzarli. Nel caso specifico dell'associazione soggetto-database è stato creato un vettore (chiamato *list\_database*), al quale si accede attraverso l'indice del soggetto (nella numerazione del *MARL-NYU*), contenente appunto il nome del database di origine (figura 4.1).

sistema	$angolo_1$	$angolo_2$
interaurale polare	azimuth	elevazione
verticale polare	elevazione	azimuth

**Tabella 4.1:** Tabella riassuntiva sulla ridefinizione degli angoli.

### 4.3.1 Strutture di raccolta coordinate

In fase di analisi si può essere interessati ad osservare le HRTF relative solo ad un certo intervallo di angolazioni; quindi non è richiesta la scansione completa del vettore *data* e pertanto se ne vuole leggere solamente la porzione di interesse. Purtroppo, provenendo da convenzioni diverse, le HRIR dei vari soggetti oltre a non presentare misurazioni per le stesse angolazioni, non sono neppure ordinate con il medesimo criterio all'interno del vettore (taluni soggetti sono ordinati per elevazione prima e per azimuth poi, altri al contrario). Quindi, non conoscendo il criterio di ordinamento delle misurazioni, la ricerca di una gamma precisa di angolazioni deve avvenire in maniera sequenziale con conseguente aumento dei tempi di elaborazione.

Per maggior precisione, i sistemi di riferimento e gli angoli di azimuth ed elevazione scelti per le misurazioni, più che dal soggetto, dipendono dal database di origine e altrettanto si può affermare per l'ordinamento delle misurazioni: pertanto le risposte impulsive di tutti i soggetti di uno stesso database sono ordinate allo stesso modo (ad esempio per piani orizzontali da  $-90^\circ$  a  $90^\circ$ , scorrendo poi l'azimuth del piano da  $-180^\circ$  a  $180^\circ$ ).

Sempre mirando all'ottimizzazione degli script si è eseguita una fase di pre-processing nella quale sono stati catalogati tutti i possibili angoli di azimuth ed elevazione, raggruppati per database di origine. Prima di descrivere il sistema di organizzazione delle coordinate si è pensato di ridefinire gli angoli affinché si potessero generalizzare gli algoritmi senza dover discriminare il sistema di riferimento: si parlerà quindi di  $angolo_1$  ed  $angolo_2$ . L' $angolo_1$  è legato alla determinazione del piano (indipendentemente dal fatto che esso sia sagittale o orizzontale), mentre l' $angolo_2$  è la coordinata che permette di scorrere all'interno del piano. La tabella 4.1 mostra come gli angoli vengano ridefiniti in base al sistema di riferimento. Con questa soluzione l' $angolo_1$  assumerà valori nell'intervallo  $[-90^\circ, 90^\circ]$ , mentre l' $angolo_2$  nell'intervallo  $[-180^\circ, 180^\circ]$ .

La struttura di raccolta delle coordinate è una *table look-up* a due livelli: nel primo si accede mediante il nome del database di origine, mentre nel secondo si entra con la coppia database- $angolo_1$ . Con il primo livello si ottengono tutti gli angoli dei piani (sagittali nell'interaurale polare, orizzontali nel verticale polare) per il database richiesto (figura 4.2); dal secondo livello, invece, risultano tutti i valori possibili dell' $angolo_2$  a parità di database e di  $angolo_1$  specificati

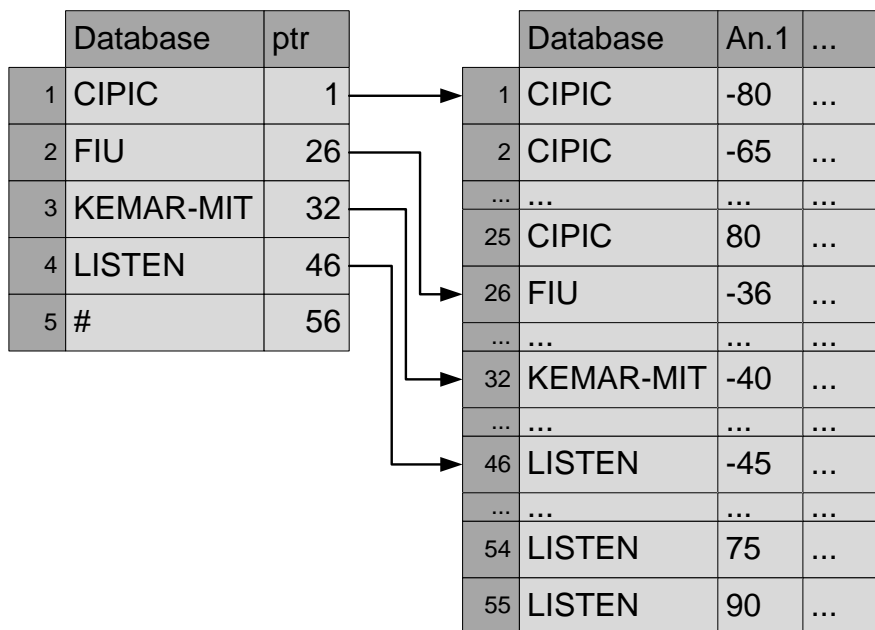
	Database	Angle 1
1	CIPIC	1 x 25: [-80,-65,...,-5,0,5,...,65,80]
2	FIU	1 x 6: [-36,-18,0,18,36,54]
3	KEMAR-MIT	1 x 14: [-40,-30,...,-10,0,10,...,80,90]
4	LISTEN	1 x 10: [-45,-30,-15,0,15,...,75,90]

**Figura 4.2:** *Tabella info\_angle1: primo livello della table look-up.*

	Database	An.1	Angle 2
			Data location
1	CIPIC	-80	1 x 50: [-50.625,-45,...,0,...,229.375,235]
			1 x 50: [1,26,...,201,...,1225,1250]
2	CIPIC	-65	1 x 50: [-50.625,-45,...,0,...,229.375,235]
			1 x 50: [2,27,...,202,...,1224,1249]
...	...	...	...
54	LISTEN	75	1 x 6: [-180,-120,-60,0,60,120]
			1 x 6: [184,183,182,181,186,185]
55	LISTEN	90	1 x 1: [0]
			1 x 1: [187]

**Figura 4.3:** *Tabella info\_angle2: secondo livello della table look-up.*

(figura 4.3). Il secondo livello della tabella è ordinato prima per database e poi per l'*angolo*<sub>1</sub>; per aumentare ulteriormente l'efficienza nella lettura della tabella è possibile avvalersi di un puntatore che indica la prima entry relativa al database richiesto (figura 4.4). Nel secondo livello della tabella sono contenuti soprattutto gli indici da utilizzare per accedere direttamente al vettore *data*.



**Figura 4.4:** Tabella *info\_angle2\_DBptr* (sulla sinistra) relazionata alla tabella *info\_angle2* (sulla destra).



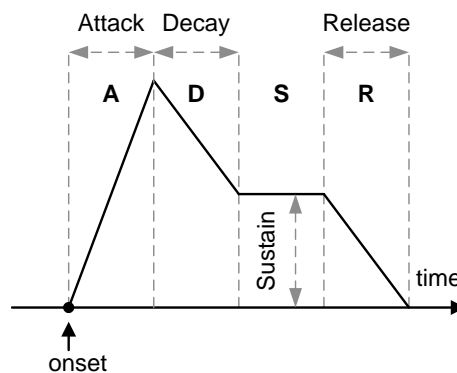
# Capitolo 5

## Metodologie

In questo capitolo saranno presentate le metodologie utilizzate nella fase di analisi, descrivendo le caratteristiche principali relativamente allo scopo per cui sono state impiegate.

### 5.1 Calcolo degli onset

L'onset si riferisce all'istante in cui un evento ha inizio, quindi caratterizza l'avvio della fase di attacco nella modellizzazione ADSR (*Attack, Delay, Sustain, Release*) (figura 5.1). Più precisamente è il tempo che l'intensità sonora impiega per passare da 0 al suo valore massimo; considerando una risposta impulsiva, nell'intervallo che precede l'onset il segnale si trova in un transitorio iniziale silenzioso.



**Figura 5.1:** Modellizzazione ADSR dell'involucro spettrale.

### 5.1.1 Onset detection

Il rilevamento e la localizzazione degli onset è molto utile nelle tecniche di analisi per i segnali musicali [12] in quanto fornisce l'informazione sull'inizio effettivo di un suono (ad esempio una nota). Il metodo usuale per rilevare gli onset ricerca le regioni transitorie nel segnale; a queste zone seguono improvvisi aumenti di energia, cambiamenti nel breve tempo dello spettro del segnale o nelle proprietà statistiche. Le tecniche di onset detection si attuano sia nel dominio del tempo che nel dominio della frequenza.

### 5.1.2 Onset detection nel dominio della frequenza

Nel dominio spettrale, gli aumenti di energia legati ai transitori tendono ad apparire come un evento a banda larga. Poiché l'energia del segnale di solito è concentrata alle basse frequenze dove maschera l'effetto dei transitori, i cambiamenti dovuti proprio ai transitori sono più evidenti alle alte frequenze. Si definisce una funzione HFC (*High Frequency Content*) definita per ogni frame <sup>1</sup> come:

$$D_H[j] = \sum_k k |X_j[k]| \quad (5.1)$$

dove  $|X_j[k]|$  è il modulo dello spettro del frame  $j$ -esimo. La HFC enfatizza le alte frequenze e mostra le ampiezze maggiori durante l'attacco del suono. Questa tecnica si presta bene all'identificazione degli onset nei suoni di tipo percussivo.

### 5.1.3 Onset detection attraverso l'energia locale

I metodi nel dominio del tempo sono basati sul calcolo di una funzione differenziale del primo ordine dell'involuppo di ampiezza del segnale; si considera quindi il massimo gradiente di salita come onset. L'approccio comune utilizza come funzione di detection la derivata nel tempo dell'energia del segnale:

$$D(t) = \frac{dE(t)}{dt} \quad (5.2)$$

---

<sup>1</sup>Il frame è un segmento del segnale nel tempo di lunghezza predefinita



cosicché improvvisi incrementi di energia siano trasformati in picchi molto stretti della derivata.

## 5.2 Finestratura temporale

La finestratura è una semplice moltiplicazione nel dominio del tempo (che diventa una convoluzione nel dominio della frequenza) per una funzione finestra:

$$h_{win}(n) = h(n)w(n) \quad (5.3)$$

La finestra ha un'ampiezza in numero di campioni pari ad  $N$  e, dato l'intervallo  $[0, N - 1]$ , assume precisi valori al suo interno mentre è nulla al di fuori. Esistono molte diverse tipologie di finestre, le quali si differenziano nella definizione all'interno dell'intervallo  $[0, N - 1]$ .

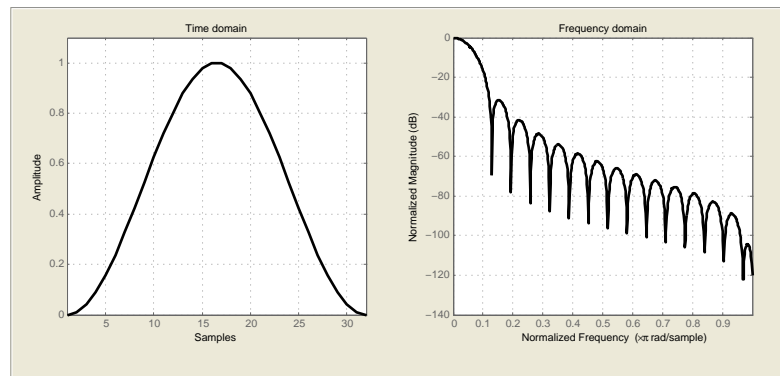
$$w(n) = \begin{cases} w_0(n), & \text{se } n \in [0, N - 1] \\ 0, & \text{se } n \notin [0, N - 1] \end{cases} \quad (5.4)$$

dove  $w_0(t)$  caratterizza la finestra all'interno dell'intervallo.

La finestratura causa nel dominio della frequenza una distorsione, detta dispersione (*leakage*) [18], inversamente proporzionale alla durata della finestra. L'andamento nel tempo della finestra determina l'ampiezza dei lobi laterali della dispersione visibili nel dominio della frequenza.

La dispersione permette di avere risposte in frequenza più uniformi e smussate. Tale forma è necessaria affinché si possano rilevare i punti notevoli più significativi, quali flessi, massimi o minimi. Con i dati grezzi ci sarebbero troppi punti di flesso che a loro volta comporterebbero molti poli e zeri inutili.

Purtroppo l'applicazione di una finestra ha anche lo spiacevole effetto di spostare drasticamente gli zeri; al contrario, i poli mantengono la loro posizione originale.



**Figura 5.2:** *Finestra di Hann.*

### 5.2.1 La finestra di *Hann*

La finestra di *Hann*<sup>2</sup>, insieme alla finestra di *Hamming*<sup>3</sup>, appartiene alla famiglia delle finestre a “coseno rialzato”; tale nome deriva dalla definizione della finestra nell’intervallo  $[0, N - 1]$ :

$$w_0(n) = \frac{1}{2} \left( 1 - \cos \left( \frac{2\pi n}{N - 1} \right) \right) \quad (5.5)$$

Dal punto di vista della frequenza, la finestra di *Hann* presenta un lobo centrale largo  $4\Delta f$  (dove  $\Delta f = 1/T$ ,  $T$  è il periodo di campionamento) e dei lobi laterali (larghi  $\Delta f$ ) più bassi. Nella figura 5.2 viene rappresentata la finestra di *Hann* sia nel dominio del tempo che nel dominio della frequenza.

La finestra di *Hann* è usata tipicamente nell’elaborazione di segnali digitali per selezionare un sottoinsieme di campioni al fine di eseguire una trasformata di Fourier o altri calcoli. Il vantaggio nell’utilizzo della finestra di *Hann* si riscontra nel poco aliasing a fronte di una sola leggera perdita di risoluzione (che a livello spettrale corrisponde ad un allargamento del lobo principale).

<sup>2</sup>La finestra di *Hann* è stata definita da Julius von Hann

<sup>3</sup>La finestra di *Hamming* è stata definita da Richard Wesley Hamming

## 5.3 Peak filter

I peak filter sono speciali *weighting filter*<sup>4</sup>, basati sui classici filtri passa-basso, passa-alto, passa-banda, elimina-banda e su un percorso diretto [27]. A differenza dei filtri a banda limitativa come i passa-banda e gli elimina-banda, i peak filter sono usati per potenziare (*boosting*) o tagliare (*cutting*) alcune frequenze desiderate. Un peak filter nel dominio continuo può essere ottenuto dalla connessione parallela di una traiettoria diretta ed filtro passa-banda secondo la formulazione:

$$H(s) = 1 + H_{BP}(s) \quad (5.6)$$

dove  $H_{BP}(s)$  è la funzione di trasferimento del filtro passa-banda.

Con l'aiuto di una funzione di trasferimento del secondo ordine per il filtro passa-banda

$$H_{BP}(s) = \frac{(H_0/Q_\infty) s}{s^2 + \frac{1}{Q_\infty s+1}} \quad (5.7)$$

la funzione di trasferimento del peak filter diventa

$$H(s) = 1 + H_{BP}(s) = \frac{s^2 + \frac{1+H_0}{Q_\infty s+1}}{s^2 + \frac{1}{Q_\infty s+1}} = \frac{s^2 + \frac{V_0}{Q_\infty s+1}}{s^2 + \frac{1}{Q_\infty s+1}} \quad (5.8)$$

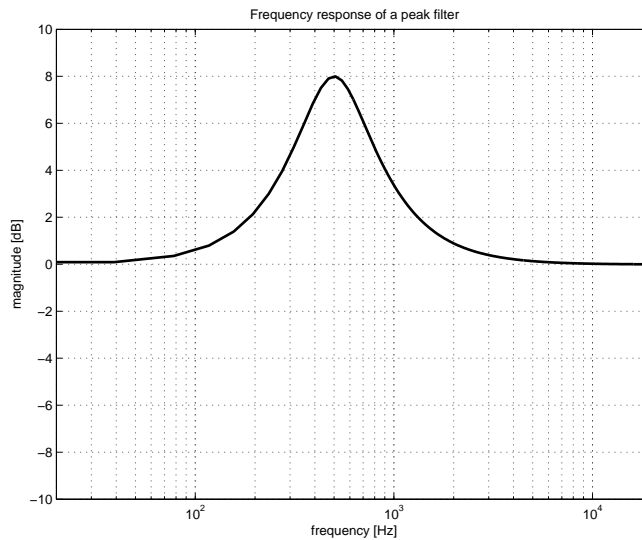
dove  $Q_\infty$  è il fattore di qualità<sup>5</sup>,  $H_0$  è il guadagno e  $V_0 = H_0 + 1$  è un parametro che determina il valore della funzione di trasferimento  $H(s)$  al tendere all'infinito della sua parte immaginaria  $\omega$  ( $s = \sigma + j\omega$ ). La massima ampiezza nella risposta in frequenza, in corrispondenza della frequenza centrale  $f_c$ , è determinata dal parametro  $V_0$ ; la relativa ampiezza di banda è fissata dal fattore  $Q_\infty$ . La simmetria geometrica della risposta in frequenza in relazione alla centrale rimane costante per  $H(s)$ .

### 5.3.1 Coefficienti del peak filter

Per implementare un filtro digitale, il filtro progettato nel dominio  $S$  con funzione di trasferimento  $H(s)$  deve essere convertito nel dominio  $Z$  con l'aiuto di una trasformazione atta ad ottenere

<sup>4</sup>I *weighting filter* sono filtri che tendono ad enfatizzare o reprimere alcuni aspetti o alcuni fenomeni di un segnale

<sup>5</sup>Il fattore di qualità è un parametro adimensionale che descrive quanto sottosmorzato è un sistema



**Figura 5.3:** Risposta in frequenza di un peak filter con  $G = 8 \text{ dB}$ ,  $f_c = 500 \text{ Hz}$  e  $f_b = 400 \text{ Hz}$ .

la funzione di trasferimento  $H(z)$ . La mappatura avviene grazie alla seguente trasformazione bilineare:

$$s = \frac{2}{T} \frac{z - 1}{z + 1} \quad (5.9)$$

dove  $T$  è il periodo di campionamento (ricavato come  $T = 1/f_s$ ,  $f_s$  è la frequenza di campionamento). La funzione di trasferimento del secondo ordine nel dominio  $Z$  assume la forma

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}} \quad (5.10)$$

La struttura dei filtri parametrici (tra i quali il peak filter) consente un accesso diretto ai parametri della funzione di trasferimento, come frequenza centrale o di cutoff, ampiezza di banda e guadagno. Per definire i parametri si introduce una variabile di ausilio  $K = \tan(\omega_c T/2)$  (dove  $\omega_c = 2\pi f_c$  è la frequenza angolare centrale); il guadagno  $V_0$  si calcola a partire dal guadagno  $G$  espresso in decibel.

Nel caso di peak filter per il potenziamento (*boosting*) il guadagno si definisce  $V_0 = 10^{G/20}$  mentre i parametri di  $H(z)$  sono i seguenti:

$$\begin{aligned}
 b_0 &= \frac{1 + \frac{V_0}{Q_\infty} K + K^2}{1 + \frac{1}{Q_\infty} K + K^2} \quad , \quad b_1 = \frac{2(K^2 - 1)}{1 + \frac{1}{Q_\infty} K + K^2} \quad , \quad b_2 = \frac{1 - \frac{V_0}{Q_\infty} K + K^2}{1 + \frac{1}{Q_\infty} K + K^2} \\
 a_1 &= \frac{2(K^2 - 1)}{1 + \frac{1}{Q_\infty} K + K^2} \quad , \quad a_2 = \frac{1 - \frac{1}{Q_\infty} K + K^2}{1 + \frac{1}{Q_\infty} K + K^2}
 \end{aligned}
 \tag{5.11}$$

Nel caso di peak filter per il taglio (*cutting*) il guadagno si definisce  $V_0 = 10^{-G/20}$  mentre i parametri di  $H(z)$  sono i seguenti:

$$\begin{aligned}
 b_0 &= \frac{1 + \frac{1}{Q_\infty} K + K^2}{1 + \frac{V_0}{Q_\infty} K + K^2} \quad , \quad b_1 = \frac{2(K^2 - 1)}{1 + \frac{V_0}{Q_\infty} K + K^2} \quad , \quad b_2 = \frac{1 - \frac{1}{Q_\infty} K + K^2}{1 + \frac{V_0}{Q_\infty} K + K^2} \\
 a_1 &= \frac{2(K^2 - 1)}{1 + \frac{V_0}{Q_\infty} K + K^2} \quad , \quad a_2 = \frac{1 - \frac{V_0}{Q_\infty} K + K^2}{1 + \frac{V_0}{Q_\infty} K + K^2}
 \end{aligned}
 \tag{5.12}$$

## 5.4 Partial tracking

Le fasi di *peak detection* e *notch detection* selezionano rispettivamente i picchi massimi e minimi che corrispondono a sinusoidi stazionarie presenti in un dato frame. Il *partial tracking* raggruppa i picchi attraverso i frame consecutivi, creando le tracce per le parziali <sup>6</sup>. Si ottengono così le tracce dei picchi di risonanza e quelle dei notch. McAulay e Quatieri hanno proposto un algoritmo di *partial tracking* per l'analisi e la sintesi del parlato nella rappresentazione sinusoidale [17].

L'analisi di un segnale viene effettuata per frame consecutivi; per ognuno di essi vengono rilevati i picchi a diversi valori di frequenza. Se il numero dei picchi fosse costante da frame a frame, l'associazione di un picco con il corrispondente del frame successivo si ridurrebbe ad un ordinamento in frequenza dei picchi. In pratica i picchi compaiono e scompaiono a causa dei complicati fenomeni acustici presenti nel segnale.

L'algoritmo di tracking riceve in ingresso i picchi selezionati per ogni frame. Nel generico frame  $k$  sono stati selezionati  $N$  picchi con frequenze  $f_0^k, f_1^k, \dots, f_{N-1}^k$  mentre nel frame  $k + 1$  ne sono stati selezionati  $M$  (in generale  $N \neq M$ ) con frequenze  $f_0^{k+1}, f_1^{k+1}, \dots, f_{M-1}^{k+1}$ . La procedura che consente di associare il picco  $n$  (con  $0 \leq n \leq N$ ) del frame  $k$  con il picco  $m$  (con  $0 \leq m < M$ )

<sup>6</sup>Per suoni acuti l'energia spettrale si concentra a poche frequenze discrete (linee spettrali) lentamente temporali; queste corrispondono a diverse componenti sinusoidali chiamate parziali

del frame  $k + 1$  si suddivide in tre passi.

Nel primo passo si supponga che sia stato trovato un match per le frequenze  $f_0^k, f_1^k, \dots, f_{n-1}^k$  e si cerca di trovarlo anche per  $f_n^k$ . Si definisce l'intervallo di corrispondenza  $\Delta$  per  $f_n^k$  come l'intervallo in cui ricercare le frequenze del frame  $k + 1$  da associare a  $f_n^k$ . Se tutte le frequenze del frame  $k + 1$  stanno al di fuori dell'intervallo di corrispondenza  $\Delta$  per  $f_n^k$ :

$$|f_n^k - f_m^{k+1}| \geq \Delta \text{ per ogni } m \quad (5.13)$$

allora la track associata alla frequenza  $f_n^k$  viene dichiarata morta: le viene assegnata ampiezza nulla nel frame  $k + 1$  e non viene più considerata nella prosecuzione dell'algoritmo. Questo primo passo viene ripetuto per tutte le frequenze, continuando da  $f_{n+1}^k$ . Se esiste  $f_m^{k+1}$  nel frame  $k + 1$  che sta all'interno di  $\Delta$  ed è la frequenza più vicina a  $f_n^k$ :

$$|f_n^k - f_m^{k+1}| < |f_n^k - f_i^{k+1}| < \Delta \text{ per ogni } i \neq m \quad (5.14)$$

si otterrà un match che si potrà dire definitivo solamente se risulterà il migliore alla fine dell'esecuzione dei tre passi.

Nel secondo passo un tentativo ha associato  $f_n^k$  con  $f_m^{k+1}$ ; se  $f_m^{k+1}$  non ottiene candidati migliori quello che gli è stato associato viene considerato definitivo. È il caso in cui

$$|f_m^{k+1} - f_n^k| < |f_m^{k+1} - f_{i+1}^k| < \Delta \text{ per ogni } i > n \quad (5.15)$$

e quindi l'analisi per  $f_n^k$  e  $f_m^{k+1}$  si conclude per proseguire con  $f_{n+1}^k$ . Se invece questa condizione non è verificata,  $f_m^{k+1}$  ha un match migliore con  $f_{n+1}^k$  e si creano due sottocasi. Nel primo sottocaso la frequenza minore adiacente a  $f_m^{k+1}$  (vale a dire  $f_{m-1}^{k+1}$ ) non rientra nell'intervallo di corrispondenza di  $f_n^k$ ; allora la traccia associata a  $f_n^k$  viene dichiarata morta all'entrata del frame  $k+1$ . Nel secondo sottocaso la frequenza  $f_{m-1}^{k+1}$  giace all'interno dell'intervallo di corrispondenza di  $f_n^k$  e viene definito il match. In entrambi i casi viene ripetuto il primo passo usando  $f_{n+1}^k$  come frequenza da analizzare.

Nel terzo passo, quando tutti i picchi del frame  $k$  sono stati assegnati a tracce vive o morte, nel frame  $k + 1$  potrebbero rimanere delle frequenze libere. Ipotizzando che  $f_m^{k+1}$  sia una di queste, si può dire che è nata una nuova possibile parziale nel frame  $k$  di ampiezza zero.

## 5.5 Algoritmo di separazione

Al fine di isolare i notch spettrali nelle PRTF si sfrutta un algoritmo che restituisce una stima delle componenti risonante e riflettente separate. L'idea è quella di compensare iterativamente lo spettro del modulo della PRTF con un filtro multi-notch fino a che non si rilevano più notch significativi. Una volta raggiunta la convergenza (all'iterazione  $i$ ), la  $H_{res}^{(i)}$  conterrà la componente risonante, mentre la combinazione  $H_{refl}^{(i)}$  dei filtri multi-notch fornirà la componente riflettente. Le condizioni iniziali dell'algoritmo influenzano pesantemente il risultato finale; devono essere scelti tre parametri:

- $N_{ceps}$ , il numero di coefficienti cepstrali usati per stimare l'involuppo spettrale della PRTF ad ogni iterazione;
- $D_{min}$ , la soglia minima di profondità affinché uno notch sia considerato tale;
- $\rho$ , il fattore di riduzione della banda passante dei notch che compongono il filtro multi-notch.

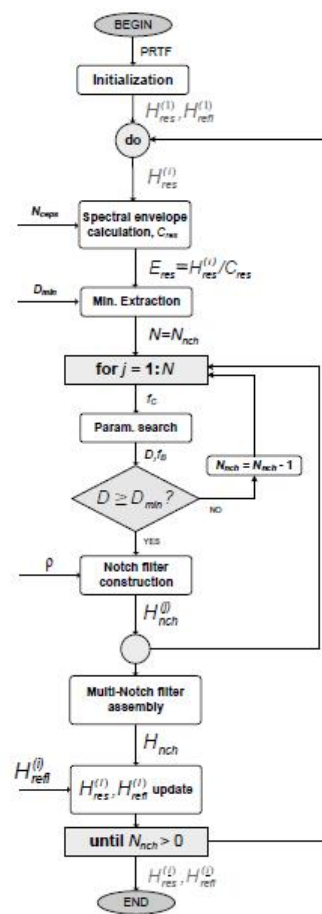
Nella fase di inizializzazione  $H_{res}^{(i)}$  viene uguagliata alla PRTF e  $H_{refl}^{(i)}$  viene impostata a 1.

L'algoritmo poi attraverso una struttura iterativa ripete i seguenti passi:

- ricerca di un nuovo notch nel segnale residuo (che alla prima iterazione coinciderà con la PRTF)
- calcolo del residuo (ottenuto sottraendo l'ultimo notch trovato)
- costruzione del filtro multi-notch

Il ciclo si conclude nel momento in cui non ci sono più notch da estrarre; a questo punto il segnale residuo conterrà solamente le risonanze, mentre nel filtro multi-notch si troverà traccia di tutti i notch trovati. Nella figura 5.4 è illustrato il diagramma di flusso dell'algoritmo.

L'algoritmo infine richiama la procedura di *partial tracking* per costruire le tracce dei notch e dei picchi di risonanza.



**Figura 5.4:** Diagramma di flusso dell'algoritmo di separazione.



# Capitolo 6

## Azioni sui dati

L'attività di standardizzazione è stata necessaria per la fase di analisi avviata con l'introduzione del nuovo formato, il lavoro di questa tesi è stato svolto sui file tradotti nel formato *MARL-NYU*; ciò consente naturalmente di poter usare gli stessi script prodotti in MATLAB a prescindere dal soggetto studiato, ovvero dal database di origine. Per le problematiche accennate nella sezione 4.1 la versione dei file utilizzata è quella modificata, con l'aggiunta degli onset e del tipo di sistema di riferimento.

Nella figura 6.1 vengono schematizzate le fasi del lavoro di tesi:

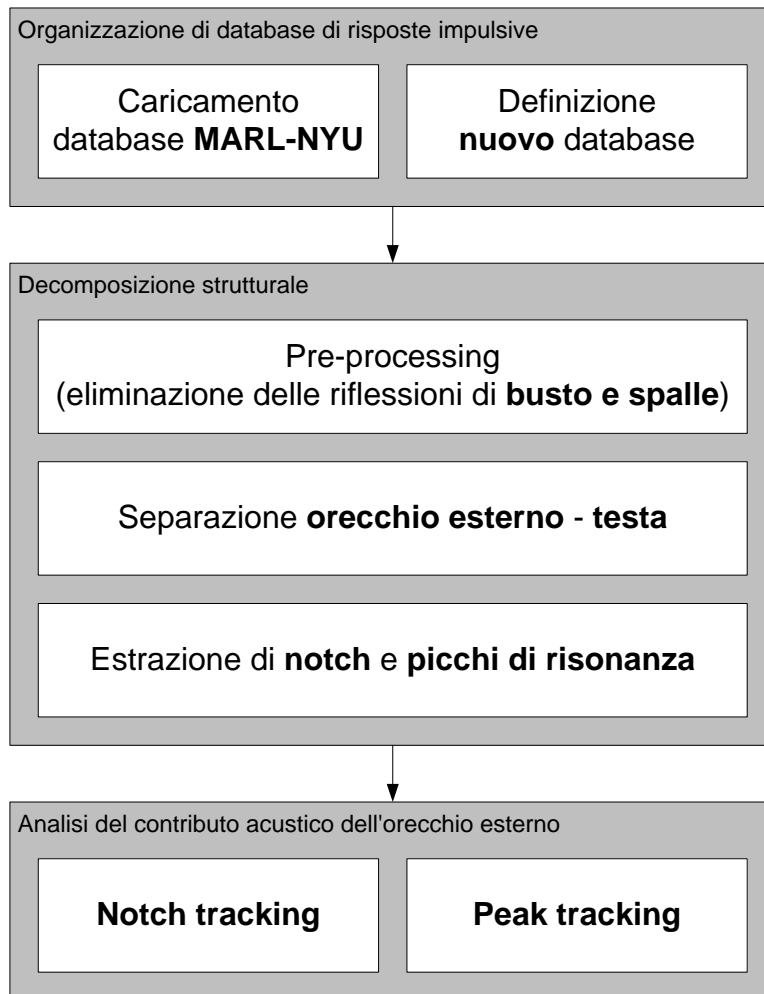
**Organizzazione di database di risposte impulsive** in questa fase sono state svolte le azioni descritte nel capitolo 4 riguardanti lo studio del formato *MARL-NYU* e la definizione di uno standard da utilizzare per riorganizzare la struttura del repository di risposte impulsive;

**Decomposizione strutturale** considerando la composizione strutturale della HRTF, in questa fase passo dopo passo è stata isolata la componente che descrive il contributo dell'orecchio esterno da quelle relative a busto, spalle e testa; una volta ricavata la PRTEF, sono state estratte le feature caratterizzanti, ovvero i notch e i picchi di risonanza;

**Analisi del contributo acustico dell'orecchio esterno** in questa fase, a seguito del *partial tracking*, sono stati analizzati i comportamenti di notch e picchi di risonanza al variare delle coordinate angolari.

### 6.1 Procedura di calcolo della HRTF

Dai file MAT contenenti tutti i dati relativi ad un singolo soggetto si deve calcolare la HRTF a partire dalle risposte impulsive (HRIR). La procedura è la seguente:



**Figura 6.1:** *Workflow del lavoro di tesi*

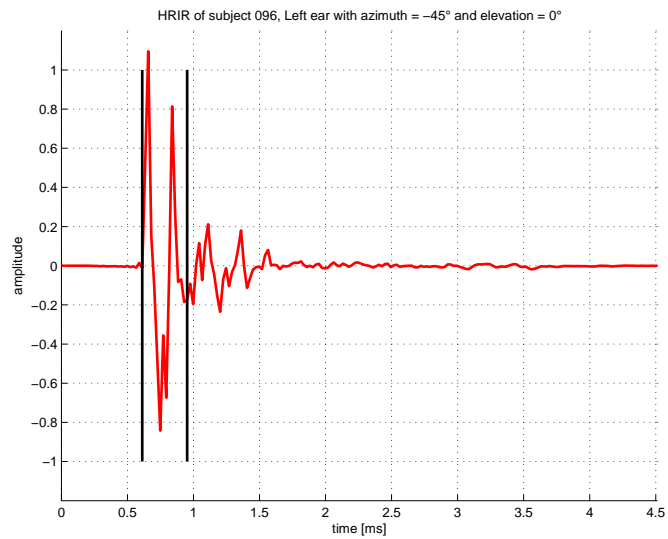
- caricamento della HRIR,
- allineamento della HRIR all'onset,
- normalizzazione della HRIR,
- finestrazione della HRIR con finestra di *Hann*,
- calcolo della HRTF usando la trasformata di Fourier,
- estrazione della HRTF secondo il range di frequenze desiderato.

### 6.1.1 Caricamento della HRIR

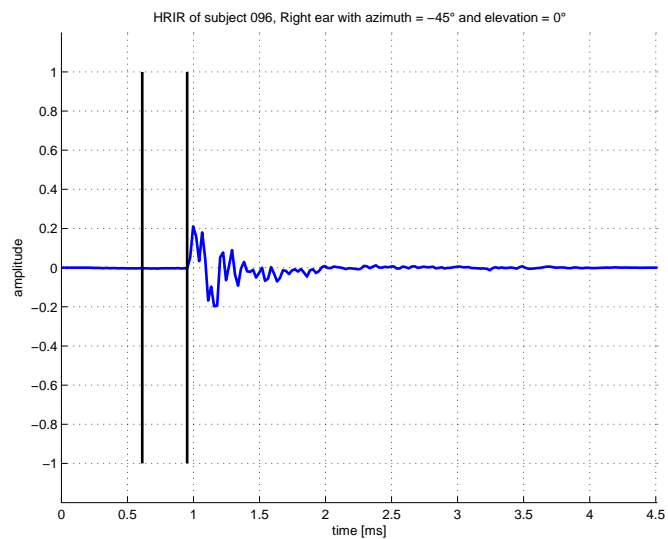
Per ogni soggetto esistono diverse HRIR registrate, le quali differiscono le une dalle altre per la posizione della sorgente sonora. Tale posizione è composta della terna azimuth  $\theta$ , elevazione  $\varphi$ , distanza  $r$  e discrimina appunto in modo univoco le HRIR di uno stesso soggetto. Grazie alle strutture di dati ausiliarie, noti il database di origine, l'azimuth e l'elevazione, si ricava direttamente l'indice del vettore *data* da dove è possibile caricare la HRIR corrispondente, interrogando il campo *IR*.

In questo segnale è possibile osservare l'istante in cui l'impulso sonoro raggiunge il microfono adagiato nell'orecchio del soggetto: tale istante di tempo è l'onset, da cui ha inizio la fase di attacco del suono che culmina con il valore di picco. Nell'intervallo che precede l'onset si ha un transitorio iniziale silenzioso caratterizzato da una leggera oscillazione del segnale attorno allo zero. Il segnale campionato memorizzato nel database è sufficientemente lungo da contenere l'intera risposta all'impulso di stimolo: questa, a seguito del raggiungimento del valore di picco (dove è concentrata la maggior parte dell'energia), è caratterizzata da oscillazioni che vanno via via attenuandosi.

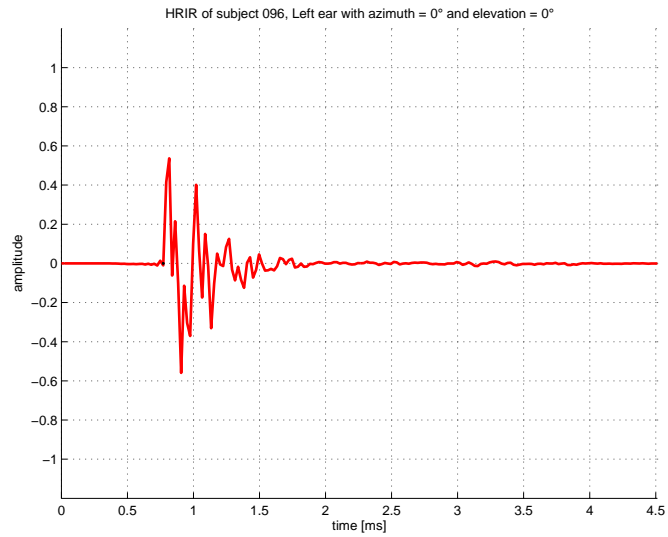
Nelle figure 6.2 e 6.3 sono mostrate le HRIR relative rispettivamente all'orecchio sinistro e a quello destro. In questo esempio la posizione della sorgente sonora corrisponde ad azimuth  $\theta = -45^\circ$  ed elevazione  $\varphi = 0^\circ$  (nella convenzione del sistema interaurale polare): l'orecchio sinistro sarà quindi quello ipsilaterale, mentre l'orecchio destro sarà quello controlaterale. Confrontando le due risposte si nota come il suono giunga prima all'orecchio ipsilaterale. Il ritardo riconducibile all'orecchio controlaterale è proprio l'ITD. A titolo di esempio nelle figure 6.4 e 6.5 si illustrano le risposte impulsive registrate con la sorgente sonora posta nel piano mediano (equidistante dalle due orecchie) ad un'elevazione  $\varphi = 0^\circ$ : in questa condizione l'ITD è approssimativamente nullo e gli onset delle HRIR sinistra e destra sono pressoché uguali.



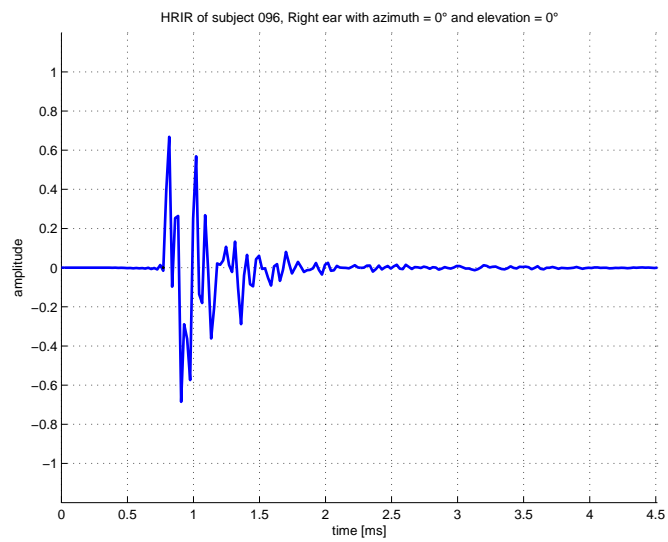
**Figura 6.2:** Risposta impulsiva all'orecchio sinistro per  $\theta = -45^\circ$  e  $\varphi = 0^\circ$ ; le linee verticali in nero evidenziano la lunghezza dell'ITD



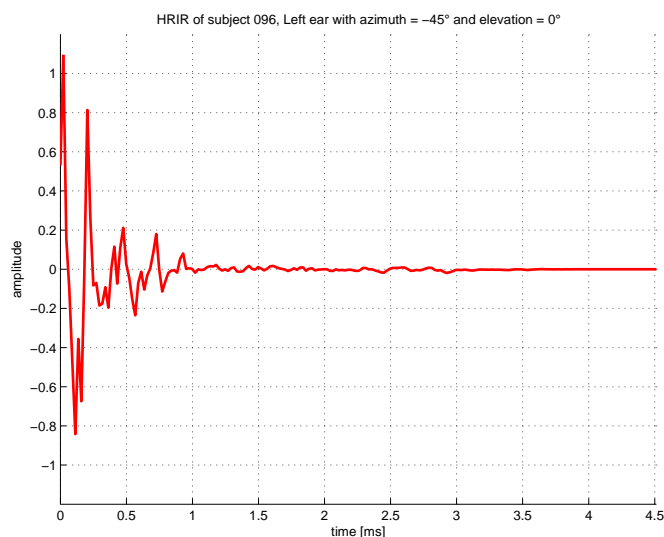
**Figura 6.3:** Risposta impulsiva all'orecchio destro per  $\theta = -45^\circ$  e  $\varphi = 0^\circ$ ; le linee verticali in nero evidenziano la lunghezza dell'ITD.



**Figura 6.4:** Risposta impulsiva all'orecchio sinistro per  $\theta = 0^\circ$  e  $\varphi = 0^\circ$ .



**Figura 6.5:** Risposta impulsiva all'orecchio destro per  $\theta = 0^\circ$  e  $\varphi = 0^\circ$ .

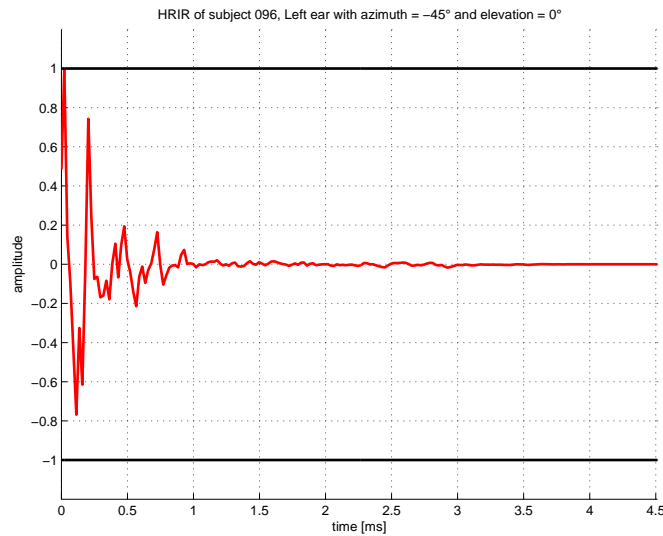


**Figura 6.6:** *Risposta impulsiva allineata all'onset.*

### 6.1.2 Allineamento della HRIR all'onset

La parte silenziosa della HRIR non porta informazione circa l'ampiezza del segnale; pertanto una volta caricata la HRIR, si ricava anche il valore di onset relativo all'orecchio in questione e si procede ad allinearla a tale valore: così facendo, il segnale risultante inizia con la fase di attacco. Al termine di questa procedura è preferibile che la "nuova" HRIR abbia la medesima lunghezza, in termini di numero di campioni, dell'originale; si procede quindi ad aggiungere in coda tanti campioni nulli quanti ne precedevano la fase l'attacco (figura 6.6). Tale operazione serve ad assicurare che le HRIR abbiano la stessa lunghezza qualora le si dovessero confrontare.

Essendo le HRIR dei segnali campionati contenenti le risposte ad un singolo stimolo sonoro, esiste una semplice procedura alternativa ai metodi di onset detection descritti nella sezione 5.1.1. Questa può essere impiegata per ricavare l'informazione di onset in quei soggetti che non la presentavano nella loro versione originale. La formulazione di questa procedura è stata ricavata dall'analisi degli onset noti (nei soggetti *CIPIC*) e prevede la ricerca dello zero precedente il picco del segnale. Poiché le HRIR sono dei segnali campionati è probabile che non esista il campione corrispondente a tale punto; pertanto l'algoritmo ricerca i due campioni che individuano il cambio di segno (un campione negativo e uno positivo) e per interpolazione lineare ricava la posizione (anche se non intera) in cui si dovrebbe trovare lo zero. Quest'ultimo valore viene poi memorizzato come onset.



**Figura 6.7:** Risposta impulsiva normalizzata.

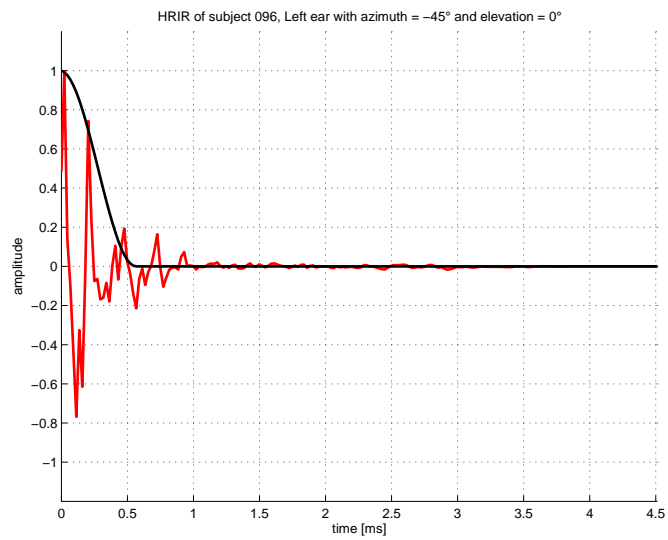
### 6.1.3 Normalizzazione della HRIR

A seguito dell'allineamento la HRIR viene normalizzata: si ricerca il valore di picco e si divide proprio per tale valore ogni campione del segnale. Al termine di questa fase tutti i campioni avranno un valore compreso nell'intervallo  $[-1, 1]$  (figura 6.7); il campione in corrispondenza del valore di picco varrà proprio 1.

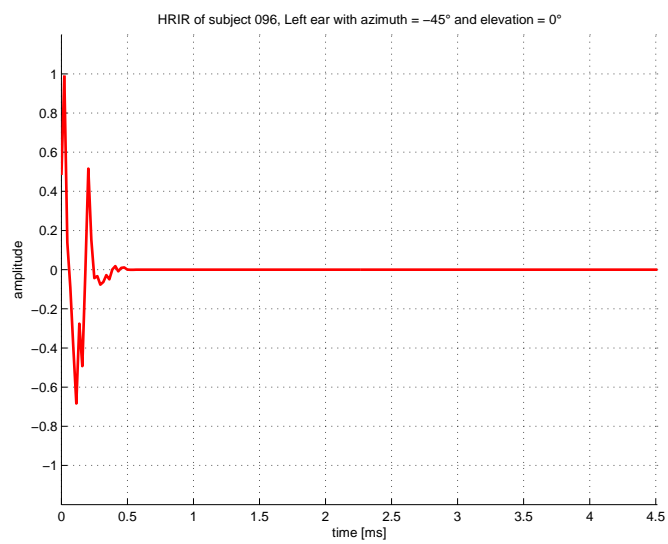
### 6.1.4 Finestratura della HRIR con finestra di *Hann*

Le onde sonore impiegano un certo tempo per coprire la distanza dalla sorgente sonora alle orecchie. Mediante la finestratura temporale è possibile selezionare quelle onde che raggiungono il microfono posto nelle orecchie entro un certo lasso di tempo. Tuttavia l'uso di una semplice finestra rettangolare comporterebbe artifici nel segnale risultante e pertanto si preferisce usare una finestra che attenui gradualmente il segnale, quale quella di *Hann*.

La finestra di *Hann* si presenta simmetrica rispetto all'origine; nella pratica però si sfrutta la metà destra della finestra allineandola alla HRIR normalizzata come illustrato nella figura 6.8. La finestra garantisce il mantenimento pressoché invariato dell'attacco del suono, mentre attenua gradualmente le oscillazioni residue dovute alle riflessioni più ritardate; nella figura 6.9 se ne possono osservare gli effetti. Per i motivi che saranno successivamente descritti nella sezione 6.2, la durata della finestra è stata impostata a 0.6 ms.



**Figura 6.8:** Risposta impulsiva con sovrapposta la finestra di Hann da 0.6 ms.



**Figura 6.9:** Risposta impulsiva finestrata.



### 6.1.5 Calcolo della HRTF usando la trasformata di Fourier

A seguito delle operazioni eseguite direttamente sulla HRIR, il segnale risultante viene trasformato dal dominio del tempo a quello della frequenza attraverso la trasformata di Fourier, la quale restituisce la HRTF, dalla quale sarà possibile ricavare le feature spettrali. La frequenza di campionamento è un parametro di riferimento importante e definisce la risoluzione spettrale della HRTF. La lunghezza in termini di campioni della HRTF in generale può essere diversa dalla HRIR: si è soliti utilizzare una potenza di 2 al fine di sfruttare i vantaggi computazionali della FFT (nello specifico si è impostata la lunghezza a 512 campioni).

Il risultato della trasformata si colloca nel campo complesso; si pone l'attenzione soprattutto sul modulo della HRTF (da qui in avanti con HRTF se ne intenderà il modulo), il quale evidenzia in maniera più nitida ed intuitiva le feature che si vogliono ricercare.

Le operazioni eseguite sulla HRIR producono delle modifiche sensibili che sono apprezzabili anche nel dominio della frequenza. Infatti il segnale in frequenza è meno soggetto ad oscillazioni e appare più smussato dal punto di vista grafico: ciò garantisce maggior precisione ed efficienza computazionale nella ricerca dei punti salienti, quali massimi e minimi locali, strettamente correlati ai picchi di risonanza e ai notch. Il confronto tra le figure 6.10 e 6.11 mostra come la finestrazione abbia risaltato nella HRTF i punti estremanti del segnale.

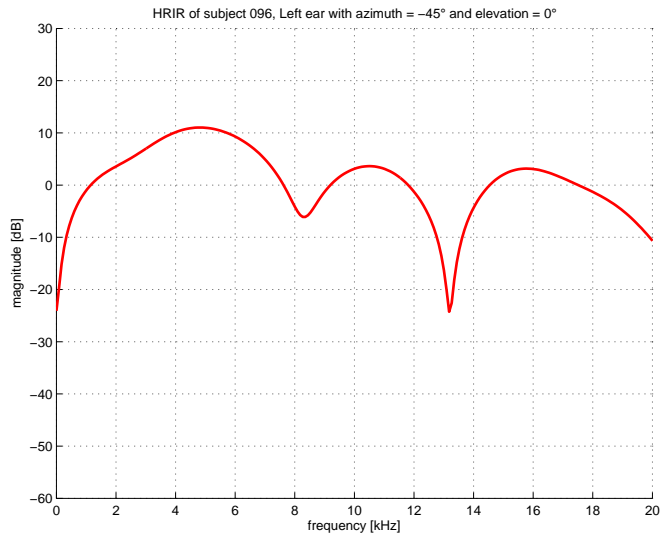
Sia in letteratura che nella pratica le risposte in frequenza si prestano ad una migliore analisi sotto la scala logaritmica. Quindi a seguito dell'operazione di trasformata, le HRTF sono convertite in decibel secondo la legge:

$$H_{dB} = 20 \log_{10}(H) \quad (6.1)$$

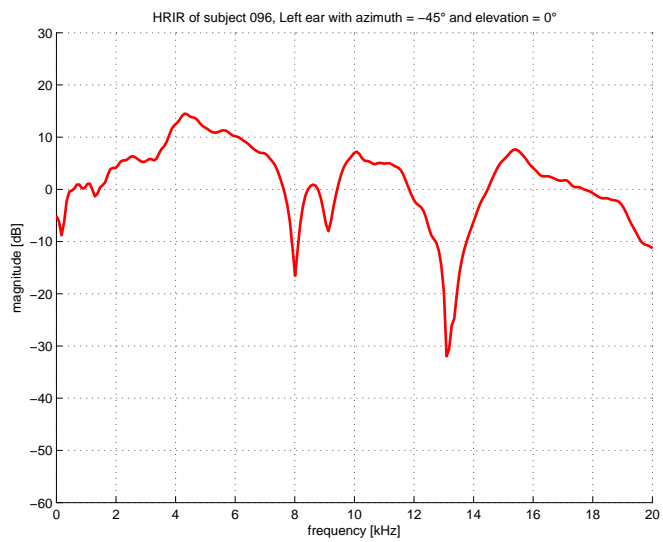
### 6.1.6 Estrazione della HRTF secondo il range di frequenze desiderato

Come già anticipato, la frequenza di campionamento  $f_s$  determina la risoluzione spettrale che avrà la HRTF a seguito della trasformata di Fourier. Nello specifico i campioni corrisponderanno a dei punti uniformemente distribuiti nell'intervallo  $[0, f_s/2[$ , espresso in Hz.

L'orecchio umano è in grado di udire i suoni nell'intervallo compreso tra 20 Hz e 20 kHz; tuttavia molte persone non sono in grado di udire frequenze al di sopra dei 16 kHz. Per questo motivo le HRTF vengono considerate solamente entro un certo range di frequenze tale da essere significativo per la percezione umana. Quindi dall'intera HRTF se ne estrae la porzione associata alle frequenze udibili, la quale sarà oggetto di analisi ed elaborazioni.



**Figura 6.10:** *Risposta in frequenza del segnale finestrato.*



**Figura 6.11:** *Risposta in frequenza del segnale non finestrato.*

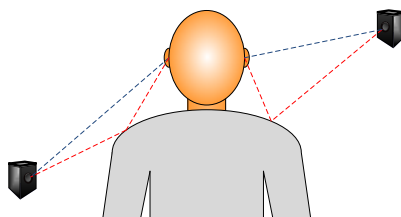


Figura 6.12: Riflessioni dovute alle spalle.

## 6.2 Eliminazione del contributo acustico di busto e spalle

Pur se registrate in un ambiente anecoico, le HRIR contengono molte riflessioni dovute all'interazione delle onde sonore con il corpo dei soggetti, in particolare il busto e le spalle. Per svolgere uno studio effettivo sul contributo acustico dell'orecchio esterno è necessario rimuovere quello relativo a busto e spalle; tale operazione si esegue nel dominio del tempo, introducendo una finestra temporale.

In quanto riflesse, tali onde impiegano un tempo maggiore per giungere all'orecchio rispetto alle onde dirette. L'applicazione di una finestrata temporale di *Hann* della durata opportuna permette di filtrare le onde in modo da:

- mantenere pressoché invariata l'onda diretta,
- attenuare le onde riflesse che accusano un ritardo contenuto,
- annullare le onde riflesse eccessivamente ritardate.

Nel secondo punto tuttavia sono incluse tutte quelle onde riflesse che giungono all'orecchio entro la durata della finestra. Le onde che presentano un lieve ritardo rispetto a quella diretta sono poco attenuate dalla finestra, mentre le onde più ritardate subiranno un'attenuazione maggiore.

Il ritardo delle onde causate dalle riflessioni del busto e delle spalle dipende in particolare dalla posizione della sorgente sonora (figura 6.12):

- per suoni provenienti dal basso, l'onda viene “deviata” dalla spalla verso l'orecchio;
- per suoni provenienti dall'alto, l'onda deve raggiungere la spalla da cui viene “rimbalzata” verso l'orecchio.

Insieme alle onde riflesse da busto e spalle, anche le risonanze e i notch prodotti dall'orecchio esterno raggiungono il canale uditivo con un certo ritardo: tuttavia questi sono eventi acustici

che si verificano sull'onda diretta entro 1 ms, quindi con un ritardo relativamente lieve rispetto all'onda diretta. La finestra temporale quindi deve avere una durata tale da:

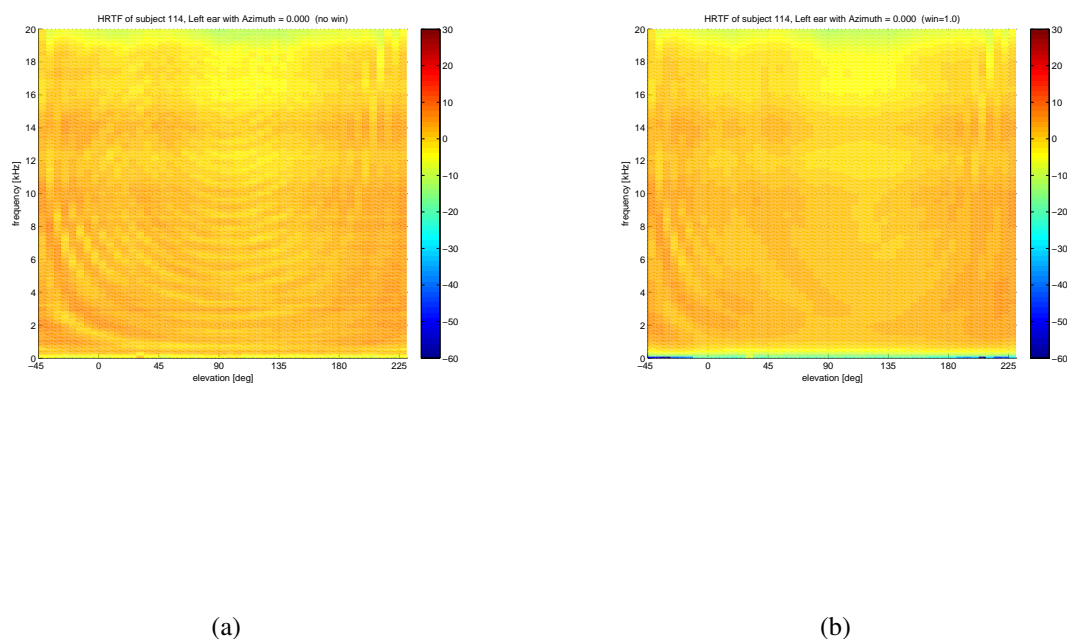
- assicurare il mantenimento di risonanze e notch prodotti dall'orecchio esterno,
- annullare le riflessioni prodotte da busto e spalle.

### **6.2.1 Analisi della HRTF priva del contributo dell'orecchio esterno al variare della finestra temporale**

La linea di demarcazione tra le onde riflesse che vengono mantenute e quelle che vengono annullate è sottile ed è stata effettuata una ricerca variando la durata temporale della finestra applicata alla HRIR (allineata all'onset). A tale scopo è stato preso in considerazione il soggetto KEMAR *pinnaless* sul quale testare delle finestre di diverse lunghezze per osservare il livello di attenuazione delle riflessioni di busto e spalle. In prima istanza è stato scelto questo soggetto affinché si potessero distinguere gli effetti di busto e spalle senza la presenza dei contributi dell'orecchio esterno. Poiché la testa non svolge un ruolo importante nella determinazione dell'altezza della sorgente sonora, analizzando il piano mediano del soggetto KEMAR *pinnaless* è possibile apprezzare i contributi forniti da busto e spalle al variare dell'angolo di elevazione. Nelle figure 6.13 e 6.14 sono illustrate le HRTF sul piano mediano a varie lunghezze della finestra temporale.

### **6.2.2 Analisi della HRTF completa al variare della finestra temporale**

In linea teorica, nella situazione in cui il contributo acustico di busto e spalle sia stato completamente rimosso, ci si aspetterebbe una risposta piatta per le ragioni di cui sopra. Dalla figura questo si verifica soprattutto per elevazioni al di sopra del piano orizzontale ( $\varphi \in [0, 180]$ ) dove la HRTF non supera i 2 dB. Invece per le elevazioni al di sotto del piano orizzontale la HRTF riporta valori superiori dovuti a delle leggere riflessioni e rifrazioni, causate da lievi deviazioni di alcune onde sonore venute a contatto con le spalle; a parità di angolo di elevazione tale fenomeno si verifica maggiormente alle basse frequenze, attenuandosi invece man mano che questa aumenta. La spiegazione deriva dalla relazione tra la lunghezza d'onda e le spalle: per le basse frequenze, alle quali corrisponde una lunghezza d'onda maggiore, il contatto con le spalle produce deviazione e rifrazione delle onde, mentre per le alte frequenze, alle quali corrisponde una lunghezza d'onda minore, le spalle agiscono principalmente come ostacolo acustico ponendo in ombra l'orecchio. L'analisi svolta sul soggetto KEMAR *pinnaless* porterebbe a scegliere la finestra più breve (di durata pari a 0.3 ms) come la più adatta per l'eliminazione del contributo acustico di busto e

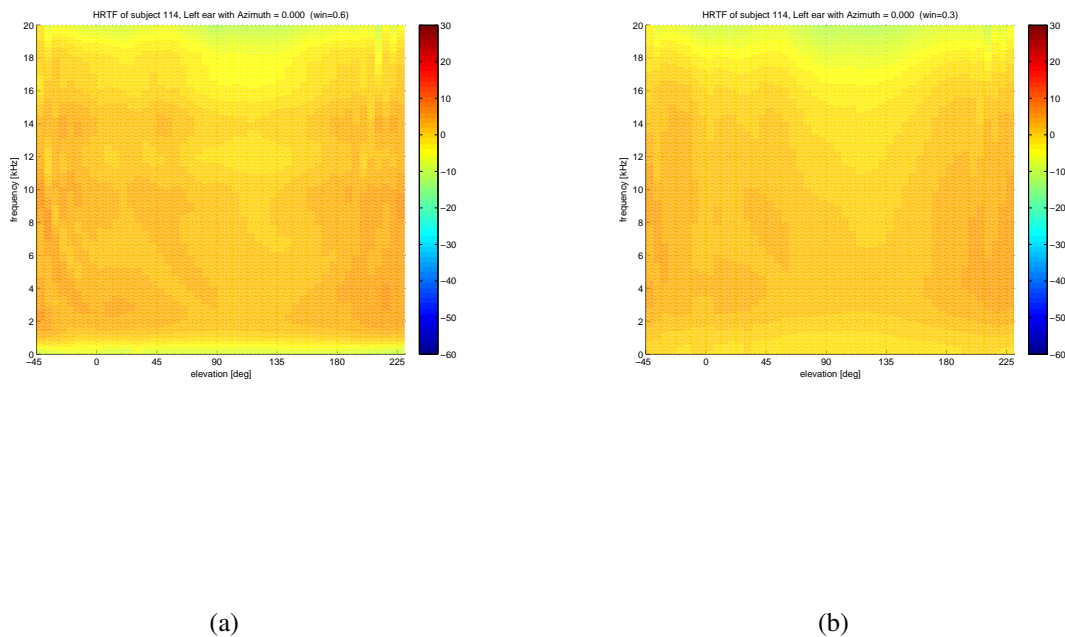


**Figura 6.13:** *HRTF del soggetto KEMAR pinnaless nel piano mediano senza finestrazione (a) e con finestrazione da 1 ms (b).*

spalle. Tuttavia in questa fase è stata trascurata la componente acustica relativa all'orecchio esterno sfruttando la peculiarità del soggetto in questione: così facendo non si conoscono gli effetti della finestrazione su risonanze e notch prodotti dalla pinna. Diventa opportuno ripetere l'analisi svolta sul KEMAR *pinnaless* per un altro soggetto, dotato di orecchie, al fine di osservare le ripercussioni che la finestrazione può avere sulle riflessioni prodotte dall'orecchio esterno. Il soggetto preso in esame è il KEMAR e ne sono state osservate le HRTF nelle stesse condizioni dello studio precedente (figure 6.15 e 6.16).

### 6.2.3 Risultati sulla HRTF al variare della finestra temporale

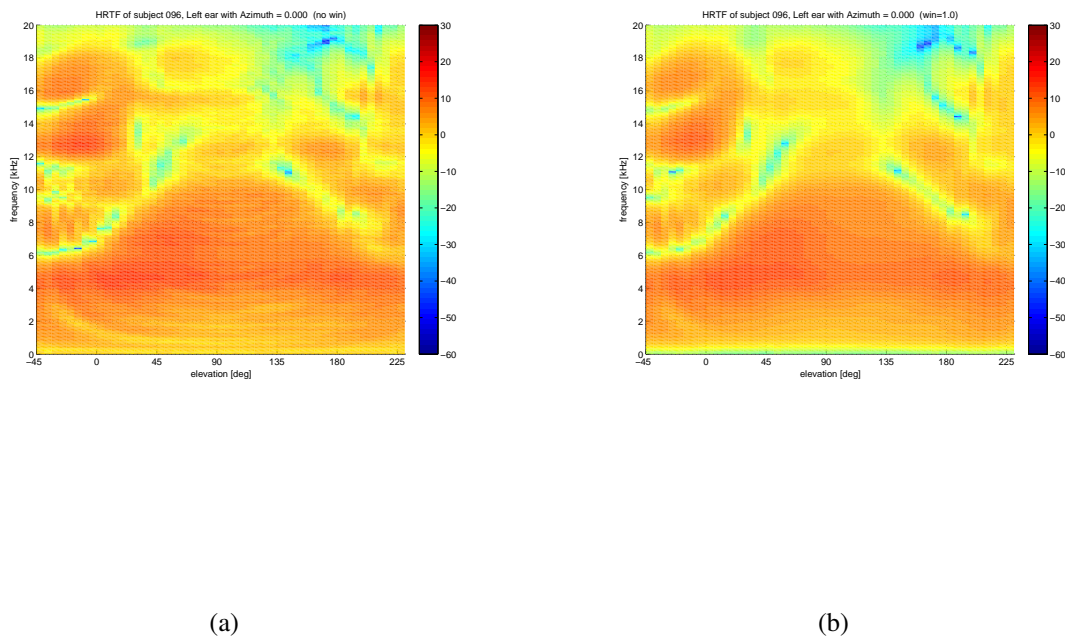
Dal punto di vista teorico le riflessioni prodotte dall'orecchio esterno agenti sull'onda diretta in generale raggiungono il microfono (posto nell'orecchio del soggetto) prima delle riflessioni



**Figura 6.14:** *HRTF del soggetto KEMAR pinnaless nel piano mediano con finestra da 0.6 ms (a) e da 0.3 ms (b).*

dovute all'interazione delle onde sonore con busto e spalle. Per tale motivo bisogna trovare una finestra della durata opportuna in grado di preservare i contributi acustici dell'orecchio esterno ed annullare quelli di busto e spalle.

Osservando la figura qqq si nota subito come la finestra più breve, eletta inizialmente dall'analisi del soggetto KEMAR *pinnaless* come la migliore nel filtrare le riflessioni indesiderate, purtroppo attenua eccessivamente anche le risonanze e i notch generati dalla pinna: le risonanze dei modi 4,5 e 6 si fondono in una sola, la risonanza della conca perde molta della sua ampiezza, mentre i notch vengono parzialmente cancellati e risultano anche meno evidenti; pertanto la finestra di 0.3 ms deve essere scartata. La finestra di 1.0 ms mantiene inalterati i contenuti della pinna ma non filtra completamente quelli di busto e spalle che rimangono visibili. La soluzione ideale si riscontra nella finestra di 0.6 ms: questa rappresenta il giusto compromesso tra le riflessioni da mantenere e quelle da annullare. In realtà rimane una leggera traccia del busto e delle spalle per

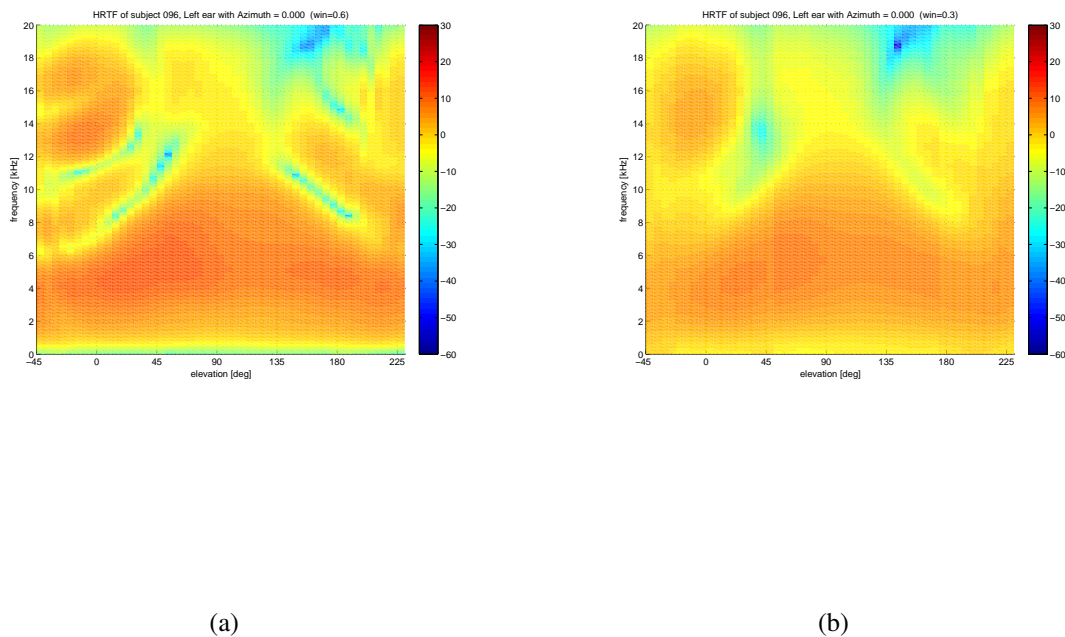


**Figura 6.15:** *HRTF del soggetto KEMAR nel piano mediano senza finestrazione (a) e con finestrazione da 1 ms (b).*

i suoni provenienti dal basso, ma considerando l'intera HRTF la si può considerare trascurabile poiché le feature della pinna risultano predominanti.

### 6.3 Eliminazione del contributo acustico della testa

Dopo aver sottratto gli effetti di busto e spalle la HRTF risultante sarà composta dai contributi della testa e dell'orecchio esterno. In questa fase si vuole estrarre la PRTF che conterrà solamente le feature spettrali della pinna. Pertanto si deve procedere ad eliminare il contributo acustico della testa, mantenendo intatto quello dell'orecchio esterno. Lo studio di questo lavoro di tesi si focalizza sull'analisi dei piani sagittali. Saranno quindi presi in considerazione quei soggetti le cui misurazioni sono state effettuate secondo il sistema di riferimento interaurale polare. Ogni piano sarà contraddistinto da un valore fisso di azimuth mentre l'elevazione può variare su un



**Figura 6.16:** *HRTF del soggetto KEMAR nel piano mediano con finestrazione da 0.6 ms (a) e da 0.3 ms (b).*

range di  $360^\circ$ . Considerando tutte le HRTF ad azimuth  $\theta$  si ottiene la HRTF del piano sagittale relativo a tale azimuth.

### 6.3.1 Eliminazione dei contributi omnidirezionali

All'interno di un piano sagittale (quindi ad azimuth  $\theta$  costante) sotto l'assunzione di testa sferica, per ogni valore di frequenza è ragionevole approssimare gli effetti della testa costanti al variare dell'elevazione. I contributi che si rivelano fissi a parità di frequenza (al variare dell'elevazione) si definiscono omnidirezionali; questi devono quindi essere ricavati per ogni valore di frequenza, per poi essere sottratti dalla HRTF del piano.

Ogni contributo omnidirezionale viene calcolato eseguendo la RMS per ogni valore di frequenza (al variare dell'elevazione) della HRTF del piano sagittale. Di seguito è data una formulazione



del calcolo della RMS per la generica frequenza  $f$  nel piano sagittale di azimuth  $\theta$ :

$$RMS_{HRTF}(\theta, f) = \sqrt{\frac{1}{|\Phi|} \sum_{\varphi \in \Phi} HRTF^2(\theta, \varphi, f)} \quad (6.2)$$

dove  $\Phi$  è l'insieme discreto dei possibili valori di elevazione. Una volta calcolata la RMS per tutte le  $N$  frequenze discretizzate, si può comporre il vettore RMS che sarà relativo al piano sagittale in questione:

$$RMS_{HRTF}(\theta) = [RMS_{HRTF}(\theta, f_1), RMS_{HRTF}(\theta, f_2), \dots, RMS_{HRTF}(\theta, f_N)] \quad (6.3)$$

questo vettore ha la medesima lunghezza di una generica HRTF.

Nella figura 6.17 si pongono a confronto la HRTF del soggetto KEMAR *pinnaless* e l'approssimazione del contributo della testa calcolato con RMS. Dai grafici si osserva che l'approssimazione è buona nell'ipsilaterale, mentre nel controlaterale l'attenuazione dovuta all'effetto di ombra della testa è molto meno pronunciata, anche se comunque si osserva un andamento decrescente nel modulo del segnale.

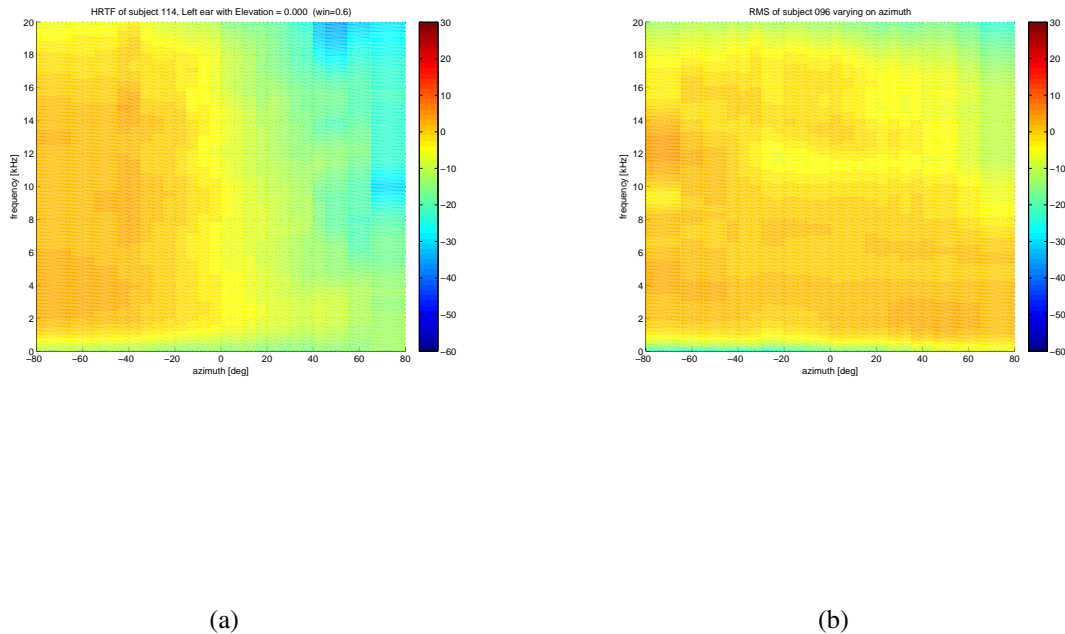
Noto il contributo omnidirezionale, lo si sottrae da ogni HRTF di azimuth  $\theta$ , ottenendo così un nuovo segnale che non è ancora la PRTF effettiva e per tale motivo viene definito PRTF'. In effetti la procedura di sottrazione dei contributi omnidirezionali, oltre ad eliminare gli effetti della testa, ha rimosso anche il modo di risonanza dovuto alla conca, il quale si presentava ad ogni angolo di elevazione. Diventa opportuno pertanto recuperare tale risonanza per aggiungerla alla PRTF'.

### 6.3.2 Compensazione del contributo acustico della conca

Per ristabilire completamente le feature della pinna all'istante antecedente la sottrazione della RMS, bisogna aggiungere la risonanza della conca; questa viene approssimata attraverso un peak filter che deve modellare solamente tale risonanza: dall'insieme di tutti i peak filter ricavati dalle HRTF di un piano sagittale si ottiene un segnale definito  $H_{concha}$ .

I coefficienti del peak filter si calcolano a partire dai parametri della risonanza della conca:

- $G$ : il valore in decibel del picco della risonanza;
- $f_c$ : la frequenza centrale corrispondente al picco;



**Figura 6.17:** Confronto tra la HRTF del soggetto KEMAR pinnaless (a) e l'approssimazione del contributo della testa calcolato con RMS (b).

- $f_b$ : l'ampiezza di banda della risonanza.

Dagli studi di Shaw la conca mediamente esercita una risonanza omnidirezionale attorno alla frequenza di 4.2 kHz con massimo intorno a 11.8 dB; tuttavia analizzando vari soggetti questi valori non rimangono costanti e pertanto bisogna cercarli per ogni soggetto all'interno delle HRTF. Inoltre anche l'ampiezza di banda della risonanza varia da soggetto a soggetto.

Per ricavare quindi i parametri che contraddistinguono la risonanza della conca (ossia  $G$ ,  $f_c$ ,  $f_b$ ) è stata predisposta una procedura basata su metodi statistici. Questa procedura lavora sulla HRTF di un intero piano sagittale e sulla relativa RMS. In una prima fase si ricercano i picchi e le corrispondenti frequenze centrali sulla HRTF del piano; successivamente, con l'ausilio della RMS si definiscono anche le ampiezze di banda.

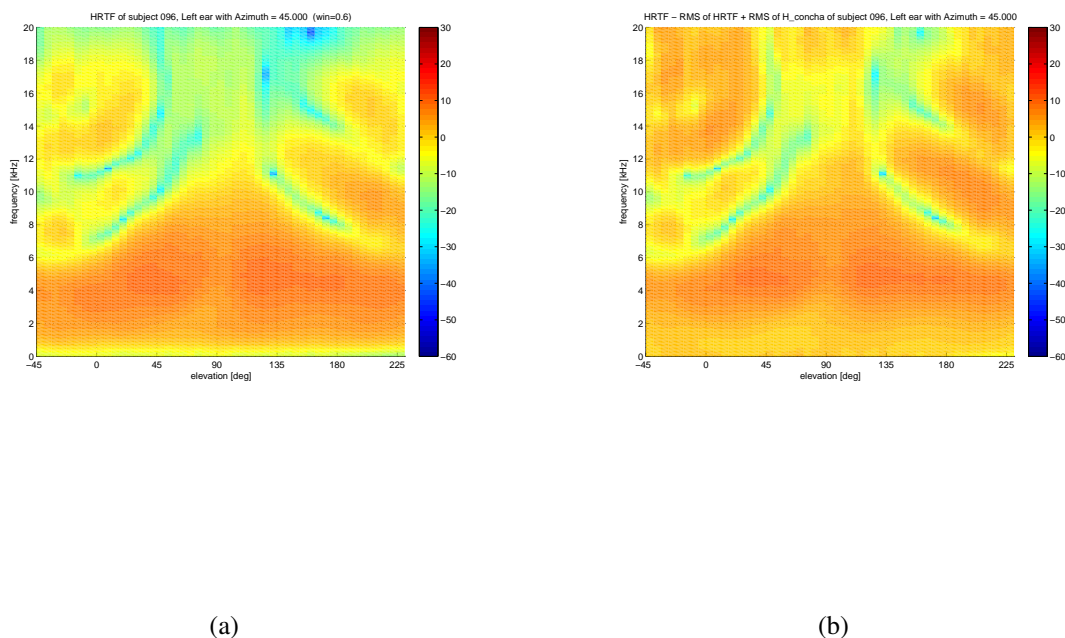
Nella prima fase la procedura ricerca per ogni HRTF (quindi per ogni elevazione) il picco massimo da cui ricava anche la frequenza corrispondente; nella maggioranza dei casi il picco trovato coincide con quello effettivamente ricercato (ovvero nella risonanza della conca), tuttavia talvolta esso appartiene ad altre risonanze, ben distante dalle frequenze della conca. Pertanto si ripete ciclicamente la ricerca del picco riducendo di volta in volta l'intervallo di ricerca attorno alla media delle frequenze centrali trovate al passo precedente; naturalmente alla prima iterazione l'intervallo comprende l'intera gamma di frequenze. Questa tecnica sfrutta una variante della rappresentazione grafica *box-plot*, impiegata in statistica per descrivere le distribuzioni e osservarne la dispersione. In particolare il *box-plot* non è stato impiegato semplicemente per analizzare bensì per ridurre la dispersione forzando la ricerca attorno alla media; così facendo si sono eliminati i falsi positivi legati alla ricerca dei massimi.

Noti la media  $\mu$  e la deviazione standard  $\sigma$  della distribuzione delle frequenze centrali ad ogni iterazione, si è scelto l'intervallo  $[\mu - 2\sigma, \mu + 2\sigma]$  come intervallo di ricerca. In effetti si sarebbe potuta adottare una tecnica più semplice, svolgendo una sola ricerca (quindi senza la presenza di cicli) all'interno di un intervallo prefissato che contiene le frequenze tipiche della conca. Questa alternativa è stata effettivamente testata ma la differenza tra i soggetti non assicurava che l'intervallo fosse valido in tutti i casi: pertanto bisognava impostare un intervallo di ricerca più ampio che però poteva portare a comprendere al suo interno anche frequenze appartenenti ad altre risonanze. In definitiva è stata adottata la prima soluzione, la quale garantisce sempre una ricerca corretta a fronte di un aumento della complessità computazionale quasi impercettibile; sono sufficienti poche iterazioni, nell'ordine delle unità, perché l'algoritmo converga.

Nella seconda fase si determinano le larghezze di banda: per ogni HRTF a partire dalla frequenza centrale si cerca il punto in cui il segnale scende al di sotto di una certa soglia; la loro differenza determina la larghezza di banda. Tuttavia questa strategia non può essere applicata a tutte le HRTF perché in alcune di esse (quelle corrispondenti alle elevazioni intorno ai  $90^\circ$ ) si trovano altre risonanze che disturbano la ricerca. Per tale motivo anziché utilizzare le HRTF si esegue la ricerca sulla RMS del piano sagittale, la quale, essendo legata ai contributi omnidirezionali, non contiene informazioni circa le altre risonanze. La strategia per determinare la larghezza di banda è la medesima: si cerca il punto in cui la RMS scende al di sotto del 50% del valore massimo della RMS (che si trova in corrispondenza della frequenza centrale della risonanza della conca). Per ogni HRTF la differenza tra la frequenza centrale e la frequenza di soglia ricavata nella seconda fase ne determina la larghezza di banda.

Una volta ricavati i parametri della risonanza della conca ( $G, f_c, f_b$ ) per una particolare elevazione si costruisce il peak filter, calcolandone i coefficienti come descritto nella sezione 5.3.1.

A questo punto l'insieme di tutti i peak filter ricavati dalle HRTF di un piano sagittale costituisce il segnale  $H_{\text{concha}}$ ; di questo se ne calcola la RMS che rappresenterà il contributo di compensazione da somministrare a tutte le PRTF' del piano per recuperare la risonanza della conca e ottenere così le relative PRTF. Confrontando i due grafici della figura 6.18 si può notare come la risonanza della conca sia stata ripristinata.



**Figura 6.18:** Analisi del soggetto KEMAR nel piano sagittale di azimuth  $\theta = 45^\circ$  all'orecchio sinistro: HRTF (a) e PRTF (b).

## 6.4 Estrazione e analisi delle feature spettrali dell'orecchio esterno

Per analizzare i contributi forniti dell'orecchio esterno per la localizzazione del suono è necessario descrivere il comportamento dei notch e dei picchi di risonanza al variare della po-

sizione della sorgente sonora nello spazio. Quindi queste componenti caratterizzanti le HRTF (e di conseguenza anche le PRTF) consentono di studiare gli indicatori di localizzazione forniti dall'orecchio esterno.

### 6.4.1 Estrazione delle feature spettrali dell'orecchio esterno

L'algoritmo impiegato per l'estrazione e la separazione dei notch e dei picchi di risonanza di una PRTF è una versione modificata di quello proposto da Geronazzo [11] per lo studio nel piano mediano. In questo piano il contributo della testa era considerato trascurabile e quindi non era necessaria una sua sottrazione dalla HRTF, la quale a seguito di una finestatura di *Hann* da 1 ms (atta a rimuovere le riflessioni di busto e spalle) era pronta per la rilevazione delle feature spettrali. Questo lavoro di tesi vuole ampliare il campo di studio a tutti i piani sagittali, dall'ipsilaterale al controlaterale.

Una prima modifica impressa nell'algoritmo originale è quindi l'allargamento del range di azimuth da analizzate per coprire tutti i piani sagittali. Al di fuori del piano mediano però gli effetti della testa non possono più essere considerati trascurabili; da qui nasce la necessità di rimuoverli e la relativa procedura descritta nella sezione 6.3 rappresenta la porzione aggiunta all'algoritmo originale per ottenere la PRTF.

### 6.4.2 Analisi delle feature spettrali dell'orecchio esterno

I risultati espressi in forma grafica relativi alle parziali dei notch e dei picchi di risonanza mostrano una buona corrispondenza tra le feature spettrali estratte e la PRTF originale (figura 6.19).

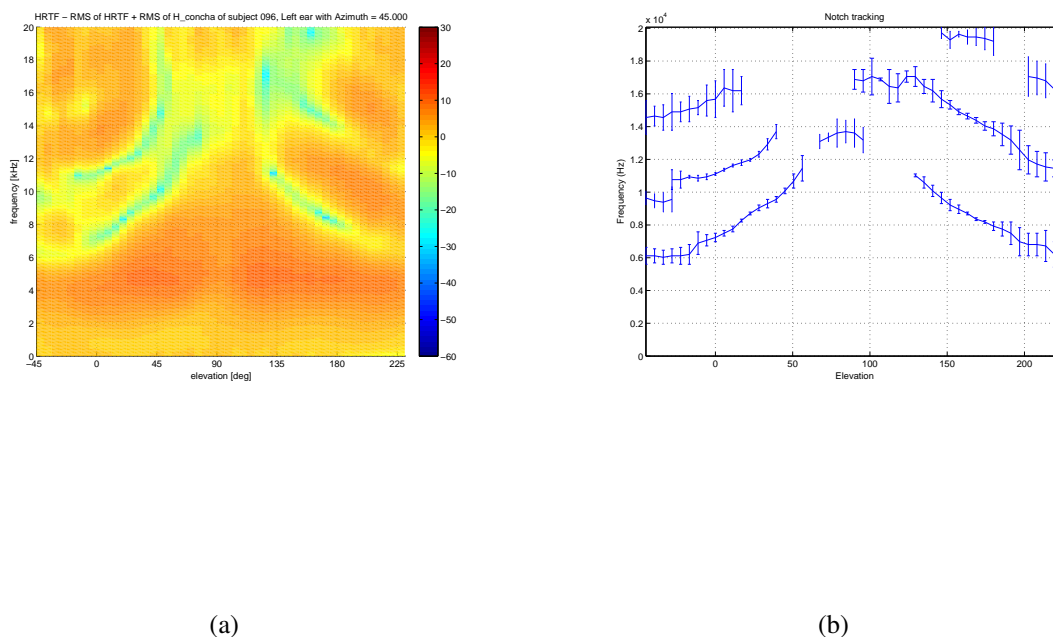
Lo studio a questo punto deve considerare l'evoluzione dei notch e dei picchi di risonanza attraverso i vari piani sagittali.

Per i notch l'attenzione ricade in particolare sulle elevazioni comprese nell'intervallo  $[-45^\circ, 45^\circ]$ . L'analisi in questa fase richiede l'osservazione dell'evoluzione delle tracce delle parziali. Le figure 6.20, 6.21 e 6.22 si riferiscono a tre principali tracce di notch:

**traccia 1** tra 6 kHz e 10 kHz,

**traccia 2** tra 10 kHz e 16 kHz,

**traccia 3** tra 14 kHz e 17 kHz.



**Figura 6.19:** Analisi dei notch per il soggetto KEMAR nel piano sagittale di azimuth  $\theta = 45^\circ$  all'orecchio sinistro: PRTF (a) e notch tracking (b).

In linea teorica i contributi dell'orecchio esterno sono correlati alle variazioni di elevazione, mentre non ci si attendono relazioni al mutare dell'azimuth. Analizzando i grafici si osserva che all'aumentare dell'elevazione le tracce si alzano in frequenza, mantenendo un certo parallelismo tra di esse: questi risultati evidenziano quindi una relazione tra i notch e la localizzazione verticale del suono. Per quanto riguarda invece la relazione tra notch e azimuth, ci si aspetterebbe che le tracce fossero orizzontali dal momento che l'orecchio esterno non dovrebbe fornire indicazione circa la localizzazione orizzontale. Quest'ultima caratteristica non viene confermata per gli angoli di azimuth ipsilaterali estremi. In particolare per il primo *tracking* (figura 6.20) si osserva una convergenza verso l'intervallo compreso tra 8 kHz e 10 kHz, mentre nel secondo (figura 6.21) le tracce si alzano in frequenza di circa 3 kHz.

Secondo Lopez-Poveda e Meddis [16] a causa della diffrazione del suono, si generano delle feature spettrali, dovute alla conca, dipendenti dall'azimuth per le posizioni della sorgente sonora

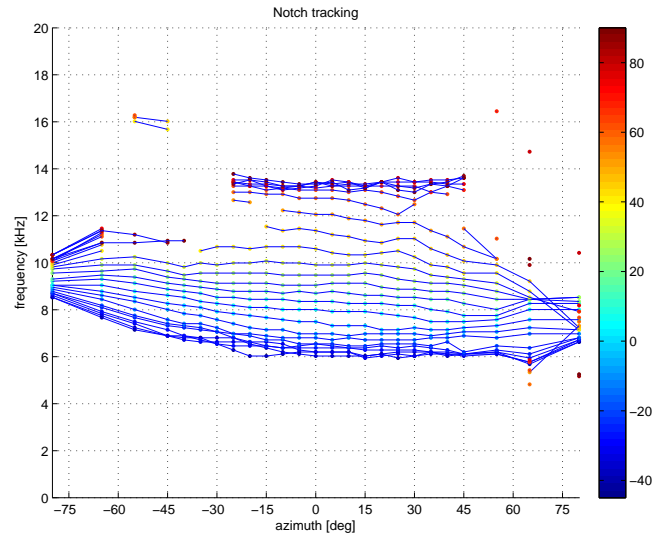


Figura 6.20: Notch tracking no.1 per l'orecchio sinistro.

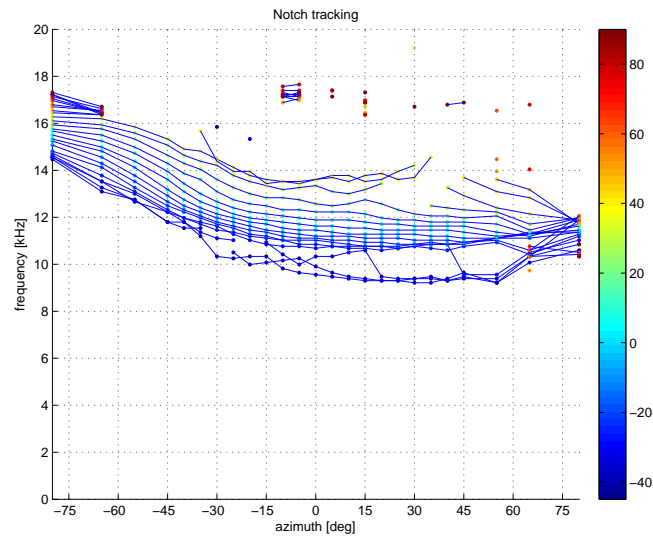
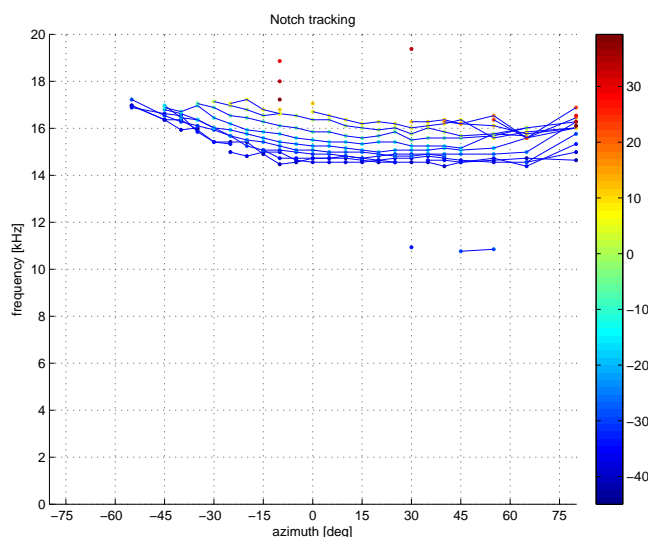
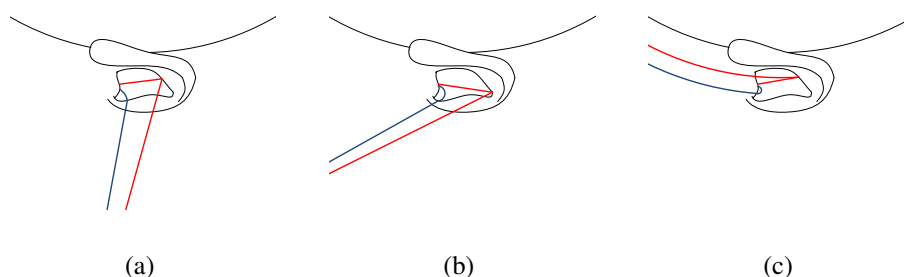


Figura 6.21: Notch tracking no.2 per l'orecchio sinistro.



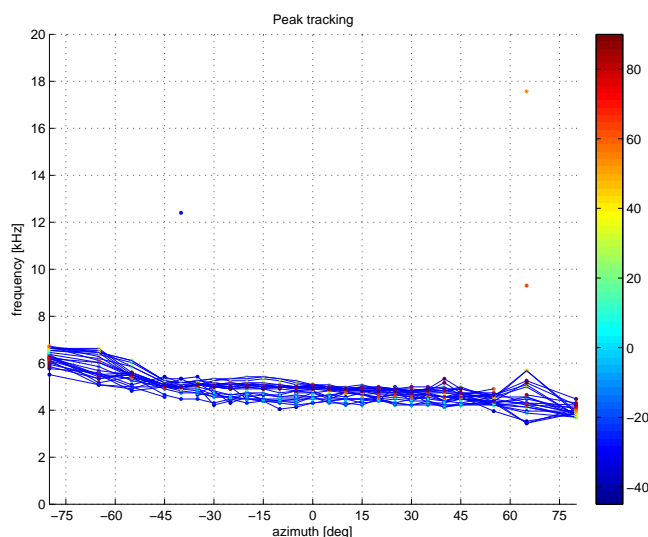
**Figura 6.22:** *Notch tracking no.3 per l'orecchio sinistro.*

più estreme nell'emisfero ipsilaterale. Considerando che più la riflessione distruttiva è ritardata rispetto all'onda diretta e più in basso in frequenza si verifica il notch, una possibile interpretazione della dipendenza dei notch dall'azimuth nell'estremo ipsilaterale può essere legata all'angolo di incidenza delle onde sonore. Nella figura 6.23 sono mostrate le rappresentazioni di tre riflessioni (con prospettiva dall'alto), nelle quali è possibile apprezzare la variazione del ritardo dell'onda riflessa sulla diretta in base all'angolo di incidenza: per suoni provenienti dall'estremo ipsilaterale, l'onda riflessa deve percorrere una traiettoria più breve; man mano che ci si sposta verso sorgenti poste nell'emisfero controlaterale la traiettoria aumenta poiché le onde devono “rimbalzare” su una parete dell'orecchio per poi tornare indietro. Per tutte le posizioni controlaterali, le onde sono soggette a diffrazione attorno alla testae quindi giungono all'orecchio con la medesima incidenza.



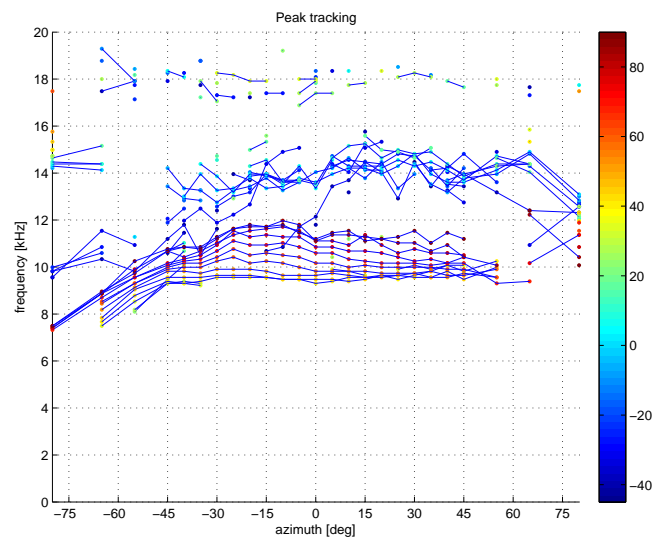
**Figura 6.23:** *Riflessioni generate nell'orecchio al variare dell'azimuth:  $\theta = -80^\circ$  (a),  $\theta = -40^\circ$  (b) e  $\theta = 40^\circ$  (c).*





**Figura 6.24:** *Peak tracking no.1 per l'orecchio sinistro.*

Dagli studi di Shaw si possono classificare tre tipi di risonanze, omnidirezionale, verticale ed orizzontale, caratterizzate da sei modi di risonanza. Nella figura 6.24 (a) sono stati tracciati i picchi relativi alla risonanza della conca; il grafico evidenzia l'omnidirezionalità di tale risonanza che si verifica intorno ai 5 kHz per tutti i valori di azimuth ed elevazione considerati. In realtà i picchi di questa risonanza non rimangono costanti per tutti i piani sagittali: per gli angoli di azimuth nell'estremo emisfero ipsilaterale le tracce dei picchi si alzano di circa 1 kHz, analogamente a quanto accade per i notch. Nella figura 6.24 (b) emergono le risonanze verticali: infatti dal grafico si possono osservare le tracce per le elevazioni comprese tra  $50^\circ$  e  $90^\circ$ , le quali si spostano in frequenza mantenendosi approssimativamente parallele. Questi picchi di risonanza si alzano in frequenza all'aumentare dell'elevazione (analogamente ai notch), ma al variare dell'azimuth mostrano un andamento calante per gli angoli dell'estremo ipsilaterale, opposto a quanto osservato fino ad ora.



**Figura 6.25:** *Peak tracking no.2 per l'orecchio sinistro.*

# Conclusione

Gli obiettivi di questo lavoro di tesi mirano a studiare gli effetti forniti dall'orecchio esterno nella localizzazione verticale del suono, contribuendo ad ampliare l'analisi al di fuori del piano mediano (già ampiamente trattata in letteratura). Le considerazioni e i risultati ottenuti forniscono indicazioni per migliorare la resa di modelli strutturali sintetici impiegati nella riproduzione di audio binaurale in cuffia. In questo lavoro di tesi è stata proposta una nuova metodologia per separare gli effetti dell'orecchio esterno dai contributi delle altre parti del corpo.

Una considerevole porzione di tale lavoro di tesi è stata impiegata per analizzare le strutture dati del repository di HRTF aggiungendo quelle informazioni mancanti importanti per la corretta elaborazione dei segnali. È stato per questo motivo ridefinito un nuovo formato per accogliere i database di HRTF, secondo le necessità riscontrate, che prendesse in considerazione anche altre informazioni utili sia in fase di analisi che in fase di sintesi. Il formato è stato descritto nell'articolo *A standardized repository of Head-Related and Headphone Impulse Response data* accettato alla 134<sup>th</sup> Audio Engineering Society Convention del 2013.

La metodologia proposta per ricavare i contributi dell'orecchio esterno esegue in successione la rimozione degli effetti delle altre parti del corpo. In particolare sono stati eliminati dapprima gli effetti di busto e spalle agendo con delle finestrate nel tempo; in secondo luogo, sotto opportune approssimazioni, è stato eliminato anche il contributo acustico della testa mediante l'uso di filtri.

I risultati estratti dalle PRTF sono in linea con le considerazioni già note in letteratura per quanto riguarda il contributo dell'orecchio esterno nella localizzazione verticale del suono. Visivamente si sono osservate delle relazioni con la coordinata orizzontale che hanno trovato conferma in letteratura per quanto riguarda però lo studio del solo piano orizzontale. È stata proposta pertanto una possibile interpretazione dei risultati per tutte le posizioni spaziali analizzate, tentando di relazionarli direttamente con le caratteristiche anatomiche dell'orecchio esterno.



## Sviluppi futuri

La metodologia proposta comprende delle fasi sperimentali che sono state trattate con delle approssimazioni (come ad esempio l'assunzione di testa sferica) in parte derivanti dalla letteratura e in parte teorizzate in questa tesi. Poiché le analisi delle HRTF variano da individuo a individuo un miglioramento sostanziale si avrebbe prendendo in considerazione anche gli aspetti antropometrici dei soggetti in esame cercando di relazionarli con i contributi forniti delle varie parti del corpo. Per avere un'ancora più completa visione dei risultati sarebbe interessante analizzare e studiare gli effetti prodotti singolarmente dalle sporgenze e dalle depressioni che contraddistinguono l'orecchio esterno (attraverso simulazioni acustiche), per poi contestualizzare con i risultati ottenuti da questo lavoro di tesi.

Una promettente applicazione dell'analisi proposta riguarda la composizione di modelli ibridi customizzati, composti prendendo in considerazione i contributi singoli delle varie parti del corpo di individui diversi che abbiano caratteristiche antropometriche simili al soggetto in esame, di cui non è disponibile la HRTF individuale (proposta nell'articolo *A modular framework for the analysis and synthesis of Head-Related Transfer Functions* di Geronazo, Spagnol e Avanzini, accettato alla 134<sup>th</sup> *Audio Engineering Society Convention* del 2013).

Alcune di queste considerazioni sono già linee di ricerca prese in esame dal *Sound and Music Computing group - Binaural team*.



# **Appendici**





# Appendice A

## Strumenti statistici

### A.1 Root Mean Square (RMS)

La *Root Mean Square* (RMS), nota anche come media quadratica, è un valore notevole che evidenzia l'andamento medio di un segnale (continuo o discreto) soprattutto in presenza di valori sia positivi che negativi.

Per un segnale continuo si definisce:

$$rms(x) = \sqrt{\frac{1}{T} \int_0^T [x(t)]^2 dt} \quad (\text{A.1})$$

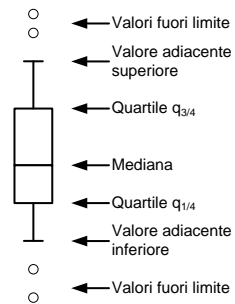
Per un segnale discreto, invece, si definisce:

$$rms(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (\text{A.2})$$

### A.2 Box-plot

Nella statistica descrittiva il *box-plot*, detto anche *box and whiskers plot* (diagramma a scatola e baffi), è una rappresentazione grafica utilizzata per descrivere la distribuzione di un campione tramite semplici indici di dispersione e di posizione.

Viene rappresentato tramite un rettangolo diviso in due parti, da cui escono due segmenti. Il rettangolo (la “scatola”) è delimitato dal primo e dal terzo quartile,  $q_{1/4}$  e  $q_{3/4}$ , e diviso al suo



**Figura A.1:** *Diagramma box-plot.*

interno dalla mediana,  $q_{1/2}$ . I segmenti (i “baffi”) sono delimitati rispettivamente dai valori minimo e massimo. I valori della distribuzione sono rappresentati come dei punti (figura A.1).

Esistono anche delle varianti nella rappresentazione del *box-plot*: una di queste definisce una lunghezza fissa per i segmenti uscenti dal rettangolo in modo tale da evidenziare la dispersione di una distribuzione qualora ci fossero dei punti al di fuori del rettangolo e dei segmenti. Altre varianti considerano valori differenti per delimitare il rettangolo.

# Appendice B

## Funzioni sviluppate in MATLAB

Di seguito sono presentate le intestazioni delle principali funzione realizzate nell'ambiente MATLAB e utilizzate per questo lavoro di tesi.

### B.1 Funzioni di gestione del repository

Sono le funzioni utilizzate per mantenere aggiornato il repository secondo le modifiche apportate al formato *MARL-NYU*.

```
function [data,specs] = convert_subject(database,subject_id)
    %
    % This function can be used to convert a subject to the modified MARL-NYU
    % format; the converted subject is stored in the .mat file
    %
    % Input:
    %     database: the HRIR/HRTF database in which the subject is stored
    %     subject_id: a number containing the id of the subject (in its
    %                 HRIR/HRTF database)
    % Output:
    %     data: the struct vector containing the data of the subject
    %     specs: the struct containing the specifications of the subject

function [data,specs] = update_subject(subject_id)
    % This built ad-hoc function can be used to update the data struct vector
    % and the specs struct of a subject in the modified MARL-NYU format.
    %
    % version 1: adding onset information
    % version 2: adding system coordinates information
```

```
%  
% Input:  
%     subject_id: a number containing the id of the subject (in the  
%                 modified MARL-NYU format)  
% Output:  
%     data: the struct vector containing the data of the subject  
%     specs: the struct containing the specifications of the subject  
%
```

## B.2 Funzioni di carattere generale

Sono le funzioni di ausilio agli algoritmi di analisi; alcune di esse implementano le operazioni più importanti delineate nel capitolo 5.

### Funzioni di conversione delle coordinate

```
function pos_int = car2int(pos_car)  
% This function can be used to convert cartesian coordinates to interaural  
%   polar coordinates  
%  
% Input:  
%     pos_car: the nx3 matrix containing the position in cartesian  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: x  
%                 col 2: y  
%                 col 3: z  
% Output:  
%     pos_int: the nx3 matrix containing the position in interaural polar  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: azimuth  
%                 col 2: elevation  
%                 col 3: distance  
%  
function pos_car = int2car(pos_int)  
% This function can be used to convert interaural polar coordinates to cartesian coo
```

```
%  
% Input:  
%     pos_int: the nx3 matrix containing the position in interaural polar  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: azimuth  
%                 col 2: elevation  
%                 col 3: distance  
% Output:  
%     pos_car: the nx3 matrix containing the position in cartesian  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: x  
%                 col 2: y  
%                 col 3: z  
%  
function pos_ver = car2ver(pos_car)  
% This function can be used to convert cartesian coordinates to vertical polar coord  
%  
% Input:  
%     pos_car: the nx3 matrix containing the position in cartesian  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: x  
%                 col 2: y  
%                 col 3: z  
% Output:  
%     pos_ver: the nx3 matrix containing the position in vertical polar  
%             coordinates  
%             each row refers to one spatial position  
%             each column refers to one coordinate:  
%                 col 1: elevation  
%                 col 2: azimuth  
%                 col 3: distance  
%  
function pos_car = ver2car(pos_ver)  
% This function can be used to convert vertical polar coordinates to cartesian coord  
%
```

```
% Input:
%     pos_ver: the nx3 matrix containing the position in vertical polar
%             coordinates
%             each row refers to one spatial position
%             each column refers to one coordinate:
%                 col 1: elevation
%                 col 2: azimuth
%                 col 3: distance
% Output:
%     pos_car: the nx3 matrix containing the position in cartesian
%             coordinates
%             each row refers to one spatial position
%             each column refers to one coordinate:
%                 col 1: x
%                 col 2: y
%                 col 3: z
%

function pos_ver = int2ver(pos_int)
% This function can be used to convert interaural polar coordinates to vertical polar
%
% Input:
%     pos_int: the nx3 matrix containing the position in interaural polar
%             coordinates
%             each row refers to one spatial position
%             each column refers to one coordinate:
%                 col 1: azimuth
%                 col 2: elevation
%                 col 3: distance
% Output:
%     pos_ver: the nx3 matrix containing the position in vertical polar
%             coordinates
%             each row refers to one spatial position
%             each column refers to one coordinate:
%                 col 1: elevation
%                 col 2: azimuth
%                 col 3: distance
%

function pos_int = ver2int(pos_ver)
% This function can be used to convert vertical polar coordinates to interaural polar
%
% Input:
```

```
%      pos_ver: the nx3 matrix containing the position in vertical polar
%      coordinates
%      each row refers to one spatial position
%      each column refers to one coordinate:
%          col 1: elevation
%          col 2: azimuth
%          col 3: distance
%
% Output:
%      pos_int: the nx3 matrix containing the position in interaural polar
%      coordinates
%      each row refers to one spatial position
%      each column refers to one coordinate:
%          col 1: azimuth
%          col 2: elevation
%          col 3: distance
%
```

## Funzioni di creazione delle strutture di ausilio

```
function create_list_database()
% This function can be used to create the list of database; this list is
% stored in config.mat file
%
% Input:
%     no input
% Output:
%     no output
%
```

```
function create_info_azimuth()
% This function can be used to create the list of angle1 values for each
% different database; these lists are stored in config.mat file
%
% Input:
%     no input
% Output:
%     no output
%
```

```
function create_info_angle2()
% This function can be used to create the list of angle2 values for each
```

```
% different database-angle1 pair; these lists are stored in config.mat
% file
%
% Input:
%     no input
% Output:
%     no output
%
```

## Funzioni di ricerca delle coordinate

```
function angle1_val = get_angle1_val(system,database,angle1_min,angle1_max)
% This function can be used to get the list of angle1 values for the
% selected database
%
% Input:
%     system: the system coordinates
%     database: the original database
%     angle1_min: the minimum threshold of returning angle1 values
%     angle1_max: the maximum threshold of returning angle1 values
% Output:
%     angle1_val: the list of angle1 values
%
```

```
function [angle2_val,position_loc] = ...
%
% This function can be used to get the list of angle2 values and the
% corresponding list of indexes (in data struct vector of modified
% MARL-NYU format) for the selected database-angle1 pair
%
% Input:
%     system: the system coordinates
%     database: the original database
%     angle1 : the value of the angle1
%     angle2_min: the minimum threshold of returning angle2 values
%     angle2_max: the maximum threshold of returning angle2 values
% Output:
%     angle2_val: the list of angle2 values
%     position_loc: the corresponding list of indexes (in data struct
% vector of modified MARL-NYU format)
%
```



## Funzioni inerenti alla teoria dei segnali

```
function x_win = windowing(x,window_type,t_length,fs)
% This function can be used to apply a desired window to the input signal
%
% Input:
%     x: the input signal
%     window_type: the type of window
%           it may be 'hamming', 'hann', 'rectwin' etc
%     t_length: the non-zero right half window length [ms]
%     fs: the sampling frequency [Hz]
% Output:
%     x_win: the windowed signal
%

function RMS_H_f = get_RMS(H_fc)
% This function can be used to get the root mean square of a 2D matrix
%
% Input:
%     H_fc: the frequency response [linear]
%           "_fc" refers to the (frequency,coordinate) indexing
%           coordinate may be azimuth or elevation
% Output:
%     RMS_H_mag_f: the RMS of H [linear]
%           "_f" refers to the (frequency) indexing
%

function [b,a] = design_peak_filter(G,fc,fb,fs)
% This function can be used to compute the coefficients of a 2nd order peak
% filter
%
% Input:
%     G: the gain (magnitude) of the peak [dB]
%     fc: the central frequency of the peak [Hz]
%     fb: the band frequency [Hz]
%     fs: the sampling frequency [Hz]
% Output:
%     b: the numerators coefficients
%     a: the denominators coefficients
%
% @bib Udo Zlder. Digital Audio Signal Processing. JohnWiley & Sons.
%       2008. 2nd edition. Helmut Schmidt University, Hamburg, Germany.
%
```

## B.3 Funzione di elaborazione dei dati

Sono le principali funzioni realizzate per l'elaborazione e l'analisi dei dati.

```
function [G,fc,fb,fs] = get_concha_resonance(f,fs,HRTF_mag_f2,RMS_HRTF_mag_dB_f)
% This function can be used to find the concha resonance of an HRTF
%
% Input:
%   f: the frequency vector [Hz]
%   fs: the sampling frequency [Hz]
%   HRTF_mag_f2: the HRTF magnitude [linear]
%       "_f2" refers to the (frequency,angle_2) indexing
%   RMS_HRTF_mag_dB_f: the RMS magnitude of HRTF [dB]
%       "_f" refers to the (frequency) indexing
% Output:
%   G: the gain (magnitude) vector of the concha peak [dB]
%   fc: the central frequency vector of the concha peak [Hz]
%   fb: the band frequency vector [Hz] of the concha
%   fs: the sampling frequency [Hz]
%

function [G,fc,fb,fs] = get_flat_resonance(f,fs)
% This function can be used to get a flat resonance
%
% Input:
%   f: the frequency vector [Hz]
%   fs: the sampling frequency [Hz]
% Output:
%   G: the gain (magnitude) vector of the concha peak [dB]
%   fc: the central frequency vector of the concha peak [Hz]
%   fb: the band frequency vector [Hz] of the concha
%   fs: the sampling frequency [Hz]
%

function analyze_HRTF(subject_id,angle1_limit,angle2_limit)
% This function can be used to analyze the resonance modes of HRTF of a
% MARL-NYU subject according to the selected angle thresholds.
%
% Input:
%   subject_id: a number containing the id of the subject (in the
%       modified MARL-NYU format)
%   angle1_limit: a 1x2 vector containing minimum and maximum thresholds
%       of angle1 scanning values or a single number containing the only
```

```
%         angle1 scanning value
%         angle2_limit: a 1x2 vector containing minimum and maximum thresholds
%         of angle2 scanning values or a single number containing the only
%         angle2 scanning value
% Output:
%         no output
%

function show_notch_tracking(subject_id,angle1_limit,angle2_limit)
% This function can be used to show the notch tracking of a subject
% according to the selected angle thresholds
%
% Input:
%         subject_id: a number containing the id of the subject (in the
%         modified MARL-NYU format)
%         angle1_limit: a 1x2 vector containing minimum and maximum thresholds
%         of angle1 scanning values or a single number containing the only
%         angle1 scanning value
%         angle2_limit: a 1x2 vector containing minimum and maximum thresholds
%         of angle2 scanning values or a single number containing the only
%         angle2 scanning value
% Output:
%         no output
%
```



# Bibliografia

- [1] Areti Andreopoulou and Agnieszka Roginska. *Documentation for the MARL-NYU file format - Description of the HRIR repository*. Music and Audio Research Laboratory - New York University, October 2011.
- [2] Federico Avanzini. *Algorithms for Sound and Music Computing*. Creative Commons, 2005.
- [3] D. W. Batteau. The role of the pinna in human localization. *Royal Society Publishing*, 168(1011):158–180, August 1967.
- [4] Corey I. Cheng and Gregory H Wakefield. Introduction to head-related transfer functions (hrtfs): Representations of hrtfs in time, frequency, and space. *Audio Engineering Society*, 49(4):231–249, April 2001.
- [5] Ana Isabel Alves Pinto Lopes da Silva. *Psychophysical and physiological assessment of the representation of high-frequency spectral notches in the auditory nerve*. PhD thesis, Universidad de Salamanca and Instituto de Neurociencias de Castilla y León, May 2007. Enrique A. Lopez Poveda (Chair).
- [6] Kathleen Brown David R. Perrott, Kouros Saberi and Thomas Z. Strybel. Auditory psychomotor coordination and visual search performance. *Perception and Psychophysics*, 48(3):214–226, May 1990.
- [7] Institute de Recherche et Coordination Acoustique/Musique. Listen hrtf database, 2003. [Online; accessed 1-December-2012].
- [8] Digital Signal Processing Lab Florida International University. Hrtf/anthropometric measurement database, 2004. [Online; accessed 1-December-2012].
- [9] Scott Foster. Impulse response measurement using golay codes. In *ICASSP 86 proceedings*, pages 929–932, April 1986.

- [10] Bill Gardner and Keith Martin. Hrtf measurements of a kemar dummy-head microphone, 1994. [Online; accessed 1-December-2012].
- [11] Michele Geronazzo. Modello strutturale di orecchio esterno per il rendering audio 3d. Master's thesis, Università degli studi di Padova, 2009.
- [12] Luca Mion Giovanni De Poli. *Algorithms for Sound and Music Computing*. Creative Commons, 2006.
- [13] Cyril M. Harris. On the absorption of sound in humid air at reduced pressures. *Acoustical Society of America*, 43(3):530–532, 1968.
- [14] The MathWorks Inc. Matlab - the language of technical computing, 2012. [Online; accessed 2-December-2012].
- [15] Bill Kapralos. *Auditory Perception and Virtual Environments*. PhD thesis, Department of Computer Science - York University - North York, Ontario, Canada, January 2003.
- [16] Enrique A. Lopez-Poveda and Ray Meddis. A physical model of sound diffraction and reflections in the human cochlear. *Acoustical Society of America*, 100(5):3248–3259, November 1996.
- [17] Robert J. McAulay and Thomas F. Quatieri Jr. Speech analysis/synthesis based on a sinusoidal representation. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 34(4):744–754, 1986.
- [18] Carlo Muscas. *Dispense del corso di Misure Elettroniche*. , 2007.
- [19] J. W. Strutt (Lord Raleigh). Our perception of sound direction. *Philosophical Magazine*, 13:214–232, 1907.
- [20] Patrick Satarzadeh. A study of physical and circuit models of the human pinnae. Master's thesis, UNIVERSITY OF CALIFORNIA, DAVIS, 2006. V. Ralph Algazi (Chair).
- [21] Edgar A.G Shaw. *Acoustical Features of the Human External Ear*, chapter 2. Lawrence Erlbaum Associates, 1997.
- [22] D.M. Thompson V.R. Algazi, R.O. Duda and C. Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics*, pages 99–102, 2001.

- 
- [23] Wikipedia. Chirp — wikipedia, the free encyclopedia, 2012. [Online; accessed 1-December-2012].
- [24] Wikipedia. Cross-correlation — wikipedia, the free encyclopedia, 2012. [Online; accessed 1-December-2012].
- [25] Wikipedia. Maximum length sequence — wikipedia, the free encyclopedia, 2012. [Online; accessed 1-December-2012].
- [26] Wikipedia. Pinna (anatomy) — wikipedia, the free encyclopedia, 2012. [Online; accessed 30-November-2012].
- [27] Udo Zölder. *Digital Audio Signal Processing*. John Wiley & Sons, 2008.