



UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE  
LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

**METODI PER LA SELEZIONE DI HRTF  
BASATI SU CONTORNO DELL'ORECCHIO IN  
AMBIENTI ACUSTICI VIRTUALI**

*Laureando:*  
Enrico Peruch

*Relatore:*  
Prof. Federico Avanzini

*Matricola:*  
1130019

*Correlatore:*  
Dott. Michele Geronazzo

Anno Accademico 2016/2017  
Data Laurea 04/07/2017



# Sommario

Il problema della personalizzazione dell'audio binaurale è un argomento molto dibattuto in questi ultimi anni. Il rendering audio 3D è sempre più richiesto in ambienti di realtà virtuale con applicazioni sempre più diffuse. Le HRTF sono alla base del rendering binaurale, in quanto un segnale audio mono convoluto con un set di HRTF in funzione della posizione spaziale permette di spazializzare il suono.

Questo lavoro di tesi propone uno studio di alcuni metodi di misura e di selezione della propria HRTF personalizzata mediante la misurazione delle caratteristiche antropometriche uniche del padiglione auricolare grazie ad un'applicazione sviluppata in Matlab. Lo strumento permette a qualsiasi utente, senza particolari conoscenze nel settore e senza necessità di attrezzature dispendiose, di selezionare in modo rapido ed intuitivo la propria HRTF personalizzata da un database.

Le metriche di selezione proposte verranno validate con dei soggetti virtuali presi da un database di HRTF. Quando essi sono dotati dei set di HRTF selezionati dall'applicazione, verranno analizzate le prestazioni di localizzazione in elevazione tramite una simulazione con modello uditivo. Seguirà un'analisi su un campione di soggetti reali: un test in realtà virtuale permetterà di paragonare le performance di localizzazione quando essi sono dotati dei set di HRTF selezionati rispetto ad una HRTF generica (KEMAR).

I risultati riportano una valida applicabilità di alcune metriche analizzate, in quanto i soggetti raggiungono delle performance con HRTF selezionata paragonabili ad HRTF individuale. Inoltre confermano un miglioramento statisticamente significativo delle prestazioni di percezione in elevazione con HRTF selezionata rispetto ad HRTF del KEMAR.



# Ringraziamenti

Vorrei scrivere qualche riga per ringraziare tutte le persone che mi hanno accompagnato e permesso di raggiungere questo grande traguardo.

Primo fra tutti vorrei dedicare questa tesi a mia mamma Sonia a cui va il mio più grande ringraziamento. In tutti questi anni mi ha sempre sostenuto con il suo immenso affetto, educandomi al meglio, trasmettendomi i suoi migliori valori e affrontando per me e con me difficoltà che avrebbero fatto crollare chiunque. Senza di lei niente sarebbe stato possibile. A mio padre Mario, che veglia su di me da sempre, le svolte positive inaspettate so che sono opera tua. A mio fratello Alberto, un tempo compagno di litigi ora di esperienze di vita, ha creato in me un forte senso di responsabilità. Ai miei zii che non mi hanno fatto mai mancare nulla e sono stati un punto di riferimento nel portare a termine gli impegni. Ai miei nonni che mi hanno cresciuto, in particolare a Gigi per tutte le passioni che mi ha trasmesso e condiviso fino alla fine. Alla Rodi, per me come una sorella maggiore per tutto l'amore e l'aiuto che ci da ogni giorno.

Un grazie speciale a tutti i miei amici, con cui ho condiviso tutti i miei migliori momenti e i veri valori della vita. Amici che conosco da sempre, amici che ho conosciuto in questi ultimi anni, dall'asilo all'università. Chi mi è stato vicino come un fratello nei momenti bui, chi è stato compagno di avventure di vita, chi ha combattuto al mio fianco, chi ha brindato ai miei successi, chi mi ha prestato un euro, chi mi ha insegnato il rispetto, chi ha migliorato il mio carattere, chi ha creduto in me e mi ha dato fiducia sempre. Ho voluto scrivere brevemente cosa avete fatto per me, non un elenco di nomi per non sminuire nessuno. Questo traguardo è merito di tutti i piccoli passi compiuti grazie alle sicurezze che mi avete trasmesso.

Un ringraziamento va ai miei professori, per tutto quello che ho appreso. In particolare al professor Federico Avanzini e Michele Geronazzo che mi hanno permesso di svolgere l'intero lavoro supportandomi con enorme disponibilità in questi ultimi mesi. Un grazie anche a Fabio, con cui ho condiviso il lavoro di tesi, una marea di progetti e praticamente tutto il percorso di laurea magistrale.



# Indice

<b>Sommario</b>	<b>i</b>
<b>Ringraziamenti</b>	<b>iii</b>
<b>1 Introduzione</b>	<b>1</b>
1.1 Percezione e localizzazione spaziale del suono . . . . .	1
1.1.1 Attributi percettivi . . . . .	1
1.1.2 Principali filtri antropometrici . . . . .	2
1.1.3 Localizzazione di una sorgente sonora . . . . .	4
1.1.4 Sistemi per rendering di audio spazializzato . . . . .	4
1.2 Come ottenere le HRTF . . . . .	5
1.2.1 Cos'è un HRTF . . . . .	5
1.2.2 Limitazioni all'utilizzo . . . . .	6
1.2.3 HRTF individuale . . . . .	7
1.2.4 HRTF generalizzata . . . . .	8
1.2.4.1 DOMISO . . . . .	8
1.2.4.2 Selezione a due passi . . . . .	9
1.2.4.3 Corrispondenza tra misurazioni antropometriche . . . . .	11
1.2.4.4 Tracciamento dei profili del padiglione auricolare . . . . .	12
<b>2 Applicazione per la selezione di HRTF</b>	<b>19</b>
2.1 Interfaccia grafica . . . . .	20
2.2 Metriche utilizzate . . . . .	23
2.2.1 Mismatch in frequenza . . . . .	24
2.2.2 Mismatch antropometrico . . . . .	25
2.3 Classifiche . . . . .	26

<b>3</b>	<b>Analisi delle metriche</b>	<b>31</b>
3.1	Acquisizione dei dati . . . . .	31
3.1.1	Soggetti dell'esperimento . . . . .	31
3.1.2	Tracciamento dei contorni . . . . .	32
3.2	Analisi delle prestazioni . . . . .	33
3.2.1	Classifiche . . . . .	33
3.2.2	Simulazione con modello uditivo . . . . .	35
3.3	Risultati . . . . .	36
3.3.1	Selezione con average mismatch . . . . .	36
3.3.2	Selezione con average rank . . . . .	38
3.3.3	Selezione con Top-5 appearance . . . . .	40
<b>4</b>	<b>Esperimento in VR</b>	<b>47</b>
4.1	Test di localizzazione . . . . .	47
4.1.1	Attrezzatura . . . . .	47
4.1.2	Scena virtuale in Unity . . . . .	48
4.1.3	Suoni . . . . .	49
4.1.4	Condizioni dell'esperimento . . . . .	50
4.1.5	Procedura . . . . .	50
4.1.5.1	Acquisizione delle foto del padiglione . . . . .	50
4.1.5.2	Selezione della HRTF . . . . .	51
4.1.5.3	Ordine dei test . . . . .	51
4.2	Risultati . . . . .	52
4.2.1	Errori in azimuth ed elevazione . . . . .	52
4.2.2	Risultati principali . . . . .	52
4.2.3	Curva di apprendimento . . . . .	55
4.2.4	Eliminazione dei soggetti . . . . .	56
4.2.5	T-Test . . . . .	56
<b>5</b>	<b>Conclusioni</b>	<b>61</b>
5.1	Analisi delle metriche . . . . .	61
5.2	Esperimento in VR . . . . .	62



# Elenco delle figure

1.1	Rendering binaurale del suono: HRTF sinistra e destra . . . . .	2
1.2	Orecchio esterno: (a) padiglione auricolare, (b) canale uditivo . . . . .	3
1.3	Effetti del torace e delle spalle: (a) riflessione, (b) occlusione . . . . .	3
1.4	(a) Coordinate polari verticali, (b) coordinate polari interaurali. . . . .	6
1.5	Configurazione per la misura di HRTF all’Acoustics Research Institute, Au- strian Academy of Sciences, Vienna. Immagine tratta da [16]. . . . .	7
1.6	Metodo DOMISO per la selezione del set personale di HRTF. Immagine tratta da [21]. . . . .	10
1.7	Front-Back error rate fra set individuale (Own), selezionato (Individualized) e casuale (Away). Analisi tratta dallo studio di Yukio Iwaya [21]. . . . .	10
1.8	Misure del padiglione auricolare nel database CIPIC. Immagine tratta da [1]. . .	12
1.9	La HRIR visualizzata come una nuvola di punti per l’orecchio destro del sogget- to 10 nel database CIPIC per azimuth costante $\theta = 0^\circ$ e per elevazione variabile da $-45^\circ$ a $+230.625^\circ$ . Si possono notare chiaramente l’impulso iniziale (4), la diffrazione di testa, gli effetti del padiglione auricolare (1) e la riflessione di torace (2) e ginocchia (3). Immagine tratta da [25]. . . . .	13
1.10	Modello basato su riflessioni per i notch in frequenza causati dal padiglione auricolare. $\phi$ è l’angolo di incidenza dove l’onda diretta viene riflessa dalla conca. La distanza di riflessione è $d$ . Immagine tratta da [25]. . . . .	14
1.11	Anatomia del padiglione e i cinque profili principali per la procedura di sele- zione. Immagine tratta da [28]. . . . .	16
1.12	Esempio di punto focale e dei tre principali contorni del padiglione auricolare. Immagine tratta da [20]. . . . .	16
2.1	Nuvola di punti per ogni posizione registrata per ogni soggetto CIPIC (a) fron- talmente e (b) lateralmente. Immagine tratta da [1]. . . . .	20
2.2	Finestra principale dell’applicazione. . . . .	22

2.3	Finestra per l'impostazione del Pixel to Meter. . . . .	22
2.4	(a) Finestra per il tracciamento delle misure antropometriche: tracciamento dell'altezza dell'orecchio. (b) Finestra per il tracciamento dei contorni dell'orecchio: Tracciamento del contorno $C_1$ . . . . .	24
2.5	Classifica di average mismatch, contorno $C_1$ , riflessione negativa per il CIPIC ID 27. Asse x: CIPIC ID; asse y: mismatch. . . . .	27
2.6	Classifica di average rank, contorno $C_1$ , riflessione negativa per il CIPIC ID 27. Asse x: CIPIC ID; asse y: rank medio. . . . .	27
2.7	Classifica di top-M appearance, contorno $C_1$ , riflessione negativa per il CIPIC ID 27. Asse x: CIPIC ID; asse y: numero di occorrenze. . . . .	28
2.8	Classifica di mismatch dell'altezza esterna dell'orecchio per il CIPIC ID 27. Asse x: CIPIC ID; asse y: mismatch(cm). . . . .	28
3.1	Schema del procedimento di validazione con le predizioni del modello uditivo.	32
3.2	Tracciamenti di contorni e canali uditivi per il soggetto CIPIC 60. In bianco il contorno $C_1$ , in celeste $C_2$ e in rosso $C_3$ . I punti bianchi sono i punti focali del canale uditivo. . . . .	33
3.3	Classifiche di Top-M Appearance al variare di $M$ . (a,b,c,d) $M = 3, 5, 7, 10$ contorno $C_1$ riflessione negativa. . . . .	34
3.4	Prestazioni medie con HRTF selezionata con mismatch minore, HRTF individuale e HRTF del KEMAR. . . . .	38
3.5	Prestazioni medie con HRTF selezionata con rank migliore, HRTF individuale e HRTF del KEMAR. . . . .	40
3.6	Prestazioni medie con HRTF selezionata con più occorrenze nella classifica top 5 appearance, HRTF individuale e HRTF del KEMAR. . . . .	42
3.7	Tracce spettrali di $f_0^{(k,n)}$ per i tracciamenti multipli di $C_1$ /punto focale per il soggetto CIPIC 60. In grigio scuro le riflessioni negative, in grigio chiaro le positive. $F_0$ della HRTF individuale e best Top-5 per la riflessione negativa (linea nera continua) e HRTF best Top-5 per la riflessione positiva (linea nera puntinata). . . . .	43
3.8	Tracce spettrali di $f_0^{(k,n)}$ per i tracciamenti multipli di $C_2$ /punto focale per il soggetto CIPIC 60. In blu le riflessioni negative, in azzurro le positive. $F_0$ della HRTF individuale (linea nera continua), HRTF best Top-5 (linea nera tratteggiata) per la riflessione negativa e HRTF best Top-5 per la riflessione positiva (linea nera puntinata). . . . .	44

3.9	Tracce spettrali di $f_0^{(k,n)}$ per i tracciamenti multipli di $C_3$ /punto focale per il soggetto CIPIC 60. In rosso le riflessioni negative, in arancione le positive. $F_0$ della HRTF individuale (linea nera continua), HRTF best Top-5 (linea nera tratteggiata) per la riflessione negativa e HRTF best Top-5 per la riflessione positiva (linea nera puntinata). . . . .	45
4.1	Condizioni dell'esperimento: il soggetto si posiziona al centro della camera insonorizzata con la luce spenta per tutta la durata del test. . . . .	48
4.2	Panoramica della scena virtuale dell'esperimento in Unity . . . . .	49
4.3	Impulso sonoro di una prova. Durata: 5 secondi. . . . .	50
4.4	Foto scattate al soggetto 1 a tre differenti distanze. . . . .	51
4.5	Regressione lineare in elevazione su tutti i soggetti per ogni condizione. . . . .	53
4.6	Regressione lineare in elevazione sui soggetti per ogni condizione (No soggetti 3, 5 e 10). . . . .	58
5.1	Prestazioni delle HRTF selezionate in ascolto nell'esperimento in realtà virtuale e nella simulazione con modello uditivo. . . . .	63



## Elenco delle tabelle

3.1	Performance della miglior metrica per ogni soggetto con selezione secondo average mismatch. . . . .	37
3.2	Numero di occorrenze per metrica con selezione secondo average mismatch. . .	37
3.3	Prestazioni medie della miglior metrica per ogni soggetto con selezione secondo average rank. . . . .	39
3.4	Numero di occorrenze per metrica con selezione secondo average rank. . . . .	39
3.5	Prestazioni medie della miglior metrica per ogni soggetto con selezione secondo top 5 appearance. . . . .	41
3.6	Numero di occorrenze per metrica con selezione secondo top 5 appearance. . .	41
4.1	Sequenze del latin square . . . . .	51
4.2	Errore medio in azimuth (giallo) ed errore medio in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. . . . .	53
4.3	Deviazione standard in azimuth (giallo) e deviazione standard in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. . . . .	54
4.4	Numero di front-back confusion per ogni condizione (sinistra). Pendenze della regressione lineare in elevazione per ogni condizione (destra). . . . .	55
4.5	AVG azimuth error (yellow) and AVG elevation (green) of the VR experiment test for each subject and number of session . . . . .	55
4.6	Errore medio in azimuth (giallo) ed errore medio in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. Rimossi i soggetti 3, 5 e 10. . . . .	56
4.7	Deviazione standard in azimuth (giallo) e deviazione standard in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. Rimossi i soggetti 3, 5 e 10. . . . .	57

4.8	Numero di front-back confusion per ogni condizione (sinistra). Pendenze della regressione lineare in elevazione per ogni condizione (destra). Rimossi i soggetti 3, 5 e 10. . . . .	57
4.9	Paired t-test fra le condizioni A e B. Errori in azimuth. Rimossi i soggetti 3, 5, 10.	58
4.10	Paired t-test fra le condizioni A e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.	59
4.11	Paired t-test fra le condizioni B e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.	59
4.12	Paired t-test fra le condizioni A e B. Errori in elevazione. Rimossi i soggetti 3, 5, 10. . . . .	59
4.13	Paired t-test fra le condizioni A e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.	60
4.14	Paired t-test fra le condizioni B e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.	60



# Capitolo 1

## Introduzione

### 1.1 Percezione e localizzazione spaziale del suono

Argomenti fondamentali per la comprensione del lavoro svolto in questa tesi sono i concetti relativi alla percezione spaziale del suono. In questa sezione sono presentati gli attributi percettivi con alcuni esempi, le principali caratteristiche fisiche del corpo umano che si comportano come un filtro per il suono ed il modo in cui un soggetto lo percepisce con le sue caratteristiche.

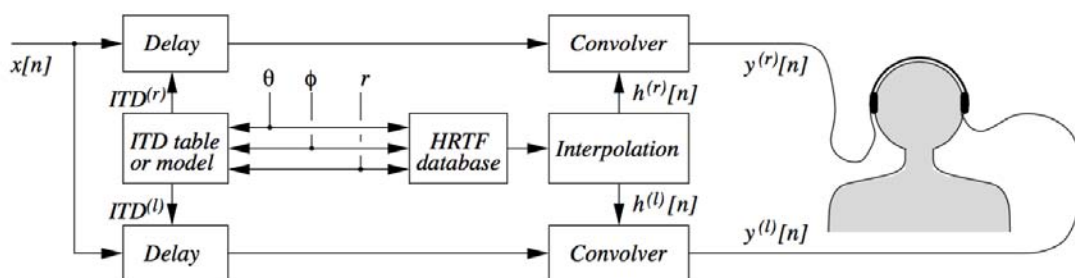
#### 1.1.1 Attributi percettivi

La percezione dell'ascoltatore degli eventi uditivi è correlata ai seguenti gruppi di attributi percettivi:

1. **Temporali:** come ritmo, durata e riverbero.
2. **Qualitativi:** come intensità, intonazione e timbro.
3. **Spaziali:** come direzione, distanza e percezione spaziale.

Gli attributi del terzo gruppo sono influenzati dalle caratteristiche fisiche del corpo umano dell'ascoltatore e l'analisi svolta in questa tesi è focalizzata proprio su di essi. La presenza del corpo è responsabile della creazione di due effetti binaurali rilevanti chiamati *Interaural Time Difference* (ITD) ed *Interaural Level Difference* (ILD) e di un effetto monoaurale come la colorazione spettrale del suono causata dal filtro dell'orecchio esterno. Questi effetti possono essere assegnati nelle funzioni di trasferimento (HRTF) relative rispettivamente all'orecchio sinistro e destro. Grazie alle HRTF individuali per un soggetto specifico è possibile renderizzare e simulare un filtro personale per la fruizione dei contenuti sonori tramite cuffie: la sintesi binaurale.





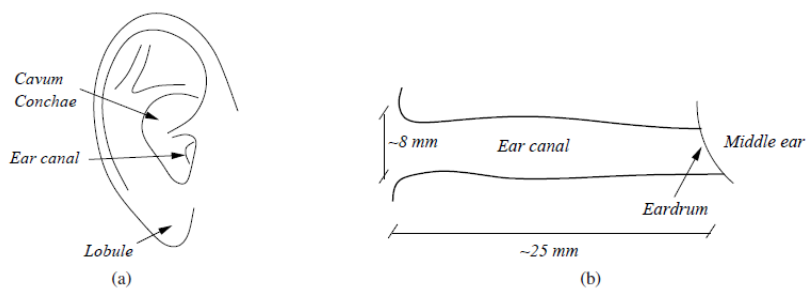
**Figura 1.1:** Rendering binaurale del suono: HRTF sinistra e destra

Il sistema completo chiamato *Virtual Auditory Display* (VAD)<sup>1</sup> utilizza per ciascun segnale una coppia di HRTF basate sulla relazione fra sorgente ed ascoltatore, ed una coppia di *Headphone Transfer Functions* (HpTFs) specifiche per le proprietà acustiche del dispositivo di riproduzione utilizzato. Per questo motivo generalmente si fa riferimento a più set di HRTF (sinistra e destra) adattate alle differenti sorgenti, ascoltatori e contesto di simulazione.

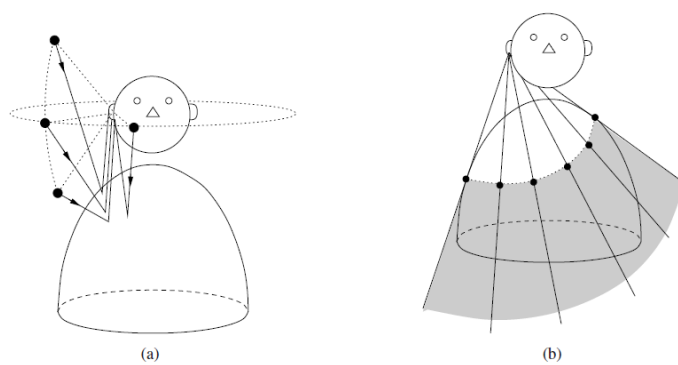
### 1.1.2 Principali filtri antropometrici

- **Testa:** è un oggetto rigido fra le orecchie ed agisce come un ostacolo per la propagazione del suono ed è la causa dei due principali effetti. *Interaural Time Difference* (ITD) a causa della distanza che deve percorrere il suono dall'orecchio più vicino alla sorgente a quello più lontano. *Interaural Level Difference* (ILD) in quanto la testa crea un'occlusione ed attenua l'intensità sonora percepita dall'orecchio mascherato.
- **Orecchio esterno:** è composto da padiglione auricolare e canale uditivo fino al timpano. Dopo questa membrana c'è l'orecchio medio ed interno. Si comporta come un'antenna acustica in grado di amplificare o attenuare alcune frequenze.
- **Torace e spalle:** contribuiscono con ulteriori riflessioni che si aggiungono al suono diretto oltre all'effetto occlusivo, come per la testa, per i suoni provenienti dal basso. Grazie al tronco ed alle spalle, un orecchio percepisce l'impulso sonoro iniziale, seguito da una serie di impulsi successivi con un ritardo direttamente proporzionale all'elevazione della sorgente (vedi fig 1.3).

<sup>1</sup>E' un sistema per generare suoni spazializzati e trasmetterli ad un ascoltatore



**Figura 1.2:** Orecchio esterno: (a) padiglione auricolare, (b) canale uditivo



**Figura 1.3:** Effetti del torace e delle spalle: (a) riflessione, (b) occlusione

### 1.1.3 Localizzazione di una sorgente sonora

- **Azimuth:** come abbiamo visto in precedenza la percezione orizzontale è influenzata da due effetti chiave: ITD e ILD. Le persone individuano la direzione orizzontale di un'onda sonora sfruttando l'effetto ITD più facilmente sulle basse frequenze. Questo perchè a frequenze  $f > 1600Hz$  la lunghezza d'onda del suono è maggiore della distanza fra le orecchie rendendo pertanto impossibile riconoscerne la differenza di fase. In questo caso subentra l'effetto ILD che è più marcato per frequenze elevate. Entrambi sono combinati per ottenere la percezione azimutale in tutta la gamma delle frequenze udibili.
- **Elevation:** ci sono alcune superfici coniche chiamate *coni di confusione* nelle quali le sorgenti sonore provenienti da quell'area producono valori quasi identici di ITD e ILD. In queste zone c'è un'alta ambiguità percettiva fra suoni provenienti frontalmente e posteriormente in quanto i due valori frontali e posteriori di ITD e ILD sono molto simili<sup>2</sup>. Analizzando i notch nello spettro delle frequenze si può identificare l'elevazione della sorgente sonora. Il padiglione auricolare causa delle riflessioni che causano interferenze distruttive a determinate frequenze, causandone un'attenuazione che si può notare tramite analisi spettrale. Questi notch forniscono informazioni riguardo alla dimensione verticale.
- **Distanza:** senza ulteriori informazioni l'intensità della sorgente sonora percepita dall'ascoltatore è il parametro maggiormente indicativo per la misura della distanza. Questo parametro è detto *loudness*. Un altro aspetto collegato a questa misura è la familiarità che abbiamo con un suono nell'identificarne la provenienza, ad esempio un tuono sappiamo perfettamente che proviene da una distanza maggiore rispetto alla caduta di un bicchiere. Possiamo quindi stimare la distanza di un suono basata sull'esperienza ad un successivo ascolto.

### 1.1.4 Sistemi per rendering di audio spazializzato

- **Stereo:** è il sistema più semplice per riprodurre un suono spazializzato. Il segnale è diviso in due, canale sinistro e destro, e ciascuno è indirizzato al relativo altoparlante creando una differenza percezione della sorgente sonora. Se l'ascoltatore è correttamente posizionato frontalmente può percepire una "sorgente fantasma" in mezzo ai due canali e consente l'ascolto di tutte le possibili combinazioni delle posizioni nell'arco di 180 gradi da sinistra a destra.

---

<sup>2</sup>A causa delle asimmetrie facciali i valori si discostano solo leggermente

- **Multicanale:** è un sistema composto da più altoparlanti posizionati in più direzioni per riprodurre più fedelmente la direzione della sorgente sonora, sia orizzontalmente che verticalmente. Il risultato è ottenuto mediante più altoparlanti di notche dimensioni posizionati strategicamente per direzionare le frequenze più acute e di pochi altoparlanti di dimensioni maggiori per le frequenze più basse non direzionate.

Sia lo **stereo** che il **multicanale**, basati su altoparlanti, soffrono dell'effetto *cross-talk*, in quanto il suono emesso da una cassa è udito da entrambe le orecchie. Tuttavia esistono alcune tecniche per cancellare questo fenomeno, utilizzate in fase di pre-processing, che aiutano a ridurre e teoricamente cancellare il segnale sonoro che arriva all'altro orecchio.

- **Cuffie:** spesso usate negli esperimenti con HRTF, grazie ai vantaggi rispetto ai sistemi sopra elencati, sono differenti proprio per il modo in cui operano. Sono notche e portatili, ogni canale ha un segnale distinto e separato che aiuta a riprodurre la tridimensionalità del suono grazie ad algoritmi specifici ed eliminano il riverbero della stanza, creati invece dagli altoparlanti. Tuttavia sono poco confortevoli da indossare per periodi prolungati e potrebbero non avere una risposta in frequenza piatta, compromettendo la renderizzazione della spazializzazione del suono.

## 1.2 Come ottenere le HRTF

### 1.2.1 Cos'è un HRTF

Una *Head-Related Transfer Function* (HRTF) è la trasformata in frequenza della *Head-Related Impulse Response* (HRIR) che è la risposta impulsiva del timpano prodotta dalla pressione del suono di una sorgente.

La HRTF è una funzione di tre coordinate spaziali e frequenze, infatti il filtro di un HRTF è unico per ogni posizione ed angolo di incidenza. In questo contesto le coordinate angolari orizzontali e verticali sono chiamate *azimuth* e *elevation* denotate da  $\theta$  e  $\phi$ , mentre il raggio da  $r$ . La notazione per indicare una HRTF è la seguente:

$$H^{(L,R)}(r, \theta, \phi, \omega) \quad (1.1)$$

Dove ( $L$ ) ed ( $R$ ) si riferiscono rispettivamente all'HRTF relativa all'orecchio sinistro e destro. Se  $r > 1$  si può assumere di essere a distanza sufficiente da rendere la HRTF indipendente da  $r$

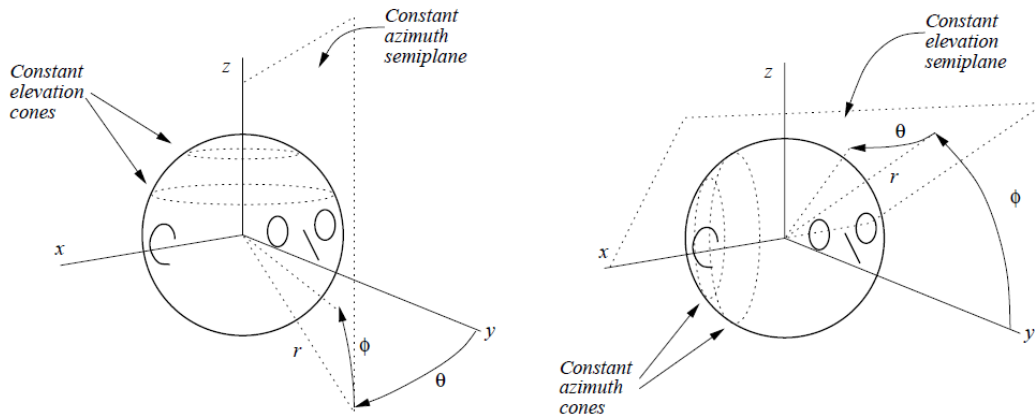
e quindi si utilizza la seguente notazione:

$$H^{(L,R)}(\theta, \phi, \omega) \quad (1.2)$$

Formalmente si può definire l'HRTF ad un orecchio come il rapporto dipendente dalla frequenza fra il *Sound Pressure Level* (SPL)  $\Phi^{(L,R)}(\theta, \phi, \omega)$  con il SPL che verrebbe prodotto in campo aperto nel centro della testa  $\Phi_f(\omega)$

$$H^{(X)}(\theta, \phi, \omega) = \frac{\Phi^{(X)}(\theta, \phi, \omega)}{\Phi_f(\omega)} \quad (1.3)$$

Dove  $X = L, R$ .



**Figura 1.4:** (a) *Coordinate polari verticali*, (b) *coordinate polari interaurali*.

Quando una HRTF non è personalizzata per un ascoltatore la precisione di localizzazione è spesso scarsa e produce vistosi errori, come frequente confusione fra percezione frontale e posteriore e mancanza di profondità del suono [29].

### 1.2.2 Limitazioni all'utilizzo

Una delle principali limitazioni per commercializzazione e diffusione della personalizzazione dell'audio binaurale è la complessità nel creare una HRTF su misura capace di rappresentare fedelmente tutti gli effetti fisici che modificano il suono per una percezione il più naturale possibile. La realizzazione di una HRTF personale per tutte le possibili direzioni richiede speciali strumenti e lunghi tempi di misurazione. Risulta spesso un compito oneroso da eseguire per ogni singolo soggetto dell'esperimento o di un test. Per questo motivo molto spesso si pren-



**Figura 1.5:** Configurazione per la misura di HRTF all'Acoustics Research Institute, Austrian Academy of Sciences, Vienna. Immagine tratta da [16].

dono in considerazione tecniche alternative, come l'assegnare ad un soggetto un set di HRTF personalizzate, ricavate ad esempio da una persona già presente in un database, ma non create individualmente. Un compromesso fra qualità e costo della misurazione [19] [9].

### 1.2.3 HRTF individuale

La configurazione standard per la misura di una HRTF individuale è una camera anecoica con una serie di altoparlanti montati su un supporto sferico, di raggio superiore ad un metro per evitare gli effetti di prossimità, ad intervalli prestabiliti di azimuth ed elevazione. Il soggetto, seduto al centro di questa sfera, ha dei microfoni nelle orecchie. A questo punto le HRIR sono misurate riproducendo degli impulsi sonori e registrando le risposte al segnale su ogni orecchio per ciascuna posizione.

L'obiettivo è di estrarre un set di HRTF per ogni utente che vuole avere la propria funzione di trasferimento personale. Ci sono numerosi aspetti delicati in questo, primo fra tutti i movimenti indesiderati del soggetto che invalidano la precisione della misura e le poche fissate posizioni, che hanno le proprie caratteristiche, utilizzate per creare il set di HRTF. Inoltre la ripetibilità della misurazione è ancora un argomento dibattuto [2].

## 1.2.4 HRTF generalizzata

Una HRTF generalizzata è scelta fra un set di HRTF già misurate invece che utilizzare la procedura per il calcolo individuale. Infatti questo metodo è basato sulla corrispondenza fra HRTF del soggetto esterno ed HRTF dei soggetti registrati precedentemente nel database. L'attenzione è focalizzata sul metodo con cui viene assegnato un particolare set di HRTF rispetto ad un altro. I ricercatori stanno trovando diversi modi per affrontare questo problema e sono presenti ad oggi una notevole varietà di alternative che utilizzano strumenti hardware e/o software più o meno dispendiosi.

Il vantaggio principale di questo approccio è che l'utente può essere guidato ad una selezione autonoma del proprio miglior set di HRTF senza bisogno di attrezzature speciali e conoscenze approfondite. Bisogna notare che l'HRTF selezionata con questo metodo non garantisce le stesse performance della propria HRTF. Tuttavia spesso offre prestazioni migliori di una HRTF generica di una testa di un manichino (KEMAR<sup>3</sup>) [7].

Di seguito elenchiamo quattro esempi di approcci per la selezione:

- **Torneo DOMISO<sup>4</sup>** [21] [30]
- **Selezione a due passi** [26]
- **Corrispondenza tra misurazioni antropometriche** [31]
- **Tracciamento dei profili del padiglione auricolare** [20] [25] [28]

### 1.2.4.1 DOMISO

Con questo metodo il soggetto esterno può scegliere la HRTF più appropriata fra le tante presenti in un database con un test di ascolto ad esclusione come se fosse un torneo ad eliminazione diretta. Il database è costruito a partire da differenti soggetti contenendo un totale di 120 set di HRTF, una per ogni ascoltatore. Il metodo è il seguente:

1. Fra tutti i 120 set di HRTF, 32 set sono scelte in modo casuale.
2. Viene preparato un campione di 13 suoni virtuali (convoluzione del rumore rosa con il set di HRTF selezionato, lunghezza pari a un secondo, 16 bit,  $f_s = 48kHz$ ) per ogni 30 gradi in senso antiorario a partire dalla fronte dell'ascoltatore nel piano orizzontale. Ciascun campione è creato con i 32 set di HRTF selezionati precedentemente.

---

<sup>3</sup>Knowles Electronic Manikin for Acoustic Research

<sup>4</sup>Determination method of OptimuM Impulse-response by Sound Orientation

3. La forma dell'orbita creata dai suoni virtuali è mostrata all'ascoltatore come un'illustrazione prima che il test inizi.
4. Gli scontri del torneo sono programmati per 32 orbite riprodotte utilizzando i 32 set di HRTF.
5. In una sessione o scontro, l'ascoltatore sceglie una delle due orbite che meglio rappresenta l'orbita mostrata precedentemente, quella selezionata è la vincente e avanza al confronto successivo.
6. Alla fine degli scontri un set di HRTF vince il torneo ed è selezionato come set personalizzato di HRTF da utilizzare per il soggetto.

In realtà il torneo non è un vero e proprio torneo ad eliminazione diretta ma ha delle notevoli variazioni. In un torneo ad eliminazione diretta è prevista una partita fra i soli vincitori che sono avanzati alla fase successiva, i perdenti vengono eliminati. La variante utilizzata in questo metodo evita l'eliminazione di due forti set di HRTF nelle prime fasi del torneo ripescando alcuni set di HRTF già scartati. Quando un set esce sconfitto da due scontri di fila viene definitivamente rimosso dal gruppo (vedi fig 1.6).

Questo metodo propone numerosi vantaggi rispetto alla misura individuale di una HRTF:

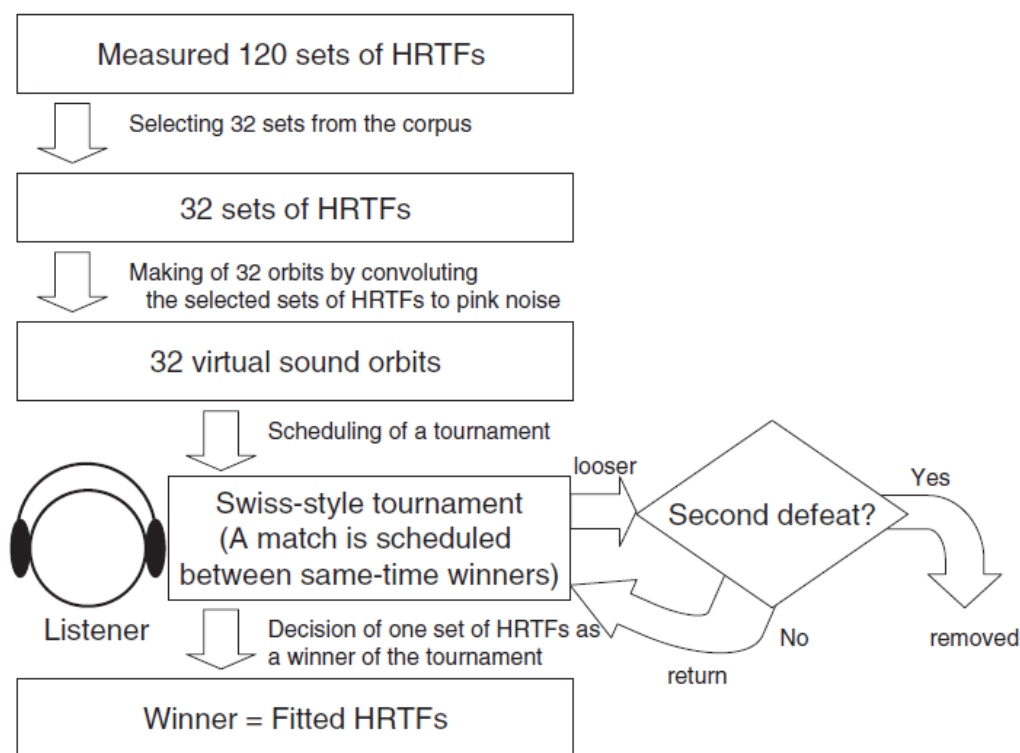
1. **Semplicità:** l'ascoltatore giudica direttamente la similarità dell'orbita percepita. Il suo incarico è molto facile.
2. **Velocità:** il dispendio in tempo per la selezione è molto contenuto (circa 15 min) rispetto alla misurazione e creazione del set di HRTF individuali (circa 2 ore) per ciascun soggetto. Tuttavia grazie alle più recenti tecniche, ideate successivamente alla pubblicazione di questo studio, il tempo necessario alla misurazione di una HRTF individuale richiede da 15 a 30 minuti [23].

Le prestazioni di questa tecnica sono state valutate (vedi fig 1.7) [21] provando che il set di HRTF selezionato tramite metodo DOMISO ha risultati simili al set di HRTF individuale del soggetto ed ha performance decisamente migliori rispetto ad un set di HRTF selezionato in modo casuale, che non avrebbe potuto vincere il torneo.

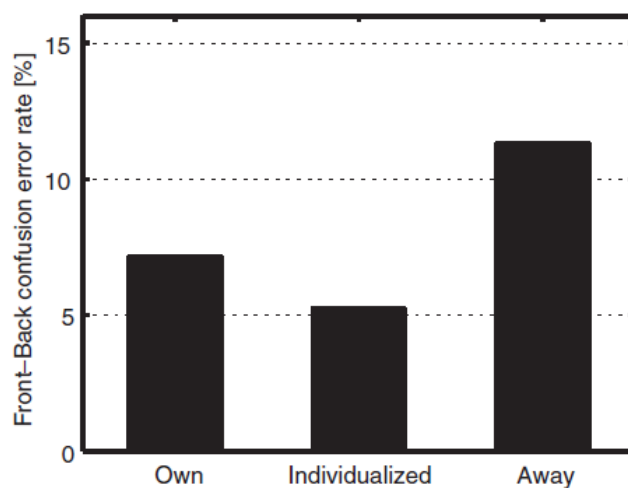
#### 1.2.4.2 Selezione a due passi

Questo metodo si basa su due step conseguenti. Solitamente nel primo passo si seleziona da un pool iniziale di HRTF un sottoinsieme, rimuovendo le peggiori dal punto di vista percettivo.





**Figura 1.6:** Metodo DOMISO per la selezione del set personale di HRTF. Immagine tratta da [21].



**Figura 1.7:** Front-Back error rate fra set individuale (Own), selezionato (Individualized) e casuale (Away). Analisi tratta dallo studio di Yukio Iwaya [21].

Nel secondo si sceglie la miglior corrispondenza fra il soggetto esterno ed il database di HRTF rimasto.

Numerosi ricercatori hanno utilizzato diversi metodi con i loro vantaggi e svantaggi. A titolo esemplificativo è citato lo studio riguardo l'effetto ventriloquismo [26]. Il metodo è diviso in due parti ciascuna svolta in giorni differenti: la selezione delle top-10 HRTF e la vincitrice.

Nella prima parte il soggetto deve ascoltare 49 impulsi sonori di rumore bianco gaussiano di durata 300ms, 5ms di fade in e fade out ed elevazione  $0^\circ$ . Il compito del soggetto è di accettare o rifiutare gli stimoli se non provengono da una traiettoria circolare regolare attorno ad esso, sentendo i suoni come se provenissero dall'esterno della testa. Con questo metodo di decisione sono scelti 10 set di HRTF.

Nella seconda parte, per ognuna delle 10 HRTF selezionate, 16 impulsi sonori simili ai precedenti ma convoluti a  $-45^\circ$ ,  $-15^\circ$ ,  $15^\circ$  e  $45^\circ$  in azimuth e a  $0^\circ$  in elevazione. Come prima il soggetto sceglie il set di HRTF migliore decidendo quali stimoli provengono dall'emisfero frontale e quali nel posteriore premendo un pulsante.

### 1.2.4.3 Corrispondenza tra misurazioni antropometriche

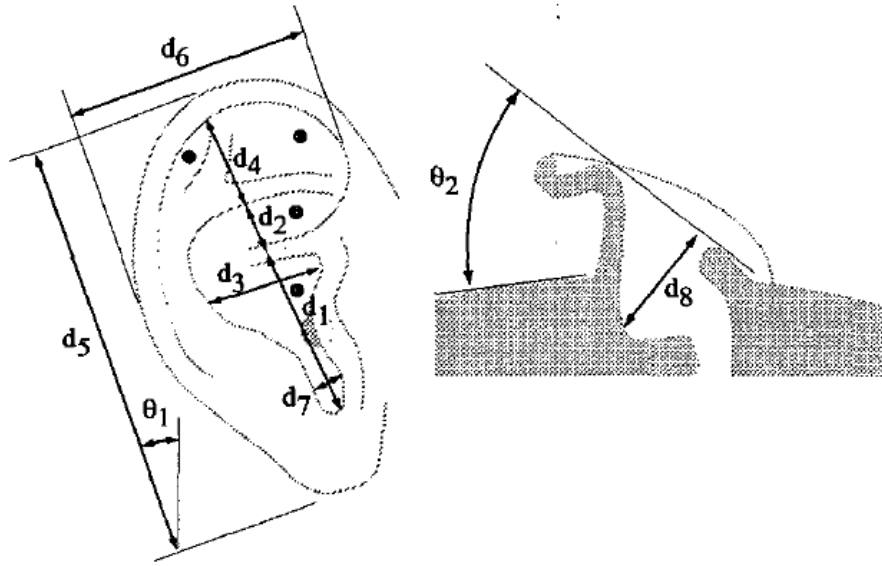
Questo metodo [31] si basa sul miglior match della HRTF nel dominio antropometrico. Questi parametri corrispondono alla forma dell'orecchio esterno utilizzando le misure disponibili, ad esempio confrontando le proprie dimensioni con le sette distanze presenti nel database CIPIC<sup>5</sup>.

La procedura inizia con l'acquisizione di foto digitali di: orecchio destro e sinistro, fronte e retro del corpo. I sette fattori da comparare sono i seguenti (fig 1.8):

- $d_1$ : altezza della conca inferiore
- $d_2$ : altezza della conca superiore
- $d_3$ : larghezza della conca
- $d_4$ : altezza della fossa triangolare
- $d_5$ : altezza del padiglione auricolare
- $d_6$ : larghezza del padiglione auricolare
- $d_7$ : larghezza dell'incisura intertragica

---

<sup>5</sup>Vedi capitolo seguente



**Figura 1.8:** Misure del padiglione auricolare nel database CIPIC. Immagine tratta da [1].

Una volta calcolati, la corrispondenza fra soggetto esterno e il  $k$ -esimo presente nel database viene calcolata con la seguente formula:

$$E^k = \sum_{i=1}^7 \frac{(\tilde{d}_i - d_i^k)^2}{\sigma_i^2} \quad (1.4)$$

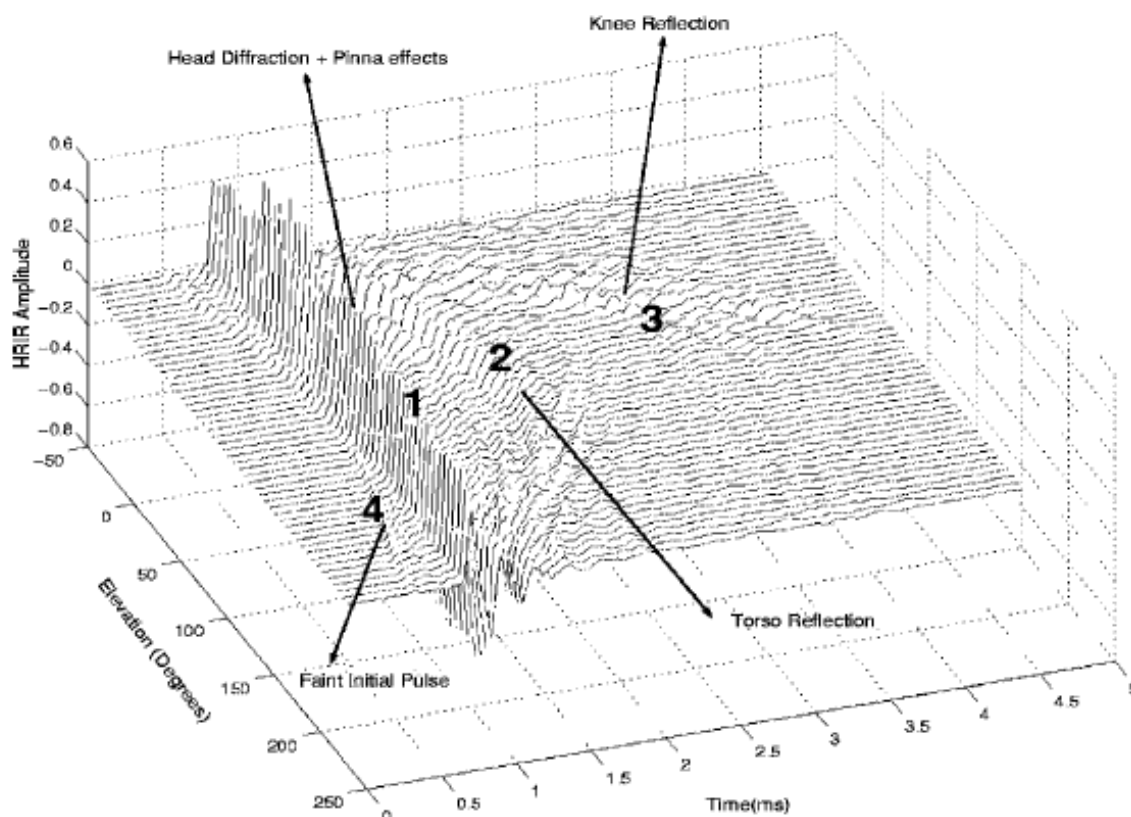
dove  $\tilde{d}_i$  con  $i \in \{1, 2, \dots, 7\}$  è il valore dell' $i$ -esimo parametro e  $d_i^k$  è il valore dello stesso parametro nel  $k$ -esimo soggetto del database CIPIC.  $\sigma_i^2$  è la varianza del  $i$ -esimo parametro fra tutti i soggetti nel database.

Ottenute queste computazioni, la miglior corrispondenza delle distanze fra tutti i fattori è il termine di indice:

$$k = \underset{k}{\operatorname{argmin}} E^k \quad (1.5)$$

#### 1.2.4.4 Tracciamento dei profili del padiglione auricolare

Un altro approccio al problema della selezione della HRTF consiste nel tracciare i profili antropometrici del padiglione auricolare e di applicare un modello riflessivo in base ai raggi descritti. Le distanze variabili dall'ingresso del canale uditivo ai bordi del padiglione auricolare definiscono riflessi che generano dei notch nello spettro di frequenza nella HRTF. Per semplicità si può assumere che il suono si comporti come un raggio più che come un'onda. Questa semplificazione è corretta fino a che la lunghezza d'onda del suono è notchla se relazionata alla



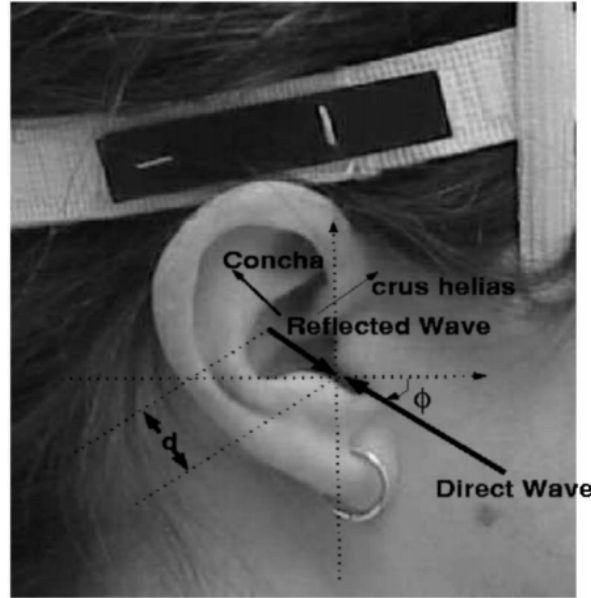
**Figura 1.9:** La HRIR visualizzata come una nuvola di punti per l'orecchio destro del soggetto 10 nel database CIPIC per azimuth costante  $\theta = 0^\circ$  e per elevazione variabile da  $-45^\circ$  a  $+230.625^\circ$ . Si possono notare chiaramente l'impulso iniziale (4), la diffrazione di testa, gli effetti del padiglione auricolare (1) e la riflessione di torace (2) e ginocchia (3). Immagine tratta da [25].

dimensioni delle superfici coinvolte. In questo caso i notch spettrali compaiono a frequenze  $f > 6kHz$ .

L'applicazione creata e descritta nel capitolo successivo utilizza principalmente questo metodo per la selezione della migliore HRTE.

### Modello basato su riflessioni

La creazione di una HRTE utilizzando le riflessioni causate dal padiglione auricolare si basa sulla fisica che influenza il suono percepito dall'orecchio interno. In particolare la distanza fra il punto di riflessione del padiglione auricolare ed il suo punto focale può essere descritta in



**Figura 1.10:** Modello basato su riflessioni per i notch in frequenza causati dal padiglione auricolare.  $\phi$  è l'angolo di incidenza dove l'onda diretta viene riflessa dalla conca. La distanza di riflessione è  $d$ . Immagine tratta da [25].

questo modo:

$$d_c(\phi) = \frac{ct_d(\phi)}{2} \quad (1.6)$$

dove  $t_d(\phi)$  dipende dall'elevazione ed è il ritardo temporale fra l'onda sonora diretta e l'onda sonora riflessa.  $c$  è la velocità del suono.

La corrispondente frequenza del notch dipende dal segno della riflessione. Assumiamo per il momento che il coefficiente di riflessione sia positivo, per interferenza il notch risulta su tutte le frequenze che hanno differenza di fase fra onda diretta e riflessa pari a  $\pi$ :

$$f_n(\phi) = \frac{2n + 1}{2t_d(\phi)} = \frac{c(2n + 1)}{4d_c(\phi)} \quad (1.7)$$

dove  $n \in \mathbb{N}$ .

Il primo notch di frequenza  $f_0(\phi)$  si trova quando  $n = 0$  con il seguente risultato:

$$f_0(\phi) = \frac{c}{4d_c(\phi)} \quad (1.8)$$

In realtà, precedenti studi [27] dimostrano che almeno l'80% dei soggetti presenti nel database CIPIC mostrano una chiara riflessione con coefficiente negativo nella HRIR. Con questa assunzione si può creare un modello più accurato, dove i notch si trovano ad una differenza di fase di

un'intera lunghezza d'onda. L'equazione che risulta è la seguente:

$$f_n(\phi) = \frac{n+1}{t_d(\phi)} = \frac{c(n+1)}{2d_c(\phi)} \quad (1.9)$$

dove  $n \in \mathbb{N}$  e

$$f_0(\phi) = \frac{c}{2d_c(\phi)} \quad (1.10)$$

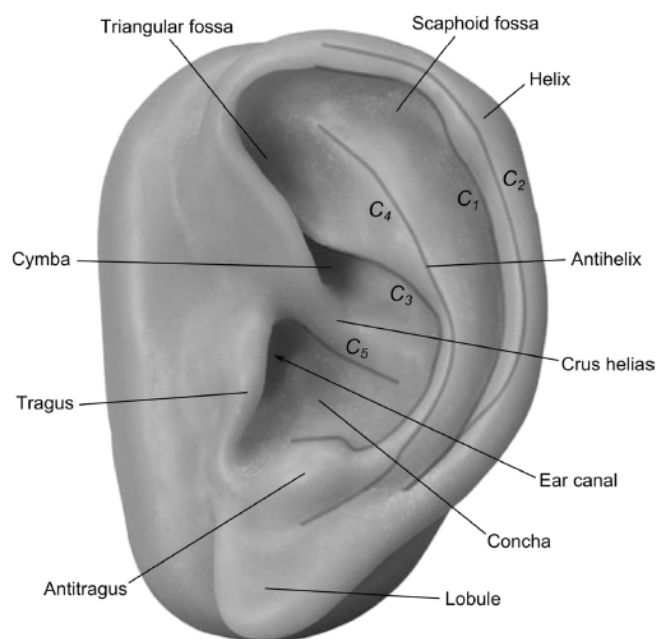
### Profili del padiglione auricolare

Il secondo elemento cruciale di questo metodo è il padiglione auricolare ed i suoi profili. In un orecchio possiamo identificare cinque differenti contorni secondo lo studio [28] (vedi fig. 1.11)

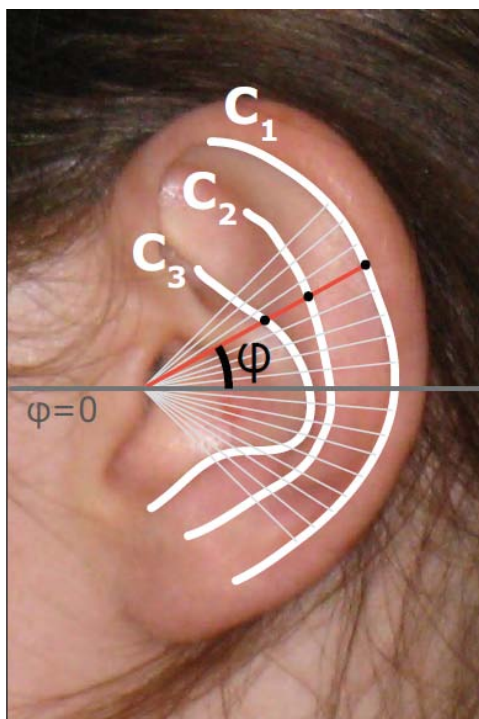
1. *Helix Border*: bordo interno dell'elice ( $C_1$ )
2. *Helix inner wall*: bordo esterno dell'elice ( $C_2$ )
3. *Concha outer border*: bordo esterno della conca ( $C_3$ )
4. *Antihelix and concha inner wall*: antielice e bordo inferiore della conca, praticamente in mezzo al padiglione ( $C_4$ )
5. *Crus helias inferior surface*: radice dell'elice e superficie sottostante ( $C_5$ )

Dopo alcuni esperimenti, i risultati [28] hanno provato che tre notch di frequenza sono associati a tre principali riflessioni associate al bordo di conca, elice ed antielice. In particolare i volumi descritti dalla geometria dell'orecchio influiscono in maniera determinante nel posizionare picchi e notch nella HRTF [24]. Il modello precedente illustrato può quindi essere semplificato a soli tre contorni (vedi fig. 1.12):

- Primo contorno: *helix border* ( $C_1$ )
- Secondo contorno: *antihelix and concha inner wall* ( $C_2$ )
- Terzo contorno: *concha outer border* ( $C_3$ )



**Figura 1.11:** Anatomia del padiglione e i cinque profili principali per la procedura di selezione. Immagine tratta da [28].



**Figura 1.12:** Esempio di punto focale e dei tre principali contorni del padiglione auricolare. Immagine tratta da [20].

### Mismatch fra i soggetti

Sfruttando i concetti descritti precedentemente, definiamo una funzione di mismatch fra le HRTF di due soggetti con la seguente equazione:

$$m = \frac{1}{n} \sum_{i=1}^n \frac{w_i}{N_\phi} \sum_{\phi} \frac{|f_0^i(\phi) - F_0^i(\phi)|}{F_0^i(\phi)} \quad (1.11)$$

dove  $n$  è il numero di contorni considerati,  $w_i$  è una combinazione convessa dei pesi di ogni  $C_i$ ,  $\phi \in [-45^\circ, +45^\circ]$  è l'angolo di elevazione frontale dove è presente l' $i$ -esimo notch in frequenza nella corrispondente HRTF e  $N_\phi \in \{1, \dots, n\}$  è il numero dei campioni considerati per ogni differente elevazione per normalizzare la sommatoria, quando  $f_0^i(\phi) \neq 0 \wedge F_0^i(\phi) \neq 0$ . Nello studio  $\phi$  assume 17 differenti valori per ogni  $\frac{90^\circ}{17-1} = 5.625^\circ$  di elevazione. Questo intervallo e passo di campionamento è sufficiente a caratterizzare i principali notch in frequenza.  $f_0^i(\phi) = \frac{c}{2d_c(\phi)}$  sono le frequenze ricavate dai contorni del soggetto esterno mentre  $F_0^i$  sono i notch estratti dalle HRTF presenti nel database grazie allo *structural decomposition algorithm* [18].

Nel capitolo successivo verranno sfruttati questi concetti per creare un'applicazione Matlab in grado di effettuare i confronti e di trovare la corrispondenza migliore fra un soggetto e le HRTF presenti nel database CIPIC.





## Capitolo 2

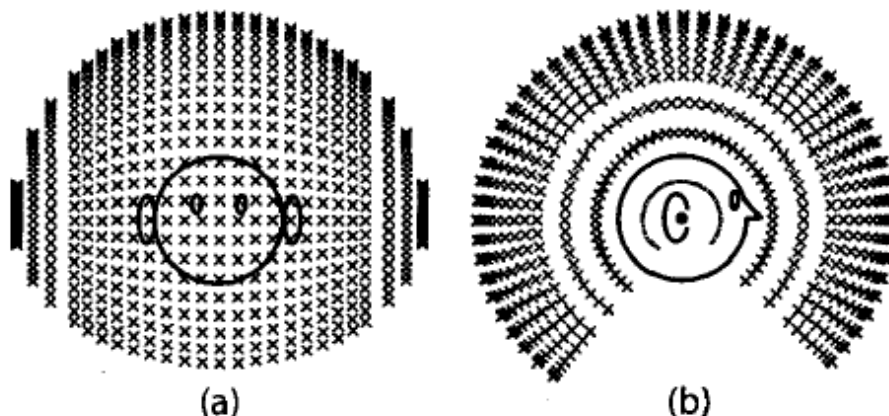
# Applicazione per la selezione di HRTF

Il primo obiettivo del lavoro è stata la creazione di un'applicazione in grado di effettuare un confronto fra HRTF riadattando completamente un programma Matlab già esistente, utilizzata in numerosi studi psicofisici svolti dal gruppo CSC-SMC (si citano a titolo di esempio [12, 10, 13, 11]). In tutti questi studi un utente esperto è stato incaricato di processare le foto dei soggetti coinvolti. Lo scopo di questa nuova applicazione è di fornire a qualsiasi utente uno strumento semplice che permetta, in autonomia e senza particolari attrezzature, la selezione di un set di HRTF personalizzate migliori di una assegnata in modo generico (ad esempio HRTF del KEMAR), partendo dal tracciamento dei contorni dell'orecchio e le misure principali. Il database di HRTF utilizzato per i confronti è il CIPIC.

### CIPIC HRTF Database

Il database CIPIC [1], creato al U.C. Davis CIPIC Interface Laboratory, è un database pubblico di HRTF ad alta risoluzione spaziale con un totale di 45 soggetti. La release contiene le HRIRs per ciascun soggetto a 25 differenti azimuth e 50 differenti elevazioni per un totale di 1250 direzioni (vedi fig. 2.1). In aggiunta sono presenti foto e misure antropometriche (vedi fig 1.8) di alcuni soggetti, utili per validare i risultati dell'applicazione.

Non è stato possibile effettuare il confronto con la totalità dei 45 soggetti a causa dell'insufficienza di dati in elevazione nelle HRTF di alcuni contorni. Nell'applicazione ne sono stati utilizzati 37 per i confronti. Per ovviare a questo inconveniente si sarebbe potuto interpolare i dati mancanti delle elevazioni nelle tracce dei notch spettrali, completandole con dei valori artificiali senza però la garanzia di corrispondenza alla realtà, motivo per cui è stato scelto di scartarli.



**Figura 2.1:** Nuvola di punti per ogni posizione registrata per ogni soggetto CIPIC (a) frontalmente e (b) lateralmente. Immagine tratta da [1].

## 2.1 Interfaccia grafica

### Dati sui soggetti

L'interfaccia principale mostra una panoramica dei dati del soggetto (vedi fig. 2.2). Nella parte a sinistra è visualizzato un elenco con tutti i soggetti inseriti con una breve descrizione di: nome, cognome, età e sesso. Con i pulsanti "Update Subject" e "Delete Subject" si possono modificare i campi o rimuovere un elemento. Tramite il pulsante "Add Subject" è possibile inserire un nuovo soggetto completando correttamente i campi relativi alle informazioni. A questo punto l'utente può inserire una foto, con il pulsante "Choose Ear Image", che servirà da maschera per guidare i tracciamenti. L'applicazione consente inoltre di effettuare semplici modifiche alle foto come ritaglio, rotazione e ridimensionamento tramite il menù a tendina "Image" presente nella parte superiore.

### Configurazioni

Nella parte destra sono presenti i seguenti parametri di configurazione:

- **Contours:** numero di linee ( $N$ ) che si andranno a tracciare seguendo il profilo di ciascun contorno  $C_1, C_2, C_3$ .
- **Ear Canals:** numero di punti ( $K$ ) che si andranno a segnare come punto focale del canale uditivo.
- **M:** parametro per la metrica *Appearance in M-Rank*. (Vedi sez. 2.3)

- **C1 Contour:** se attivo permette di tracciare ed elaborare il contorno  $C_1$ .
- **C2 Contour:** se attivo permette di tracciare ed elaborare il contorno  $C_2$ .
- **C3 Contour:** se attivo permette di tracciare ed elaborare il contorno  $C_3$ .
- **External Ear Length:** se attivo permette di tracciare ed elaborare la misura antropometrica dell'altezza del padiglione auricolare.
- **Internal Ear Length:** se attivo permette di tracciare ed elaborare la misura antropometrica dell'altezza interna del padiglione auricolare.
- **Average Mismatch:** elabora le classifiche del mismatch medio delle HRTF per ogni contorno selezionato.
- **Average Ranked:** elabora le classifiche del rank medio delle HRTF per ogni contorno selezionato.
- **Appearance in M-Rank:** elabora le classifiche del numero di occorrenze delle HRTF per ogni contorno selezionato.

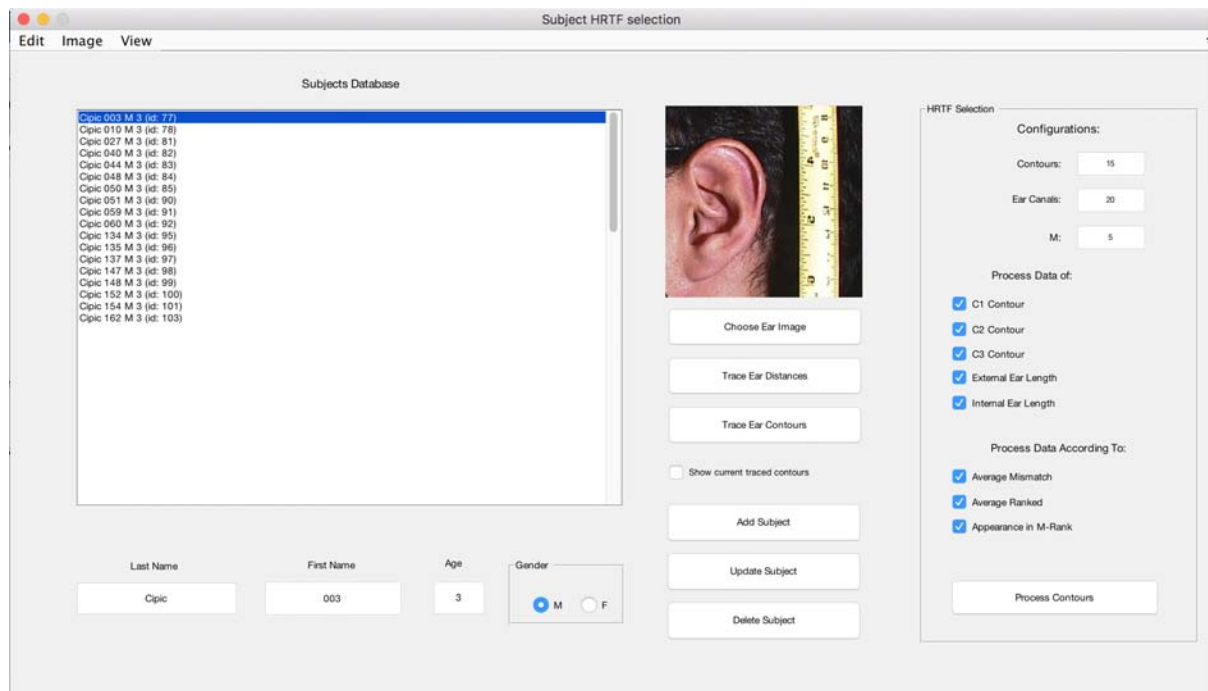
Nella parte centrale i due pulsanti "*Trace Ear Distances*" e "*Trace Ear Contours*" consentono rispettivamente di tracciare le distanze antropometriche ed i contorni dell'orecchio, in relazione ai parametri di configurazione selezionati sulla destra. Il flag "*Show current traced contours*" permette alle  $N$  linee tracciate manualmente ( $N$  per ciascun contorno  $C_1, C_2, C_3$ ) di restare visibili durante tutta la procedura. Questa funzione risulta utile quando si vuole cercare di ricoprire un'area con più linee piuttosto che un contorno preciso.

Passo fondamentale prima di procedere ai tracciamenti è l'impostazione del *Pixel to Meter* che è il fattore di conversione fra misura reale (cm) e misura tracciata nella foto (pixel). Una nuova finestra verrà aperta (vedi fig. 2.3) dove si traccia una linea retta e si inserisce il valore in centimetri relativo alla misura fisica.

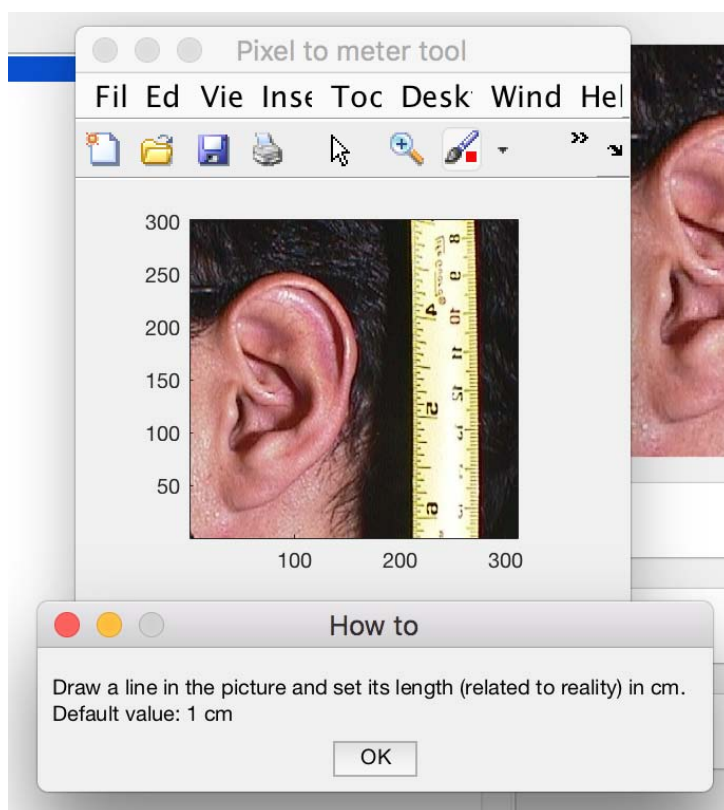
Infine il pulsante *Process Contour* genera le elaborazioni di tutte le classifiche delle HRTF relative al soggetto selezionato.

## Tracciamento delle misure antropometriche

L'applicazione, tramite il pulsante *Trace Ear Distances* consente di inserire due misure principali tracciandole graficamente sulla la foto dell'orecchio. Verranno poi confrontate con i soggetti presenti nel database CIPIC. Le due misure si riferiscono a:



**Figura 2.2:** Finestra principale dell'applicazione.



**Figura 2.3:** Finestra per l'impostazione del Pixel to Meter.

- **Altezza dell'esterno del padiglione:** misura dal punto più alto del contorno  $C_1$  alla base del lobulo. Corrisponde a  $d_5$  nel database CIPIC. Vedi fig. 1.8
- **Altezza dell'interno del padiglione:** misura dall'interno del contorno  $C_1$  alla base della conca. Corrisponde a  $d_1 + d_2 + d_4$  nel database CIPIC.

## Tracciamento dei contorni

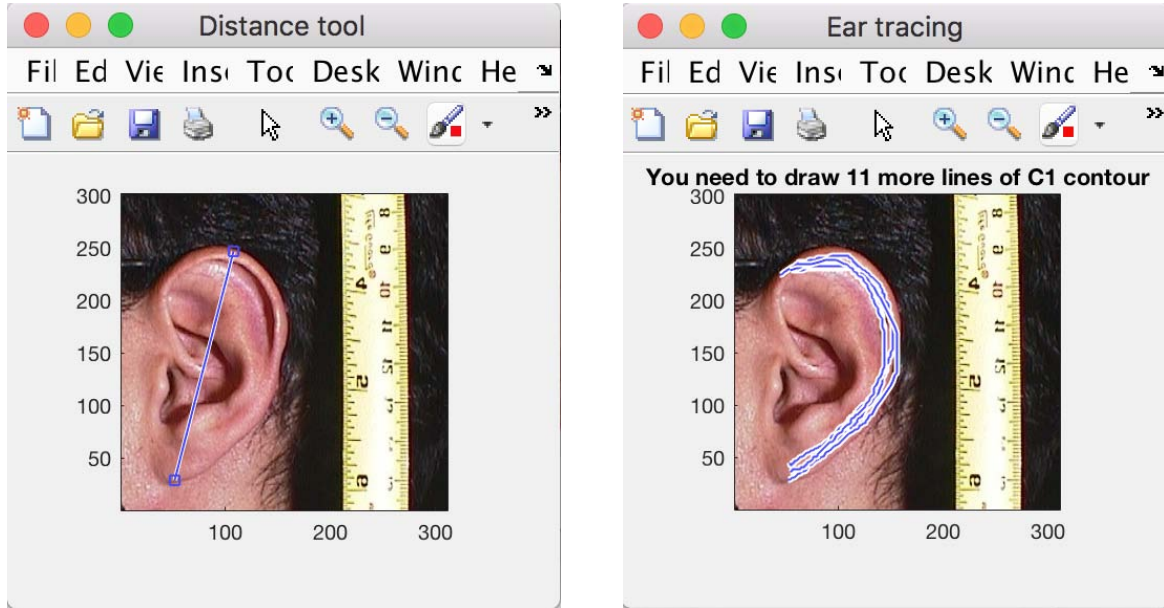
Premendo il pulsante *Trace Ear Contours* si aprirà una finestra dove si potranno tracciare  $N$  contorni ( $C_1, C_2, C_3$ ) e  $K$  punti del canale uditivo selezionati nelle impostazioni.

La motivazione alla base della scelta di un tracciamento multiplo è che, mediando i vari tentativi, si possono minimizzare gli errori dell'utente, dovuti alle tracce a mano libera. Inoltre molto spesso da un'immagine in 2D non si riescono ad identificare chiaramente i bordi dei vari contorni, pertanto si può coprire un'area con più linee piuttosto che cercare di indovinare dove sia il profilo esatto. Anche per il canale uditivo vale lo stesso principio, non si sa esattamente da una foto dove effettivamente sia il suo centro esatto, punto focale della riflessione.

Al termine del tracciamento vengono calcolati i notch in frequenza  $f_0(\phi)$  con  $\phi \in [-45^\circ, 45^\circ]$  per 17 elevazioni differenti equispaziate di  $5,625^\circ$  per ogni coppia  $(k, n)$  canale/contorno con coefficiente di riflessione positivo e negativo. Questi dati verranno poi confrontati con i notch delle HRTF presenti nel database per ogni soggetto CIPIC.

## 2.2 Metriche utilizzate

Una volta tracciate  $N$  linee per ciascun contorno ( $C_1, C_2, C_3$ ) e  $K$  punti dei canali uditivi sull'immagine 2D si può procedere al calcolo vero e proprio delle differenze fra i notch di frequenza delle HRTF. Come visto in precedenza (vedi eq. 1.8) il notch dipende sia dalla distanza fra contorno e punto focale del canale uditivo, sia dal coefficiente di riflessione (positivo o negativo) pertanto per ogni coppia  $(k, n)$  si ricavano sei differenti notch in frequenza (contorno - riflessione).



**Figura 2.4:** (a) Finestra per il tracciamento delle misure antropometriche: tracciamento dell'altezza dell'orecchio. (b) Finestra per il tracciamento dei contorni dell'orecchio: Tracciamento del contorno  $C_1$

### 2.2.1 Mismatch in frequenza

La formula per ricavare il mismatch fra le frequenze, simile a quella vista in precedenza (eq. 1.11) diventa la seguente:

$$m_{(k,n)} = \sum_{i=1}^3 \frac{w_i}{N_\phi} \sum_{\phi} \frac{|f_0^{i(k,n)}(\phi) - F_0^i(\phi)|}{F_0^i(\phi)} \quad (2.1)$$

dove  $(k, n)$  con  $(0 \leq k < K)$  e  $(0 \leq n < N)$  si riferisce ad una particolare coppia di punto focale e canale uditivo;  $i = 1, 2, 3$  sono i contorni di riferimento;  $w_i$  è una combinazione convessa dei pesi di ogni  $C_i$  tale che  $\sum_i w_i = 1$ ;  $f_0^{i(k,n)}(\phi)$  sono i notch estratti dai tracciamenti mentre  $F_0^i(\phi)$  sono i notch presenti nella HRTF nel database;  $\phi \in [-45^\circ, 45^\circ]$  è l'angolo di elevazione;  $w_i$  è una combinazione convessa dei pesi di ogni  $C_i$  e coefficiente di riflessione;  $N_\phi$  è il numero dei campioni considerati per ogni differente elevazione per normalizzare la sommatoria, quando  $f_0^i(\phi) \neq 0 \wedge F_0^i(\phi) \neq 0$

In questo studio sono state considerati dieci metriche con valori differenti di mismatch, ciascuno corrispondente ai seguenti contorni, pesi e coefficienti di riflessione:

1. Contorno  $C_1$  ( $w_1 = 1, w_2 = w_3 = 0$ ) con coefficiente di riflessione **negativo**.

2. Contorno  $C_1$  ( $w_1 = 1, w_2 = w_3 = 0$ ) con coefficiente di riflessione **positivo**.
3. Contorno  $C_2$  ( $w_2 = 1, w_1 = w_3 = 0$ ) con coefficiente di riflessione **negativo**.
4. Contorno  $C_2$  ( $w_2 = 1, w_1 = w_3 = 0$ ) con coefficiente di riflessione **positivo**.
5. Contorno  $C_3$  ( $w_3 = 1, w_1 = w_2 = 0$ ) con coefficiente di riflessione **negativo**.
6. Contorno  $C_3$  ( $w_3 = 1, w_1 = w_2 = 0$ ) con coefficiente di riflessione **positivo**.
7. Contorno  $C_1, C_2, C_3$  **equipesati** ( $w_1 = w_2 = w_3 = 1/3$ ) con coefficiente di riflessione **negativo**.
8. Contorno  $C_1, C_2, C_3$  **equipesati** ( $w_1 = w_2 = w_3 = 1/3$ ) con coefficiente di riflessione **positivo**.
9. Contorno  $C_1, C_2, C_3$  con **pesatura ottima**<sup>1</sup> ( $w_1 = 0.66, w_2 = 0.24, w_3 = 0.1$ ) con coefficiente di riflessione **negativo**.
10. Contorno  $C_1, C_2, C_3$  con **pesatura ottima** ( $w_1 = 0.66, w_2 = 0.24, w_3 = 0.1$ ) con coefficiente di riflessione **positivo**.

I valori risultanti da queste metriche verranno considerati distintamente per creare classifiche separate che riepilogheranno quali sono le HRTF che corrispondono maggiormente al soggetto preso in esame.

### 2.2.2 Mismatch antropometrico

Una volta tracciate le due misure antropometriche descritte in precedenza si procede al calcolo del mismatch fra il soggetto considerato ed i soggetti del database. In particolare il confronto per l'altezza dell'esterno del padiglione verrà effettuato confrontando il valore  $d_5$  presente nel database con la misura tracciata nell'immagine. Il valore di mismatch è calcolato con la seguente formula:

$$m_{d_1} = |d_{ext} - d_5| \quad (2.2)$$

dove  $d_{ext}$  è la misura tracciata.

Il confronto per l'altezza dell'interno del padiglione verrà effettuato confrontando la somma dei valori  $d_1, d_2, d_4$  con la misura tracciata nell'immagine. Il valore di mismatch è calcolato con

<sup>1</sup>La pesatura ottima è stata definita seguendo un lavoro in corso del gruppo di ricerca CSC-SMC.



la seguente formula:

$$m_{d_2} = |d_{int} - (d_1 + d_2 + d_4)| \quad (2.3)$$

dove  $d_{int}$  è la misura tracciata.

## 2.3 Classifiche

Calcolati tutti i valori dei mismatch in frequenza estratti dai tracciamenti nell'immagine e confrontati con i soggetti nel database, si possono ora ricavare delle metriche di distanza per classificare le HRTF dei CIPIC ID in ordine di somiglianza. In particolare ne definiamo tre:

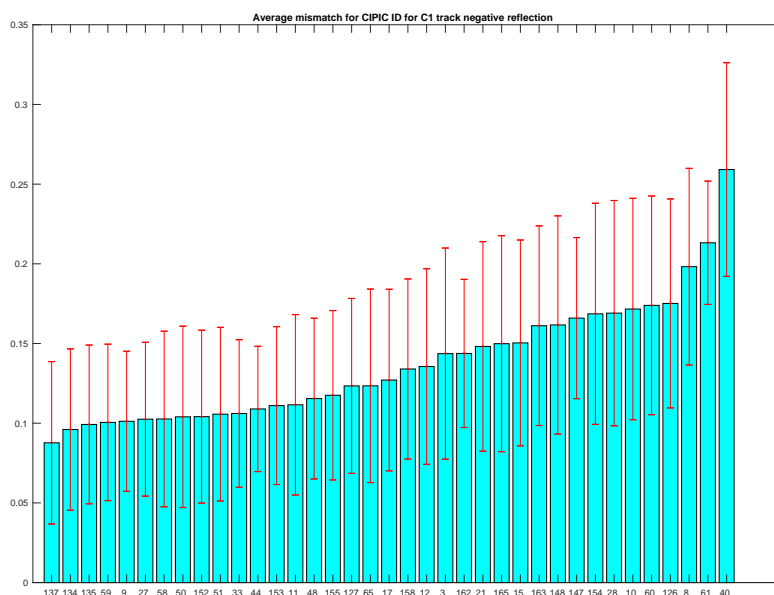
- **Average mismatch:** calcolati con l'eq. 2.1 sono i valori medi di mismatch, ordinati in ordine crescente, delle coppie  $(k, n)$  per ciascuna HRTF. Assume valori fra  $[0, 1]$ .
- **Average rank:** calcolati i mismatch per la coppia  $(k, n)$  ed ordinati in ordine crescente, si ricava un valore intero che corrisponde alla posizione in classifica della HRTF. Facendo la media di questo valore per tutte le coppie  $(k, n)$  si ha la misura di *average rank* per ogni HRTF. Assume valori da 1 al numero di HRTF presenti nel database.
- **Top-M appearance:** calcolati i mismatch per la coppia  $(k, n)$  ed ordinati in ordine crescente è il numero di volte che la HRTF appare nelle prime  $M$  posizioni in tutte coppie  $(k, n)$ . Assume valori interi fra  $[0, N \times K]$ .

Considerando dieci differenti metriche e tre modi univoci per confrontare le HRTF vengono generate un totale di 30 classifiche per ogni soggetto. Ovviamente se non sono stati tracciati uno o più contorni ( $C_1, C_2, C_3$ ) le classifiche relative alle condizioni di equipesatura (7,8) e pesatura ottima (9,10) non verranno calcolate.

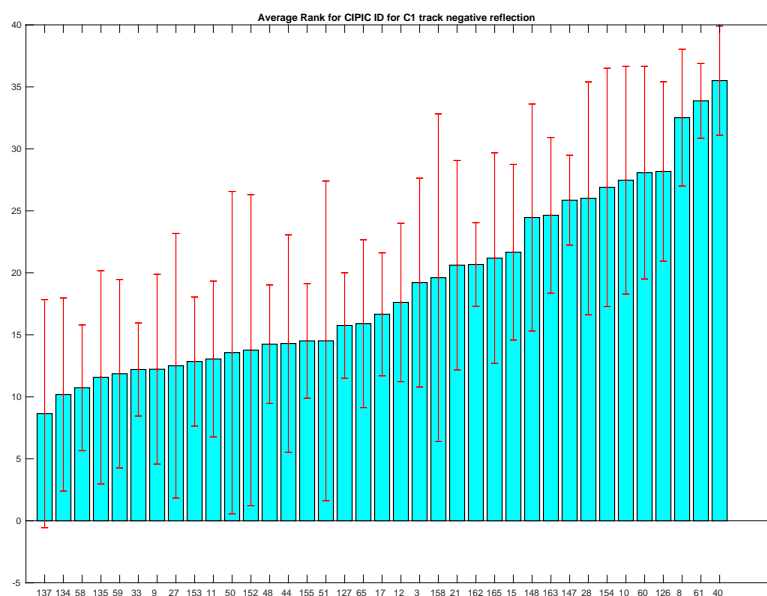
Per quanto riguarda i dati antropometrici sono stilate due classifiche:

- **Mismatch dell'altezza esterna:** raggruppa i valori calcolati con l'eq. 2.2 in ordine crescente per ogni CIPIC ID.
- **Mismatch dell'altezza interna:** raggruppa i valori calcolati con l'eq. 2.3 in ordine crescente per ogni CIPIC ID.

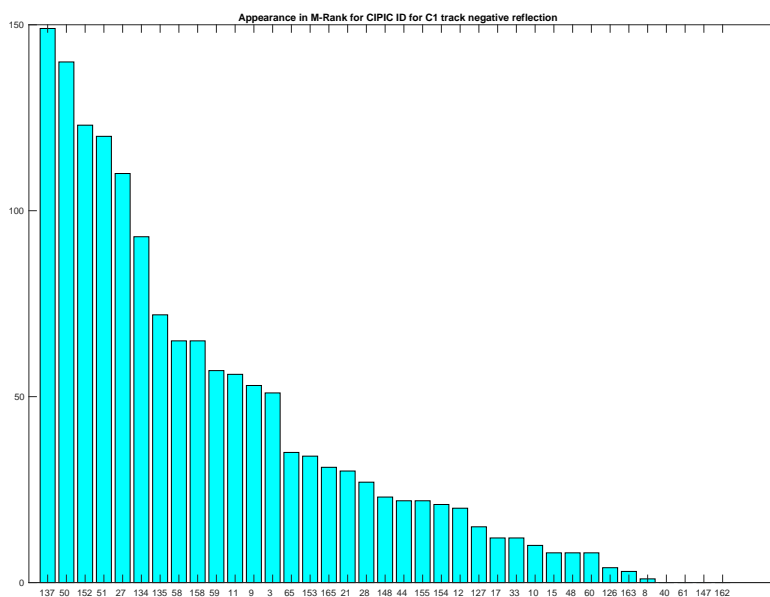
In totale al termine dell'elaborazione, se sono stati tracciati e selezionati tutti i campi possibili presenti nelle configurazioni, verranno create 32 classifiche. L'elaborazione richiede un tempo  $O(N \cdot M)$ . Nell'esperimento trattato nei successivi capitoli, con  $N = 15$  e  $K = 20$  le



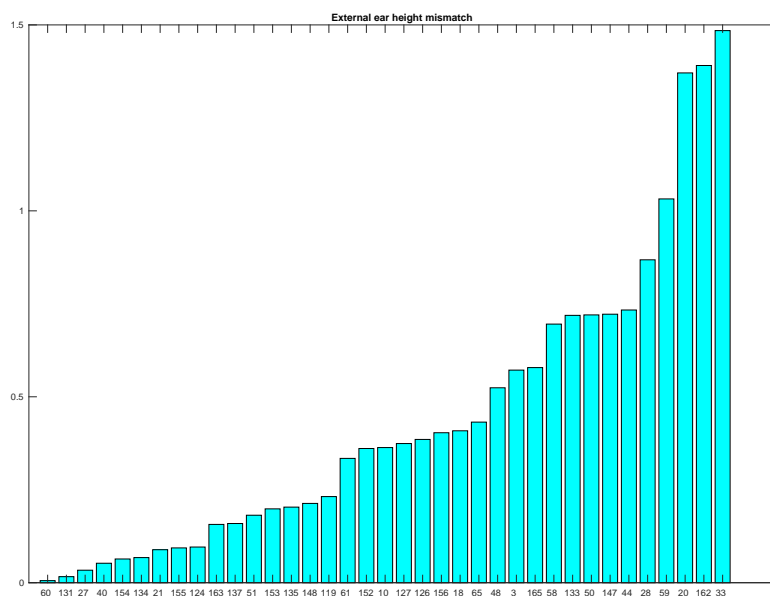
**Figura 2.5:** Classifica di average mismatch, contorno  $C_1$ , riflessione negativa per il CIPIC ID 27. Asse x: CIPIC ID; asse y: mismatch.



**Figura 2.6:** Classifica di average rank, contorno  $C_1$ , riflessione negativa per il CIPIC ID 27. Asse x: CIPIC ID; asse y: rank medio.



**Figura 2.7:** Classifica di top- $M$  appearance, contorno  $C_1$ , riflessione negativa per il CIPIC ID 27. Asse  $x$ : CIPIC ID; asse  $y$ : numero di occorrenze.



**Figura 2.8:** Classifica di mismatch dell'altezza esterna dell'orecchio per il CIPIC ID 27. Asse  $x$ : CIPIC ID; asse  $y$ : mismatch(cm).

classifiche vengono generate in poco più di 1 minuto. Per una maggiore fruibilità dei contenuti i file generati sono grafici in formato *.pdf* raccolti in una cartella per ogni soggetto. I grafici sono degli istogrammi a barre raffiguranti i valori medi con deviazione standard per le classifiche di *Average mismatch* ed *Average rank* e numero di occorrenze per le classifiche di *Top-M appearance*.



# Capitolo 3

## Analisi delle metriche

In questo capitolo valuteremo tutte le metriche discusse precedentemente per capire quali siano le più significative per la selezione della miglior HRTF. L'obiettivo è di riuscire a selezionare un set ridotto di HRTFs che più rispecchiano le caratteristiche del soggetto esterno, in modo che si possano testare più approfonditamente in un secondo momento, ad esempio con un test in realtà virtuale, per scegliere la più adatta alle caratteristiche psicoacustiche. Per valutare le performance della percezione dei suoni in elevazione con la HRTF selezionata è stata effettuata una simulazione con un modello uditivo <sup>1</sup>.

La validazione del metodo utilizzato con la metrica per il contorno  $C_1$  con coefficiente di riflessione negativa è stata argomento di un lavoro discusso in un paper sottoposto per una pubblicazione [17].

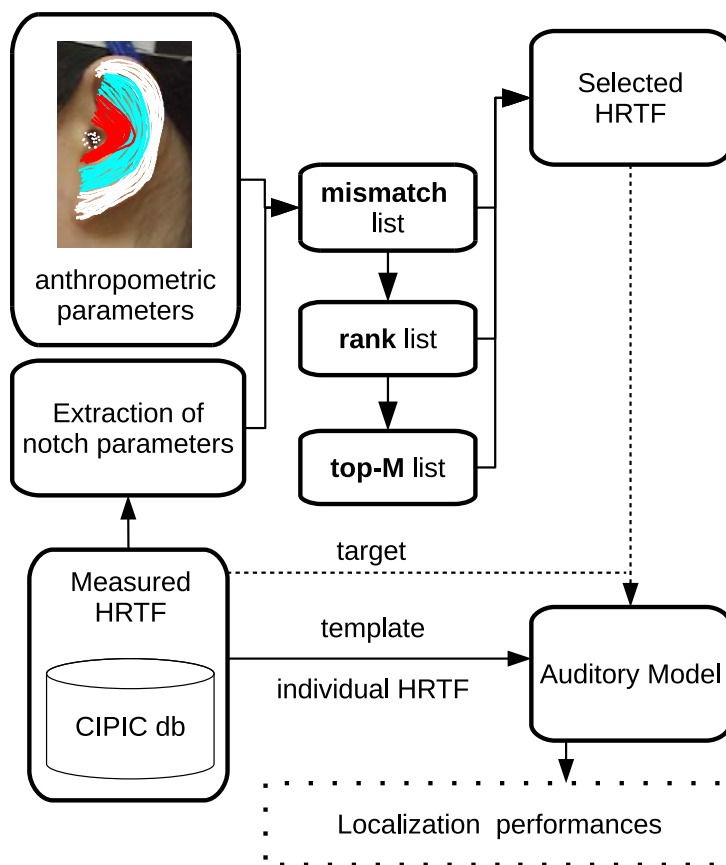
### 3.1 Acquisizione dei dati

#### 3.1.1 Soggetti dell'esperimento

I soggetti per l'esperimento sono stati presi dal database CIPIC per avere un confronto diretto fra HRTF estratta dall'applicazione e HRTF effettivamente misurata in laboratorio. Fra i 45 soggetti presenti nel database ne sono stati selezionati 18 per i quali era presente la foto laterale dell'orecchio, necessaria per i tracciamenti, le misure antropometriche, necessarie per l'impostazione del *pixel to meter factor* e tutte le tracce dei contorni  $C_1, C_2, C_3$  che formano i notch in frequenza delle HRTFs necessarie per il calcolo dei mismatch.

---

<sup>1</sup>vedi sez. 3.2.2

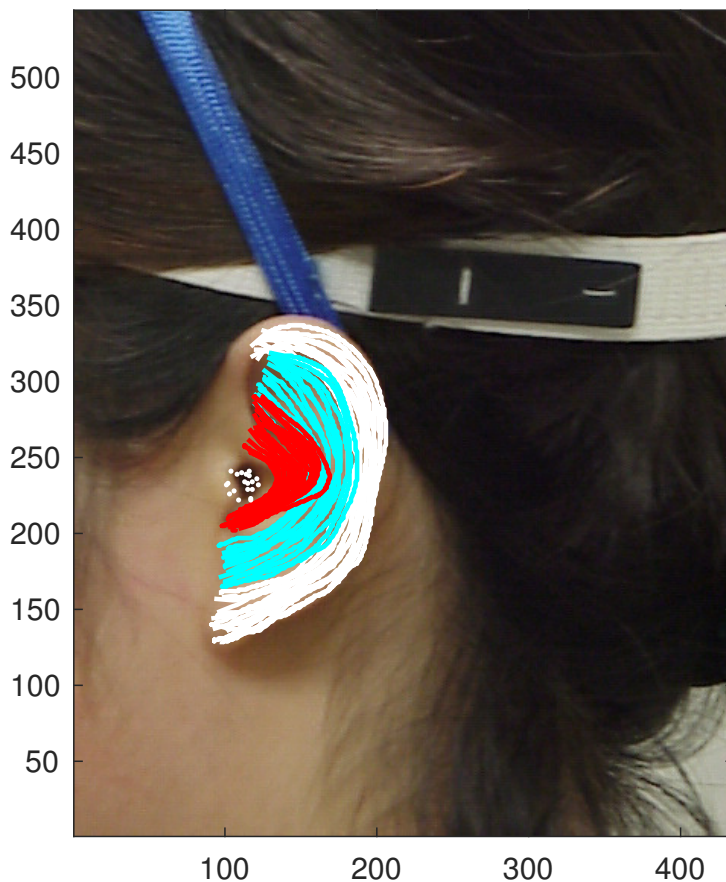


**Figura 3.1:** Schema del procedimento di validazione con le predizioni del modello uditivo.

### 3.1.2 Tracciamento dei contorni

Abbiamo scelto di tracciare  $N = 15$  linee per ogni contorno ( $C_1, C_2, C_3$ ) e  $K = 20$  punti per il canale uditivo. In questo modo si possono coprire le aree dei contorni, invece del bordo netto, in quanto dalla foto non si può capire con precisione dove sia effettivamente l'avvallamento. In aggiunta non potendo prevedere se le riflessioni del contorno siano positive o negative, ricoprire un'area più ampia consente di tenere in considerazione entrambe le possibilità [28]. I punti e le linee sono tracciati con le seguenti linee guida (vedi fig. 3.2):

- **Contorno  $C_1$ :** area dal bordo esterno dell'elice al bordo interno dell'elice, sconfinando leggermente anche all'interno della fossa.
- **Contorno  $C_2$ :** area dal bordo interno dell'elice al bordo esterno della conca.
- **Contorno  $C_3$ :** area dal bordo esterno della conca fino all'ingresso del canale uditivo.
- **Canale uditivo:** nuvola di punti che ricopre l'area del punto focale del canale uditivo.



**Figura 3.2:** Tracciamenti di contorni e canali uditivi per il soggetto CIPIC 60. In bianco il contorno  $C_1$ , in celeste  $C_2$  e in rosso  $C_3$ . I punti bianchi sono i punti focali del canale uditivo.

La procedura di tracciamento richiede circa 10 minuti per ogni soggetto analizzato. Risulteranno in totale 300 coppie  $(n, k)$  contorno/canale uditivo. Si possono vedere i grafici delle frequenze risultanti dai tracciamenti  $f_0^{(k,n)}$  nelle figure 3.7 e 3.8

Per l'impostazione del *pixel to meter factor* è stato preso come riferimento la misura antropometrica  $d_5$  presente nel database CIPIC.

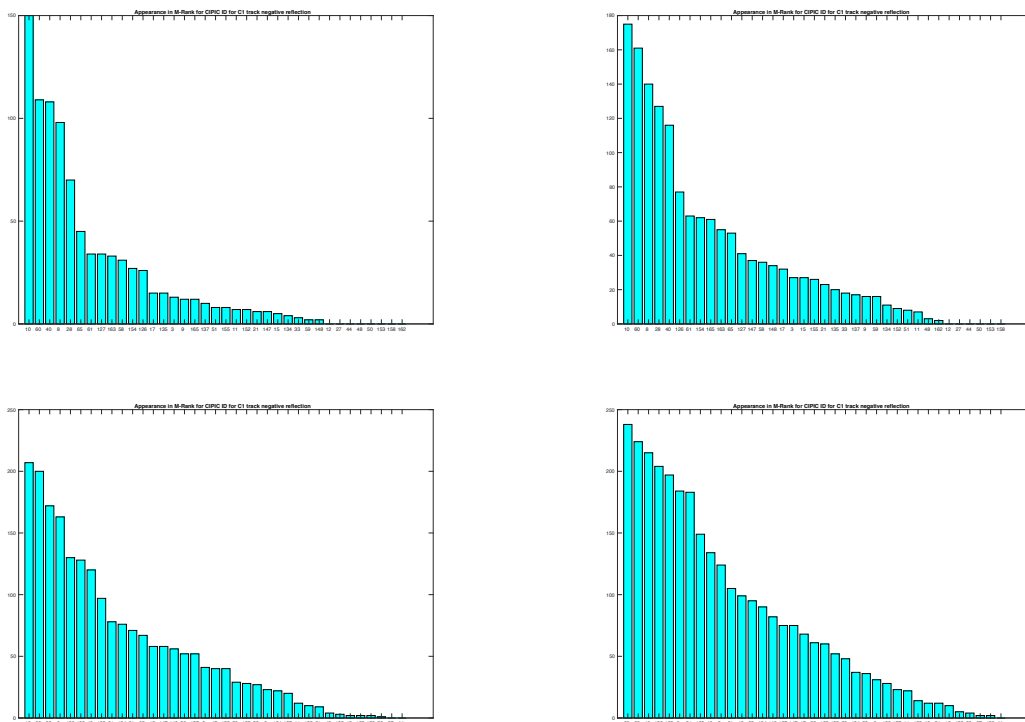
## 3.2 Analisi delle prestazioni

### 3.2.1 Classifiche

Sono state considerate tutte le 30 classifiche relative alle riflessioni dei contorni elaborate con l'applicazione tralasciando le ultime due relative alle misure antropometriche.



Il valore di  $M$  è stato impostato empiricamente a 5. Un valore più alto avrebbe creato un numero considerevole di HRTF in alcune classifiche che si piazzavano al primo posto essendo presenti in ogni coppia, mentre un valore più basso avrebbe ridotto il numero di HRTF che comparivano nella classifica, in quanto avrebbero avuto un numero di occorrenze pari a zero. Inoltre per valori elevati ( $M > 10$ ) la classifica di *Top-M appearance* degenera e diventa molto simile a quella di *Average rank*. Per ogni soggetto avremo quindi il numero di occorrenze delle HRTF che si classificano nei primi 5 posti considerando tutte le 300 coppie  $(n, k)$ .



**Figura 3.3:** Classifiche di *Top-M Appearance* al variare di  $M$ . (a,b,c,d)  $M = 3, 5, 7, 10$  contorno  $C_1$  riflessione negativa.

Per ciascuna classifica abbiamo selezionato la HRTF che si classificava al primo posto. Avremo quindi per ogni metrica<sup>2</sup> l'ID CIPIC della HRTF che meglio si è classificata per *Average mismatch*, *Average rank* e *Top-5 appearance*. Le candidate vincitrici passeranno allo step successivo di simulazione con il modello uditivo.

<sup>2</sup>Vedi sez. 2.2.1

### 3.2.2 Simulazione con modello uditivo

Utilizzando le previsioni di un modello uditivo, abbiamo simulato un esperimento virtuale dove ogni ascoltatore CIPIC avrebbe dovuto dare una valutazione sulla localizzazione dello stimolo uditivo percepito. Abbiamo utilizzato un modello recente [3], che segue un paradigma basato sul confronto tra la rappresentazione interna di un suono in arrivo sul timpano e un modello di riferimento. Le caratteristiche spettrali delle differenti HRTF sono correlate alla direzione di arrivo, creano una mappatura spettro-spaziale e una metrica percettiva per le prestazioni di elevazione.

Il modello si basa su un'elaborazione a due fasi [8]:

1. Viene creata una rappresentazione interna del suono in entrata ed il suono in uscita, ad esempio una HRTF, viene convertito in una funzione di trasferimento direzionale, *Directional Transfer Function* (DTF).
2. La nuova rappresentazione viene confrontata con il modello, ovvero le DTFs individuali calcolate dalle HRTFs, simulando così il processo di localizzazione del nostro cervello.

Per un angolo predefinito, la probabilità che l'ascoltatore virtuale lo punti esattamente definisce l'indice di similarità *Similarity Index* (SI). Il valore dell'indice deriva dalla distanza in gradi tra l'angolo reale e l'angolo puntato. E' definito da una distribuzione gaussiana di media zero e deviazione standard chiamata incertezza *Uncertainty* (U). Più basso è il valore di U, più alta è la sensibilità dell'ascoltatore virtuale nel distinguere un profilo di ascolto (HRTF) determinando una misura di probabilità e non un valore deterministico.

La simulazione è stata effettuata incrociando i soggetti CIPIC con le HRTF del database, con un valore di incertezza tarato sulla sensibilità di un umano ( $U = 1.8$ ). Abbiamo predetto le prestazioni in elevazione per ogni soggetto quando era in ascolto con: HRTF individuale, KEMAR (CIPIC ID 165) e la migliore per classifica di *Average mismatch*, *Average rank* e *Top-5 appearance* per ciascuna delle 10 metriche proposte.

La precisione della j-esima elevazione rispetto alla posizione assoluta della sorgente sonora è definita dal *Local Polar RMS Error* (PE):

$$PE_j = \sqrt{\frac{\sum_{i \in L} (\phi_i - \varphi_j)^2 p_j[\phi_i]}{\sum_{i \in L} p_j[\phi_i]}} \quad (3.1)$$

dove  $L = \{i \in N : 1 \leq i \leq N_\phi, |\phi_i - \varphi_j| \bmod 180^\circ < 90^\circ\}$  definisce le risposte locali in elevazione fra  $[-90^\circ, +90^\circ]$ , rispetto alla risposta locale  $\phi_i$  ed alla posizione assoluta  $\varphi_j$ , mentre  $p_j[\phi_i]$  denota la previsione, ovvero la funzione di probabilità.

Il valore PE medio è stato calcolato considerando le risposte in elevazione fra  $[-45^\circ, +45^\circ]$ , dove emergono differenze di percezione spaziale dei suoni da parte di un essere umano [4]. Questo valore fornirà una metrica globale per le performance di localizzazione per i soggetti dell'esperimento.

### 3.3 Risultati

Ci siamo proposti di valutare i risultati ottenuti distinguendo le tre classifiche con cui le HRTF sono state selezionate. Per ogni soggetto avremo 10 metriche, di cui confronteremo le prestazioni ottenute con la simulazione precedente, per evidenziare quale sia la più adatta per la selezione. Solo in rari casi l'applicazione riuscirà a selezionare la HRTF individuale. Lo scopo dell'analisi è capire se le performance della scelta migliore possano essere paragonabili alla selezione individuale. Verrà quindi effettuato un paired t-test per valutare se le metriche riescano a selezionare una HRTF che non ha performance con differenze statisticamente significative rispetto alla scelta della HRTF individuale.

#### 3.3.1 Selezione con average mismatch

Prendiamo ora in considerazione per ogni soggetto le migliori HRTF selezionate secondo la classifica di *Average mismatch*. Per tutti i 18 soggetti sarà considerata la metrica che ha selezionato la HRTF con migliori prestazioni nel modello uditivo. Nella tabella 3.1 si possono vedere per ogni soggetto, la metrica migliore di selezione, l'ID CIPIC della HRTF selezionata e le performance di ascolto con HRTF scelta dall'applicazione, HRTF individuale e HRTF generica (KEMAR).

La metrica 1 (Contorno  $C_1$  con coefficiente di riflessione **negativo**) con questa selezione risulta la migliore con 7 occorrenze, seguono la metrica 3 con 4 occorrenze e la 6 con 3 occorrenze. Nella tabella 3.2 sono presenti i valori delle altre metriche. In fig 3.4 si possono vedere le prestazioni medie con la miglior HRTF selezionata, individuale e KEMAR.

Per valutare se questo metodo di selezione possa essere paragonato alla scelta dell'individuale sono stati effettuati due paired t-test sul campione di 18 soggetti con i seguenti risultati:

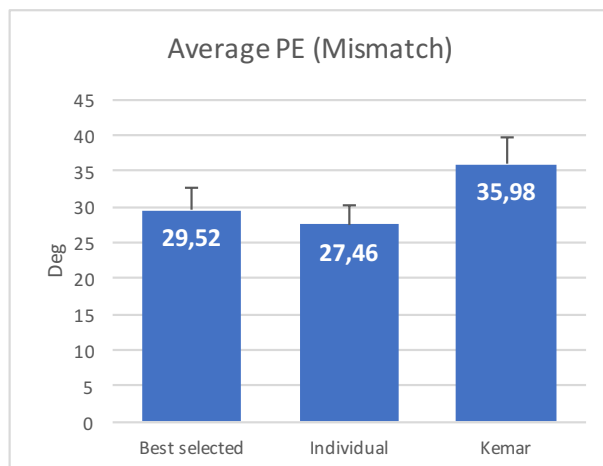
- Performance HRTF **best selected** vs performance HRTF **individuale**: c'è una differenza significativa ( $p = 0.004$ ).

Average Mismatch Performance					
Soggetto	Metrica	Best HRTF	PE Best	PE Ind	PE Kamar
3	6	163	31,83	27,77	32,62
10	3	17	28,56	23,54	32,11
27	1	137	32,58	29,94	35,64
40	1	60	30,21	26,30	35,54
44	7	51	27,63	25,27	36,01
48	5	65	25,96	26,00	31,14
50	7	51	27,39	25,43	40,50
51	1	51	29,43	29,43	46,21
59	3	135	32,57	33,51	36,57
60	1	60	26,74	26,74	38,43
134	3	17	30,54	30,11	35,76
135	1	28	32,16	27,56	35,54
137	1	163	38,64	29,90	36,32
147	6	137	26,83	22,43	31,17
148	5	40	27,52	27,44	34,83
152	6	152	24,85	24,85	38,97
154	1	40	28,54	30,10	35,26
162	3	65	29,40	27,91	35,08

**Tabella 3.1:** Performance della miglior metrica per ogni soggetto con selezione secondo average mismatch.

Metrica		Occorrenze
1	C1 neg	7
2	C1 pos	0
3	C2 neg	4
4	C2 pos	0
5	C3 neg	2
6	C3 pos	3
7	Eq W neg	2
8	Eq W pos	0
9	Opt W neg	0
10	Opt W pos	0

**Tabella 3.2:** Numero di occorrenze per metrica con selezione secondo average mismatch.



**Figura 3.4:** Prestazioni medie con HRTF selezionata con mismatch minore, HRTF individuale e HRTF del KEMAR.

- Performance HRTF **best selected** vs performance HRTF **KEMAR**: c'è una differenza significativa ( $p < 0.001$ ).

Con questi risultati possiamo affermare che questo metodo di selezione non ha dato dei risultati paragonabili alla scelta della HRTF individuale se si confrontano le performance in elevazione con il modello uditivo. Nonostante ciò l'HRTF scelta con l'applicazione garantisce in media dei risultati migliori rispetto alla scelta di un HRTF generica.

### 3.3.2 Selezione con average rank

Passiamo ora a considerare per ogni soggetto le migliori HRTF selezionate secondo la classifica di *Average rank*. Anche in questo caso verrà scelta per ogni soggetto la metrica con migliori prestazioni nel modello uditivo. Nella tabella 3.3 si possono vedere i valori medi delle performance.

In questo caso la miglior metrica è la 5 (Contorno  $C_3$  con coefficiente di riflessione **negativo**) presente un totale di 7 volte, segue la 1 con 5 presenze e la 3 con 3. Nella tabella 3.4 si possono vedere il numero di occorrenze delle altre metriche. In fig. 3.5 sono presenti i valori medi delle performance di localizzazione per HRTF selezionata, individuale e KEMAR. Come

nel caso precedente abbiamo effettuato i paired t-test con i seguenti risultati:

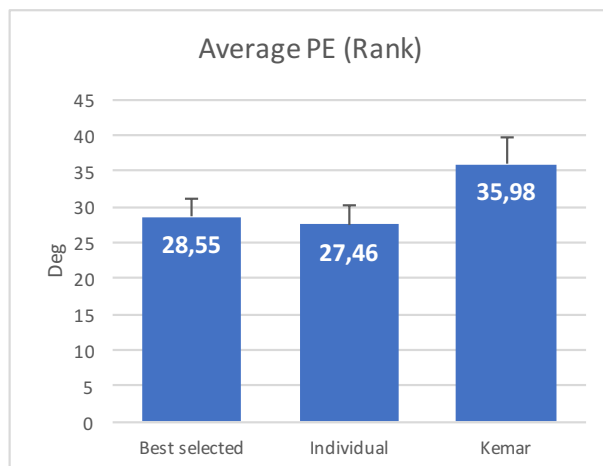
- Performance HRTF **best selected** vs performance HRTF **individuale**: non c'è differenza significativa ( $p = 0.028$ ).

Average Rank Performance					
Soggetto	Metrica	Best HRTF	PE Best	PE Ind	PE Kemar
3	5	65	25,85	27,77	32,62
10	5	65	25,71	23,54	32,11
27	1	137	32,58	29,94	35,64
40	1	60	30,21	26,30	35,54
44	1	51	27,63	25,27	36,01
48	3	65	25,96	26,00	31,14
50	7	51	27,39	25,43	40,50
51	1	51	29,43	29,43	46,21
59	3	135	32,57	33,51	36,57
60	1	60	26,74	26,74	38,43
134	3	17	30,54	30,11	35,76
135	5	40	27,87	27,56	35,54
137	5	11	34,35	29,90	36,32
147	8	137	26,83	22,43	31,17
148	5	148	27,44	27,44	34,83
152	4	152	24,85	24,85	38,97
154	5	40	28,54	30,10	35,26
162	5	65	29,40	27,91	35,08

**Tabella 3.3:** Prestazioni medie della miglior metrica per ogni soggetto con selezione secondo average rank.

	Metrica	Occorrenze
1	C1 neg	5
2	C1 pos	0
3	C2 neg	3
4	C2 pos	1
5	C3 neg	7
6	C3 pos	0
7	Eq W neg	1
8	Eq W pos	1
9	Opt W neg	0
10	Opt W pos	0

**Tabella 3.4:** Numero di occorrenze per metrica con selezione secondo average rank.



**Figura 3.5:** Prestazioni medie con HRTF selezionata con rank migliore, HRTF individuale e HRTF del KEMAR.

- Performance HRTF **best selected** vs performance HRTF **KEMAR**: c'è una differenza significativa ( $p < 0.001$ ).

Sorprendentemente questa metrica riesce a selezionare una HRTF con performance paragonabili a quella individuale in quanto non ci sono differenze statisticamente significative ( $p = 0.028$ ). Anche se in media ha prestazioni inferiori all'HRTF individuale questo metodo di selezione identifica una HRTF in grado di garantire performance decisamente migliori di una generica (KEMAR) che si avvicinano molto all'individuale.

### 3.3.3 Selezione con Top-5 appearance

Consideriamo l'ultimo set di classifiche. In questo caso per ogni coppia soggetto/metrica prenderemo le HRTF che sono apparse più frequentemente fra i primi 5 posti. Successivamente per tutti i 18 soggetti sceglieremo la metrica con prestazioni migliori nel modello uditivo. Nella tabella 3.5 sono presenti le metriche selezionate e i valori medi di PE.

Con questo criterio di selezione la miglior metrica risulta essere la 3 (Contorno  $C_2$  con coefficiente di riflessione **negativo**) presente 7 volte, seguono a parimerito le metriche 1 e 4 con 4 occorrenze. Nella tabella 3.6 si possono vedere il numero di occorrenze delle altre metriche.

Anche in questo caso abbiamo effettuato i paired t-test con i seguenti risultati:

- Performance HRTF **best selected** vs performance HRTF **individuale**: non c'è differenza significativa ( $p = 0.32$ ).

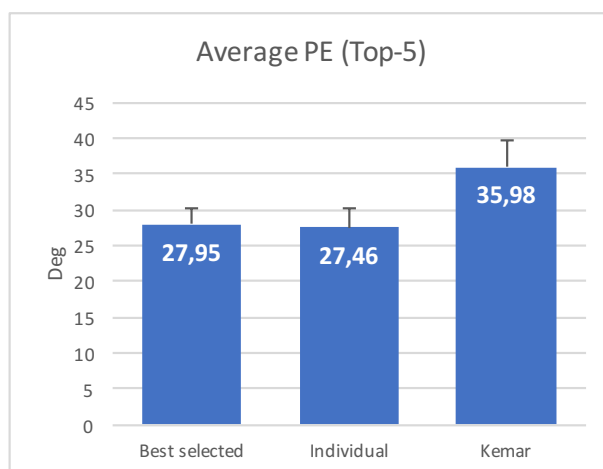
Average Top-5 Performance					
Soggetto	Metrica	Best HRTF	PE Best	PE Ind	PE Kamar
3	3	65	25,85	27,77	32,62
10	3	65	25,71	23,54	32,11
27	1	137	32,58	29,94	35,64
40	1	60	30,21	26,30	35,54
44	1	51	27,63	25,27	36,01
48	3	65	25,96	26,00	31,14
50	2	51	27,39	25,43	40,50
51	2	51	29,43	29,43	46,21
59	3	135	32,57	33,51	36,57
60	4	65	27,02	26,74	38,43
134	3	65	28,83	30,11	35,76
135	4	65	26,38	27,56	35,54
137	4	65	29,47	29,90	36,32
147	1	137	26,83	22,43	31,17
148	3	65	25,61	27,44	34,83
152	10	152	24,85	24,85	38,97
154	4	65	27,45	30,10	35,26
162	3	65	29,40	27,91	35,08

**Tabella 3.5:** Prestazioni medie della miglior metrica per ogni soggetto con selezione secondo top 5 appearance.

Metrica		Occorrenze
1	C1 neg	4
2	C1 pos	2
3	C2 neg	7
4	C2 pos	4
5	C3 neg	0
6	C3 pos	0
7	Eq W neg	0
8	Eq W pos	0
9	Opt W neg	0
10	Opt W pos	1

**Tabella 3.6:** Numero di occorrenze per metrica con selezione secondo top 5 appearance.



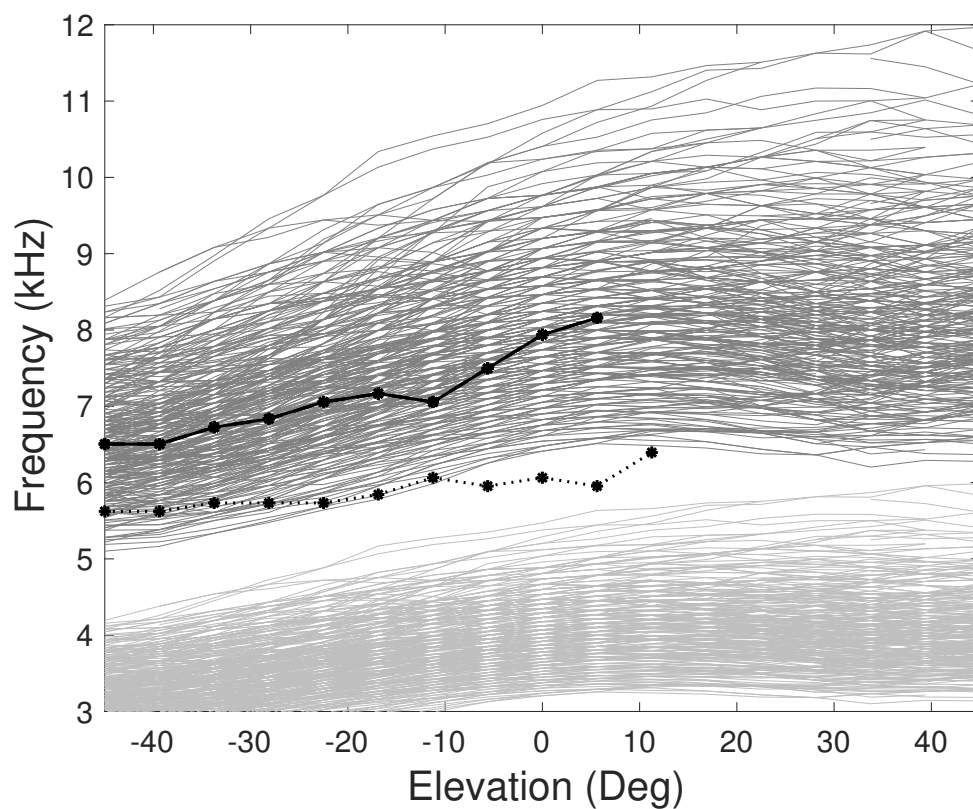


**Figura 3.6:** Prestazioni medie con HRTF selezionata con più occorrenze nella classifica top 5 appearance, HRTF individuale e HRTF del KEMAR.

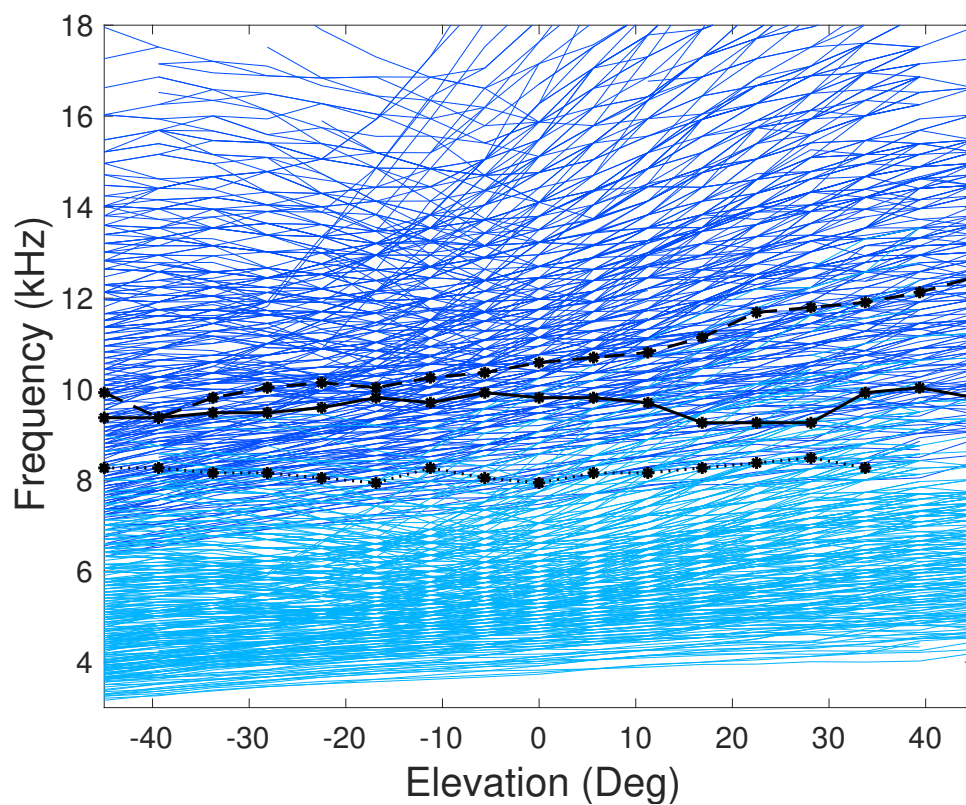
- Performance HRTF **best selected** vs performance HRTF **KEMAR**: c'è una differenza significativa ( $p < 0.001$ ).

Con questa analisi la metrica risulta davvero ottima e decisamente più precisa della precedente raggiungendo valori molto vicini alle performance dell'individuale. Non sono presenti differenze significative ( $p = 0.32$ ) ed in media le prestazioni sono paragonabili all'individuale (vedi fig. 3.6). Le performance sono decisamente migliori della HRTF generica.

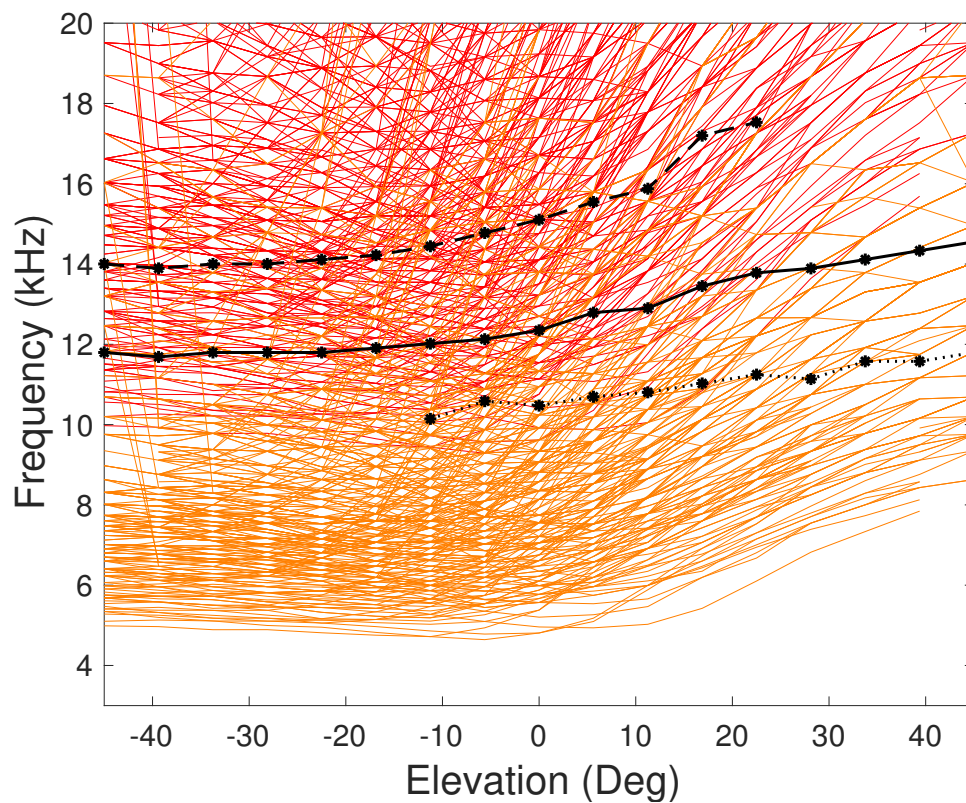
Bisogna osservare la seguente presenza della HRTF relativa al CIPIC ID 65 in 10 casi su 18 con questo metodo di selezione. Il motivo potrebbe essere che i buchi in frequenza della HRTF 65 nel contorno  $C_2$ , avendo coperto un'ampia zona del padiglione nei tracciamenti, risultano al centro di queste aree in molti soggetti, classificandosi frequentemente nelle prime 5 posizioni delle coppie  $(n, k)$ . Inoltre se la HRTF 65 ha prestazioni assolute medie maggiori rispetto alle altre, la metrica 3 o 4 verrà scelta come metrica migliore fra le 10 considerate.



**Figura 3.7:** Tracce spettrali di  $f_0^{(k,n)}$  per i tracciamenti multipli di  $C_1$ /punto focale per il soggetto CIPIC 60. In grigio scuro le riflessioni negative, in grigio chiaro le positive.  $F_0$  della HRTF individuale e best Top-5 per la riflessione negativa (linea nera continua) e HRTF best Top-5 per la riflessione positiva (linea nera puntinata).



**Figura 3.8:** Tracce spettrali di  $f_0^{(k,n)}$  per i tracciamenti multipli di  $C_2$ /punto focale per il soggetto CIPIC 60. In blu le riflessioni negative, in azzurro le positive.  $F_0$  della HRTF individuale (linea nera continua), HRTF best Top-5 (linea nera tratteggiata) per la riflessione negativa e HRTF best Top-5 per la riflessione positiva (linea nera puntinata).



**Figura 3.9:** Tracce spettrali di  $f_0^{(k,n)}$  per i tracciamenti multipli di  $C_3$ /punto focale per il soggetto CIPIC 60. In rosso le riflessioni negative, in arancione le positive.  $F_0$  della HRTF individuale (linea nera continua), HRTF best Top-5 (linea nera tratteggiata) per la riflessione negativa e HRTF best Top-5 per la riflessione positiva (linea nera puntinata).



# Capitolo 4

## Esperimento in VR

In collaborazione con un altro tesista, Fabio Prandoni, con cui ho condiviso lo sviluppo dell'applicazione per la selezione di HRTF, abbiamo effettuato dei test di localizzazione in realtà virtuale per confrontare le prestazioni delle HRTF scelte sottoponendole all'ascolto con soggetti reali. Si è cercato di mantenere un design sperimentale snello e di veloce esecuzione per i soggetti in modo da poter essere utilizzato in futuro come screening test prima di qualsiasi esperienza di realtà virtuale; un primo tentativo è già stato effettuato in [11] per ambienti di realtà virtuale non-visiva utilizzando un approccio sperimentale che fa uso di task di raggiungimento della sorgente. In questo lavoro si è cercato invece di prediligere l'aspetto ecologico della localizzazione seguendo un task di ricerca.

### 4.1 Test di localizzazione

In questo esperimento è stato chiesto ai soggetti di localizzare delle sorgente sonore provenienti da diverse direzioni. Abbiamo valutato le prestazioni di localizzazione di 15 persone quando ascoltavano con tre set differenti di HRTF (A,B,C). Durante l'ascolto i soggetti erano in una stanza insonorizzata per evitare qualsiasi interferenza esterna. Ognuno di essi doveva completare tre diverse sessioni di ascolto senza conoscere l'ordine delle HRTF caricate. Per completare l'intero test un soggetto impiegava in media 50 minuti.

La procedura del test di localizzazione è basata sulle linee guida fornite dal paper [20].

#### 4.1.1 Attrezzatura

Come strumento per effettuare il test è stato utilizzato uno smartphone Samsung S7 accoppiato con un visore Samsung Gear VR per il rendering 3D. La sorgente sonora è stata renderizzata da



**Figura 4.1:** Condizioni dell'esperimento: il soggetto si posiziona al centro della camera insonorizzata con la luce spenta per tutta la durata del test.

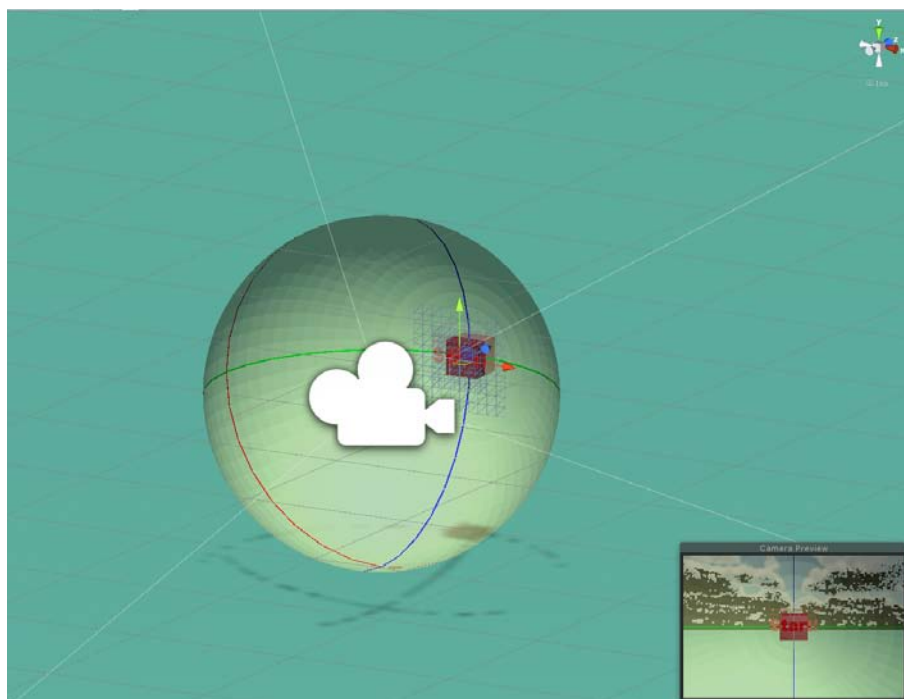
PC con HOBA<sup>1</sup> a cui sono state collegate delle cuffie Sennheiser HD600 opportunamente compensate. In particolare per quest'ultimo aspetto sono state utilizzate le risposte all'impulso della cuffia misurate su più di 100 soggetti dall'Austrian Research Institute<sup>2</sup> e disponibili in formato sofa [6]. Tali misurazione sono state utilizzate per calcolare un filtro di equalizzazione capace di rimuovere il contributo acustico medio delle cuffie e ridurre la colorazione spettrale in fase di playback [5]. La trasmissione delle coordinate spaziali della posizione puntata dal soggetto nella scena virtuale, da smartphone a PC, avveniva tramite una connessione dati wireless.

#### 4.1.2 Scena virtuale in Unity

All'interno di una sfera virtuale di un metro di raggio, il soggetto è libero di guardare in ogni direzione. Con un tocco sul touchpad del Samsung Gear VR è possibile marcare il punto da cui si ritiene provenire il suono udito. Per puntare la direzione al centro del campo visivo è presente un reticolo laser. L'ambiente circostante è composto da un cielo blu e da una distesa di erba verde senza oggetti che possano interferire. Dopo ogni tap, il soggetto deve ritornare nel punto

<sup>1</sup>Acronimo di HRTFs On-demand for Binaural Audio, è un renderer audio 3D per browser sviluppato dall'Università di Padova

<sup>2</sup><http://sofacoustics.org/data/headphones/ari>



**Figura 4.2:** *Panoramica della scena virtuale dell'esperimento in Unity*

di partenza dell'esperimento dove è presente un cubo marrone alla posizione (0, 0), attendere due secondi, seguire nuovamente il suono, identificarlo e ripetere la sequenza.

Ogni soggetto deve completare la serie completa di stimoli sonori prima di procedere al test successivo. Successivamente fare tre minuti di pausa per evitare l'affaticamento e compilare il questionario di fine sessione. Nel frattempo verrà caricata la nuova HRTF.

Dopo la serie di tre test i risultati sono pronti per essere processati. I file sono salvati nello smartphone in un file *.txt* con l'ID del soggetto e la condizione (A,B,C) applicata.

### 4.1.3 Suoni

Ciascun test è composto da una serie di 25 stimoli sonori. Ogni suono si compone di sei ripetizioni di:

- 3 impulsi consecutivi di rumore bianco gaussiano di durata  $40ms$ .
- $30ms$  di silenzio.

È stato provato che questi suoni sono più efficaci rispetto ad un treno di impulsi standard di rumore bianco [22]. Sommando la pausa di 2 secondi prima della partenza del suono, la durata complessiva della prova è di 5 secondi (vedi fig. 4.3).





**Figura 4.3:** Impulso sonoro di una prova. Durata: 5 secondi.

Questi suoni sono generati considerando tutte le possibili combinazioni di 6 valori in azimuth (da  $-180^\circ$  a  $120^\circ$  con passo di  $60^\circ$ ) e 4 valori in elevazione (da  $-28.125^\circ$  a  $56.250^\circ$  con passo di  $28.125^\circ$ ). È stata aggiunta un'ulteriore posizione a  $(0^\circ, 90^\circ)$ . Il raggio della distanza dei punti è costante ( $r = 1$ ).

#### 4.1.4 Condizioni dell'esperimento

Per ogni soggetto dell'esperimento è stata selezionata la miglior HRTF personalizzata utilizzando l'applicazione descritta nel capitolo 2 secondo i seguenti criteri:

- **A:** miglior HRTF selezionata equipesando le 3 classifiche in frequenza<sup>3</sup>, metrica 1 (Contorno  $C_1$  con coefficiente di riflessione negativo).
- **B:** HRTF del KEMAR.
- **C:** miglior HRTF selezionata per mismatch antropometrico: altezza interna eq. 2.3.

#### 4.1.5 Procedura

##### 4.1.5.1 Acquisizione delle foto del padiglione

Per aggiungere un soggetto al database dell'applicazione è necessario scattare una foto del padiglione auricolare. È importante che la procedura di acquisizione sia la stessa per ogni persona. Pertanto l'individuo si posizionerà seduto su una sedia di fronte ad uno specchio, in modo che il riflesso sia esattamente al centro fra una riga nera. Le foto dell'orecchio destro sono state scattate a tre distanze diverse<sup>4</sup> con lo smartphone Samsung Galaxy S7 utilizzato per il test in realtà virtuale. Inoltre per facilitare l'impostazione del *pixel to meter* in prossimità del padiglione è stato piazzato un metro.

<sup>3</sup>Score ottenuto sommando  $i$  punti assegnati in base alla  $i$ -esima posizione per la classifica di *Average mismatch*, *Average rank* e *Top-5 Appearance*. La HRTF con il punteggio  $i$  più basso è la candidata.

<sup>4</sup>50cm, 100cm, 150cm



**Figura 4.4:** Foto scattate al soggetto 1 a tre differenti distanze.

**Tabella 4.1:** Sequenze del latin square

	0	1	2
0	C	A	B
1	A	B	C
2	B	C	A

#### 4.1.5.2 Selezione della HRTF

Una volta inserita la foto nell'applicazione ed impostato il *pixel to meter* si può procedere alla selezione della HRTF personalizzata. Sono state tracciate  $N = 10$  linee per il contorno  $C_1$ ,  $K = 10$  punti per il canale uditivo e l'altezza interna del padiglione auricolare. Il parametro  $M$  è stato impostato a 5.

#### 4.1.5.3 Ordine dei test

È ragionevole pensare che i soggetti potessero abituarsi al test in realtà virtuale man mano che i test procedevano, con il rischio di trovare sistematicamente nella terza prova performance in media migliori. Per ovviare a ciò le 3 condizioni (A,B,C) sono state proposte in ordine variabile. Abbiamo deciso di utilizzare il metodo del *latin square*: al soggetto  $i$ , partendo da 0, è stata assegnata la sequenza  $i \bmod 3$  (vedi tabella 4.1).

## 4.2 Risultati

### 4.2.1 Errori in azimuth ed elevazione

Per il calcolo dell'errore in azimuth abbiamo deciso di prendere il valore minimo di errore fra la posizione percepita e la posizione reale o del cono di confusione:

$$X_{Err_1} = \begin{cases} ||x_{real} - x_{perc}| - 360^\circ| & \text{se } |x_{real} - x_{perc}| > 180^\circ \\ |x_{real} - x_{perc}| & \text{altrimenti} \end{cases} \quad (4.1)$$

$$X_{Err_2} = \begin{cases} ||Cone(x_{perc}) - x_{real}| - 360^\circ| & \text{se } |Cone(x_{perc}) - x_{real}| > 180^\circ \\ |Cone(x_{perc}) - x_{real}| & \text{altrimenti} \end{cases} \quad (4.2)$$

$$X_{Err} = \min\{X_{Err_1}, X_{Err_2}\} \quad (4.3)$$

dove Cone(x) da' come output il valore di azimuth del cono di confusione che ha la stessa differenza angolare in assoluto a  $x$  rispetto all'origine della semisfera ( $0^\circ$  per la frontale,  $-180^\circ$  or  $180^\circ$  per la posteriore).  $x_{real}$  è l'azimuth della sorgente del suono e  $x_{perc}$  è il valore di azimuth puntato dal soggetto.

L'errore in elevazione è calcolato come il modulo della differenza fra elevazione percepita ed elevazione reale della sorgente:

$$Y_{Err} = |y_{real} - y_{perc}| \quad (4.4)$$

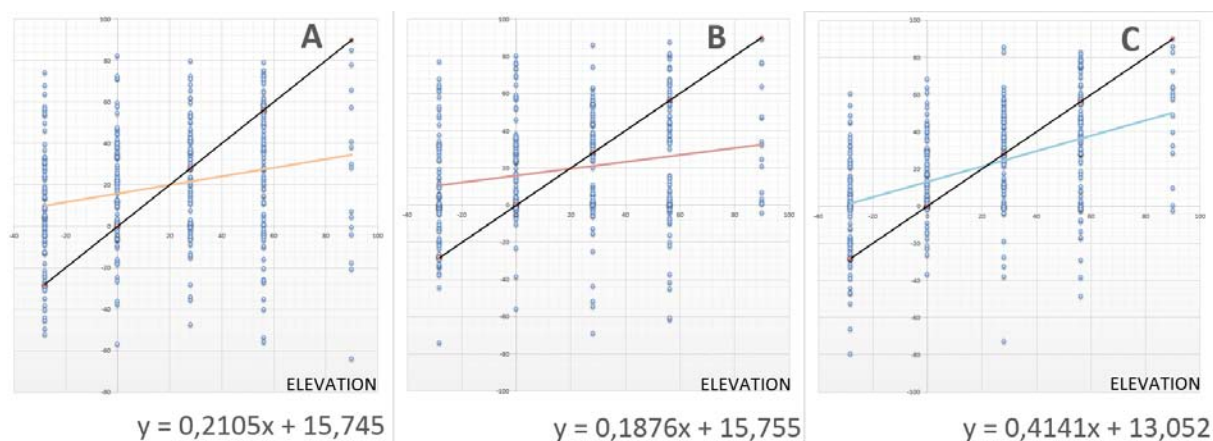
### 4.2.2 Risultati principali

Prendendo tutti i dati dei 15 soggetti, l'errore in azimuth ed elevazione sono calcolati usando le eq. 4.3 ed 4.4 definite sopra. La tabella 4.2 mostra i risultati medi di ogni soggetto per ogni condizione. La tabella 4.3 mostra la deviazione standard di tutti i valori per ogni soggetto.

La figura 4.5 mostra la regressione lineare della percezione in elevazione utilizzando i dati di tutti i soggetti per tutte e tre le condizioni. Più la pendenza è tendente a 1, più l'elevazione è percepita correttamente.

AllSubjects						
	AVG Azimuth			AVG Elevation		
	A	B	C	A	B	C
1	10,13	8,06	5,56	21,56	21,83	21,18
2	9,13	13,76	14,51	23,46	23,91	17,89
3	21,14	83,04	10,96	41,04	35,34	23,60
4	5,90	9,08	8,33	30,41	34,13	19,90
5	24,48	28,24	21,28	39,50	30,87	29,52
6	20,63	16,82	12,73	20,43	28,55	17,88
7	23,44	16,42	10,06	23,44	27,02	27,12
8	19,55	17,51	17,14	29,77	38,74	28,17
9	11,73	15,42	11,69	30,86	33,25	26,83
10	15,70	25,36	32,14	37,17	44,08	34,30
11	15,87	12,64	25,49	35,02	32,48	37,47
12	9,88	10,69	13,84	28,77	32,33	20,21
13	10,44	15,10	12,26	27,34	25,92	25,56
14	21,04	16,80	10,20	31,30	40,36	28,71
15	14,67	17,61	11,61	24,18	30,15	27,43
AVG	15,58	20,44	14,52	29,62	31,93	25,72

**Tabella 4.2:** Errore medio in azimuth (giallo) ed errore medio in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto.



**Figura 4.5:** Regressione lineare in elevazione su tutti i soggetti per ogni condizione.

AllSubjects						
	StD Azimuth			StD Elevation		
	A	B	C	A	B	C
1	14,46	17,81	5,70	16,74	21,50	15,84
2	7,81	17,21	17,09	22,58	23,43	15,07
3	20,80	55,16	13,17	28,02	21,81	17,92
4	6,07	6,61	8,03	34,15	24,03	20,96
5	21,33	22,81	26,33	30,33	23,58	25,84
6	22,67	12,81	10,57	18,15	26,06	11,84
7	17,18	16,82	13,85	17,18	24,94	20,37
8	15,65	15,09	13,64	24,39	26,93	21,06
9	11,60	14,23	8,74	19,89	24,31	19,10
10	14,08	20,34	22,53	28,23	32,40	22,76
11	19,71	12,83	29,72	26,74	29,26	30,93
12	9,56	14,62	12,29	20,94	24,40	20,01
13	6,69	18,90	9,69	16,51	18,25	18,35
14	16,23	18,81	10,34	25,59	31,10	21,41
15	15,48	21,10	13,78	18,96	21,91	17,97
AVG	14,62	19,01	14,37	23,23	24,93	19,96

**Tabella 4.3:** Deviazione standard in azimuth (giallo) e deviazione standard in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto.

	FrontBackALL				SlopesALL		
	A	B	C		A	B	C
1	1	4	1	1	0,404	0,501	0,660
2	1	0	1	2	0,245	0,254	0,547
3	1	22	2	3	0,234	0,129	0,560
4	1	1	1	4	0,136	0,367	0,488
5	6	5	5	5	0,080	0,020	0,191
6	3	3	1	6	0,462	0,324	0,817
7	5	3	1	7	0,046	0,219	0,233
8	1	2	2	8	0,281	0,011	0,276
9	3	4	1	9	0,190	0,000	0,291
10	3	7	5	10	-0,128	-0,106	-0,012
11	2	3	3	11	0,183	0,414	0,382
12	0	0	2	12	0,200	0,013	0,344
13	1	1	0	13	0,288	0,301	0,327
14	3	2	3	14	0,210	0,320	0,670
15	3	2	3	15	0,480	0,016	0,420
AVG	2,27	3,93	2,07	AVG	0,22	0,19	0,41

**Tabella 4.4:** Numero di front-back confusion per ogni condizione (sinistra). Pendenze della regressione lineare in elevazione per ogni condizione (destra).

### 4.2.3 Curva di apprendimento

Il primo aspetto da approfondire è quanto i soggetti hanno migliorato le loro prestazioni test dopo test. Facendo la media degli errori ad ogni sessione ed organizzandoli nella tabella 4.5 seguendo le sequenze del quadrato latino è stato facile verificarlo. I valori sono talmente simili da affermare che non ci sono stati miglioramenti qualitativi dal primo, al secondo e terzo test.

AVG ERR. AZIMUTH (ALL SUBJECTS-Learning)		
1	2	3
18,88	13,46	15,04
AVG ERR. ELEVATION (ALL SUBJECTS-Learning)		
1	2	3
27,92	26,39	27,50

**Tabella 4.5:** AVG azimuth error (yellow) and AVG elevation (green) of the VR experiment test for each subject and number of session

NO 3-5-10						
AVG Azi			AVG Ele			
	A	B	C	A	B	C
1	10,13	8,06	5,56	21,56	21,83	21,18
2	9,13	13,76	14,51	23,46	23,91	17,89
4	5,90	9,08	8,33	30,41	34,13	19,90
6	20,63	16,82	12,73	20,43	28,55	17,88
7	23,44	16,42	10,06	23,44	27,02	27,12
8	19,55	17,51	17,14	29,77	38,74	28,17
9	11,73	15,42	11,69	30,86	33,25	26,83
11	15,87	12,64	25,49	35,02	32,48	37,47
12	9,88	10,69	13,84	28,77	32,33	20,21
13	10,44	15,10	12,26	27,34	25,92	25,56
14	21,04	16,80	10,20	31,30	40,36	28,71
15	14,67	17,61	11,61	24,18	30,15	27,43
AVG	14,37	14,16	12,79	27,21	30,72	24,86

**Tabella 4.6:** Errore medio in azimuth (giallo) ed errore medio in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. Rimossi i soggetti 3, 5 e 10.

#### 4.2.4 Eliminazione dei soggetti

A seguito di un'analisi più approfondita dei risultati possiamo scartare alcuni soggetti che hanno avuto delle prestazioni anomale. In soggetto 3 può essere rimosso a causa del notevole numero di errori di front-back confusion nella condizione B. In aggiunta i soggetti 5 e 10 non hanno percepito alcuna differenza in elevazione in tutte e tre le condizioni. Eliminando questi soggetti sono state create delle nuove tabelle con risultati più accurati.

La tabella 4.8 mostra i valori di front-back confusion dei soggetti selezionati.

#### 4.2.5 T-Test

Sono stati effettuati dei paired t-test per ciascuna coppia di condizioni. Le tabelle seguenti (4.9-4.10-4.11-4.12-4.13-4.11) mostrano i risultati dei sei t-test. Il valore di rilevanza  $p$  dei t-test multipli seguendo la correzione Bonferroni sarà di 0.025

In **azimuth**, nessuna coppia di condizioni risulta con differenze statisticamente significative in quanto il valore di  $p > 0.025$ .

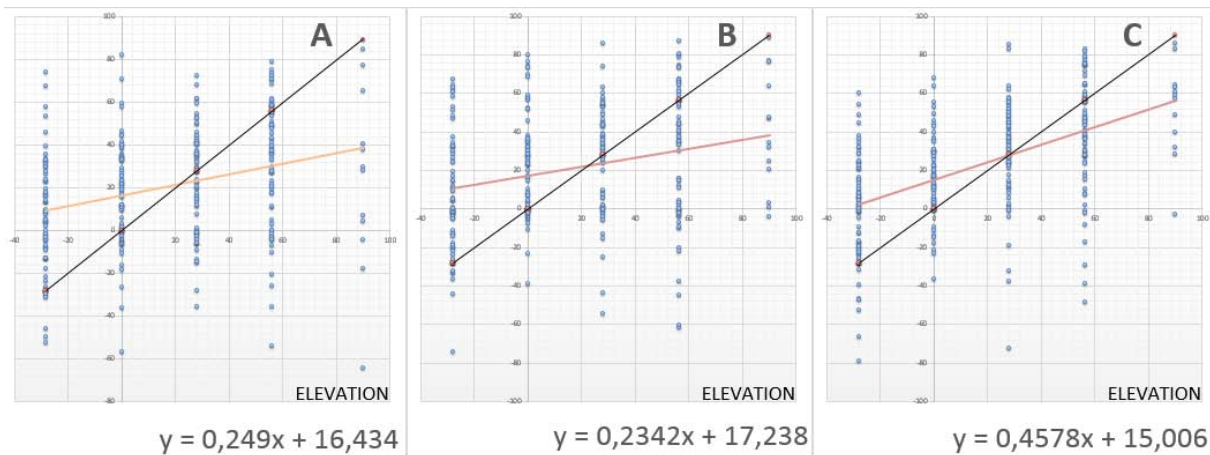
NO 3-5-10						
	StD Azimuth			StD Elevation		
	A	B	C	A	B	C
1	14,46	17,81	5,70	16,74	21,50	15,84
2	7,81	17,21	17,09	22,58	23,43	15,07
4	6,07	6,61	8,03	34,15	24,03	20,96
6	22,67	12,81	10,57	18,15	26,06	11,84
7	17,18	16,82	13,85	17,18	24,94	20,37
8	15,65	15,09	13,64	24,39	26,93	21,06
9	11,60	14,23	8,74	19,89	24,31	19,10
11	19,71	12,83	29,72	26,74	29,26	30,93
12	9,56	14,62	12,29	20,94	24,40	20,01
13	6,69	18,90	9,69	16,51	18,25	18,35
14	16,23	18,81	10,34	25,59	31,10	21,41
15	15,48	21,10	13,78	18,96	21,91	17,97
AVG	13,59	15,57	12,79	21,82	24,68	19,41

**Tabella 4.7:** Deviazione standard in azimuth (giallo) e deviazione standard in elevazione (verde) dell'esperimento per ogni soggetto del test in VR, condizioni (A,B,C). Una riga per soggetto. Rimossi i soggetti 3, 5 e 10.

	FrontBack				Slopes		
	A	B	C		A	B	C
1	1	4	1	1	0,404	0,501	0,660
2	1	0	1	2	0,245	0,254	0,547
4	1	1	1	4	0,136	0,367	0,488
6	3	3	1	6	0,462	0,324	0,817
7	5	3	1	7	0,046	0,219	0,233
8	1	2	2	8	0,281	0,011	0,276
9	3	4	1	9	0,190	0,000	0,291
11	2	3	3	11	0,183	0,414	0,382
12	0	0	2	12	0,200	0,013	0,344
13	1	1	0	13	0,288	0,301	0,327
14	3	2	3	14	0,210	0,320	0,670
15	3	2	3	15	0,480	0,016	0,420
AVG	2,00	2,08	1,58	AVG	0,26	0,23	0,45

**Tabella 4.8:** Numero di front-back confusion per ogni condizione (sinistra). Pendenze della regressione lineare in elevazione per ogni condizione (destra). Rimossi i soggetti 3, 5 e 10.





**Figura 4.6:** *Regressione lineare in elevazione sui soggetti per ogni condizione (No soggetti 3, 5 e 10).*

NO 3-5-10		
	A	B
AVG	14,37	14,16
Variance	32,26	11,12
Num Occurrences	12	12
Pearson Correlation	0,722115	
Stat t	0,181123	
P(T<=t) one tail	0,429782	

**Tabella 4.9:** *Paired t-test fra le condizioni A e B. Errori in azimuth. Rimossi i soggetti 3, 5, 10.*

In **elevazione**, le condizioni **A e B** e **B e C** risultano con differenze statisticamente significative con valori rispettivamente  $p = 0.005$  e  $p = 0.0034$  entrambi  $< 0.025$ . La condizione **A e C** invece non risulta avere differenze statisticamente significative con  $p = 0.046 > 0.025$ .

NO 3-5-10		
	A	C
AVG	14,37	12,79
Variance	32,26	24,85
Num Occurrences	12	12
Pearson Correlation	0,188078	
Stat t	0,803899	
P(T<=t) one tail	0,219242	

**Tabella 4.10:** Paired t-test fra le condizioni A e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.

NO 3-5-10		
	B	C
AVG	14,16	12,79
Variance	11,12	24,85
Num Occurrences	12	12
Pearson Correlation	0,21124171	
Stat t	0,88369988	
P(T<=t) one tail	0,19788094	

**Tabella 4.11:** Paired t-test fra le condizioni B e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.

NO 3-5-10		
	A	B
AVG	27,21	30,72
Variance	20,52	31,64
Num Occurrences	12	12
Pearson Correlation	0,720436	
Stat t	-3,09529	
P(T<=t) one tail	0,005095	

**Tabella 4.12:** Paired t-test fra le condizioni A e B. Errori in elevazione. Rimossi i soggetti 3, 5, 10.

NO 3-5-10		
	A	C
AVG	27,21	24,86
Variance	20,52	32,56
Num Occurrences	12	12
Pearson Correlation	0,650443	
Stat t	1,844141	
P(T<=t) one tail	0,046121	

**Tabella 4.13:** Paired *t*-test fra le condizioni A e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.

NO 3-5-10		
	B	C
AVG	30,72	24,86
Variance	31,64	32,56
Num Occurrences	12	12
Pearson Correlation	0,42078025	
Stat t	3,328633	
P(T<=t) one tail	0,00336377	

**Tabella 4.14:** Paired *t*-test fra le condizioni B e C. Errori in azimuth. Rimossi i soggetti 3, 5, 10.

# Capitolo 5

## Conclusioni

### 5.1 Analisi delle metriche

Al termine di questa analisi le metriche migliori, ottenute sommando il numero di occorrenze in tutte e tre le selezioni, risultano la 1 e la 3, rispettivamente con 16 e 14 presenze. Considerando la notevole predominanza nella scelta della metrica con HRTF del soggetto CIPIC ID 65, che aumenta il numero di metriche con contorno  $C_2$  scelte, questo risultato trova riscontro nello studio precedente di Michele Geronazzo et al. [20], dove era stato scelto di assegnare l'intero peso solamente al contorno  $C_1$ , e nello studio in fase di pubblicazione dove si considerava come metrica migliore la pesatura ottima (0.66, 0.24, 0.1). Con il metodo di tracciamento utilizzato per  $C_3$  possiamo notare nell'esempio di fig 3.9 che le frequenze con riflessione negativa tendono a sconfinare nel campo non udibile ( $f > 18kHz$ ), mentre quelle con riflessione positiva si discostano molto dalla traccia spettrale del contorno  $C_3$  della HRTF individuale. Inoltre si può vedere che per i 18 soggetti circa l'80% delle metriche migliori hanno coefficiente negativo, confermando le ipotesi del modello riflessivo (vedi sez. 1.2.4.4).

È di rilievo far notare che in tutti e 3 i metodi di selezione, sebbene non sia stata scelta la HRTF individuale per la maggior parte dei soggetti, le prestazioni di percezione uditiva sono risultate sempre migliori rispetto ad una HRTF generica del manichino KEMAR e molto vicine alle performance della HRTF individuale. Questo consente di ridurre il dataset di partenza HRTF ad un numero ridotto di migliori per il soggetto, per poi procedere ad una selezione più accurata delle restanti con metodi più precisi, come un test di ascolto in prima persona, più dispendiosi in tempo e attrezzature.

Per confermare con più certezza questi risultati si potrebbero effettuare dei nuovi tracciamenti con diverse modalità (ad es. coprendo aree più ridotte o più vaste), variare il parametro

$M$  per la classifica *Top-M Appearance* e fare una nuova analisi comparativa. Nonostante questi accorgimenti il campione di soggetti presi in esame (18) è di dimensioni ridotte per fornire dei dati statistici significativi. I lavori futuri avranno come principale obiettivo quello di replicare l'analisi su un database con un numero più elevato di set di HRTF (es. 40-50 soggetti).

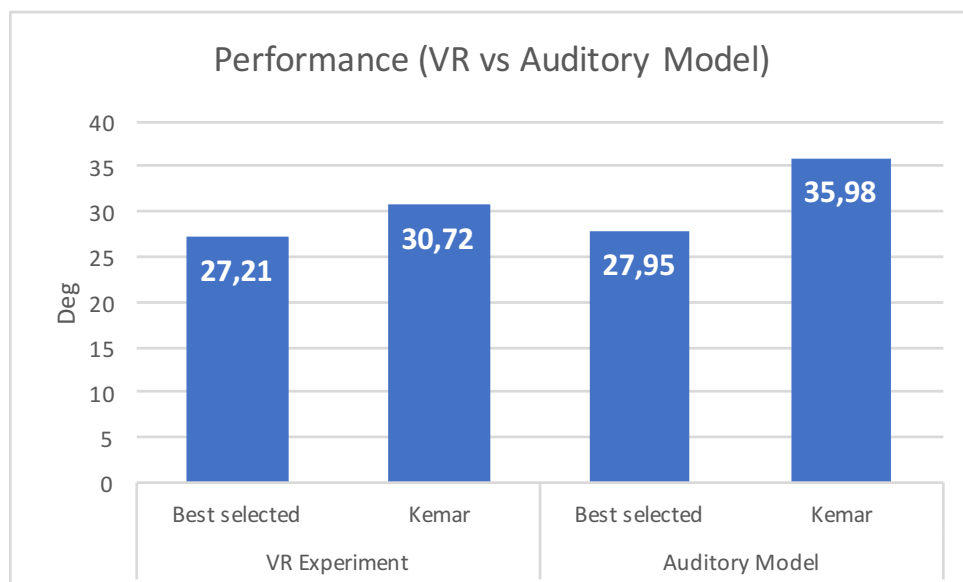
Infine questo studio può venir replicato utilizzando diverse condizioni di acquisizione foto (angolazioni e distanze) o utilizzando più misurazioni acustiche che raccolgono alcune informazioni della HRTF. Un esempio di quest'ultima direzione di ricerca è rappresentato dal seguente studio [15, 14] in cui un apparecchio mobile può venir utilizzato come sorgente sonora e dei microfoni binaurali acquisiscono un'approssimazione dell'acustica dell'orecchio esterno da cui possono venirne estratte alcune riflessioni.

## 5.2 Esperimento in VR

Riguardo la precisione della percezione in azimuth, analizzando i dati ed i valori dei t-test possiamo concludere che le tre condizioni hanno avuto le stesse prestazioni. Non ci sono state differenze significative con l'ascolto tramite una HRTF piuttosto che con un'altra. Il risultato è in accordo con la teoria per la quale le condizioni **A** e **C** sono basate sulle riflessioni del padiglione ed influenzano solo le performance in elevazione.

Per quanto riguarda i risultati in elevazione possiamo dire che la condizione **C** è risultata la migliore come prestazioni. In media gli errori in elevazione sono molto simili ma analizzando la regressione su tutti i dati mostra una pendenza doppia di **C** rispetto ad **A** e **B**. I risultati dei t-test hanno confermato che sia la condizione **A** che **C** differiscono statisticamente da **B** ma non fra di loro. Possiamo concludere che le condizioni **A** e **C** sono statisticamente migliori della condizione con HRTF generica del KEMAR.

Confrontando le performance della condizione **A** con i soggetti utilizzati nel test e le prestazioni medie dei soggetti CIPIC fornite nell'analisi con il modello uditivo con selezione di *Top-5 Appearance* (vedi fig. 5.1) vediamo che si discostano di poco. Questo risultato conferma la corrispondenza fra i valori simulati con modello uditivo nel test di localizzazione per i soggetti CIPIC con i valori del test di localizzazione in VR con soggetti reali.



**Figura 5.1:** Prestazioni delle HRTF selezionate in ascolto nell'esperimento in realtà virtuale e nella simulazione con modello uditivo.



# Bibliografia

- [1] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001.
- [2] A. Andreopoulou, D. Begault, and B. Katz. Inter-Laboratory Round Robin HRTF Measurement Comparison. *IEEE Journal of Selected Topics in Signal Processing*, PP(99):1–1, 2015.
- [3] R. Baumgartner, P. Majdak, and B. Laback. Assessment of Sagittal-Plane Sound Localization Performance in Spatial-Audio Applications. In Jens Blauert, editor, *The Technology of Binaural Listening, Modern Acoustics and Signal Processing*, pages 93–119. Springer Berlin Heidelberg, January 2013.
- [4] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, USA, 1983.
- [5] Braxton Boren, Michele Geronazzo, Fabian Brinkmann, and Edgar Choueiri. Coloration Metrics for Headphone Equalization. In *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, pages 29–34, Graz, Austria, July 2015.
- [6] Braxton B. Boren, Michele Geronazzo, Piotr Majdak, and Edgar Choueiri. PHOnA: A Public Dataset of Measured Headphone Transfer Functions. In *Proc. 137th Conv. Audio Eng. Society*. Audio Engineering Society, October 2014.
- [7] W. G. Gardner and K. D. Martin. HRTF Measurements of a KEMAR. *J. of the Acoustical Society of America*, 97(6):3907–3908, June 1995.
- [8] M. Geronazzo, A. Carraro, and F. Avanzini. Evaluating vertical localization performance of 3d sound rendering models with a perceptual metric. In *2015 IEEE 2nd VR Workshop*



on *Sonic Interactions for Virtual Environments (SIVE)*, pages 1–5, Arles, France, March 2015. IEEE Computer Society.

- [9] Michele Geronazzo. *Mixed structural models for 3D audio in virtual environments*. Ph.D. Thesis, University of Padova, Padova, Italy, April 2014.
- [10] Michele Geronazzo, Federico Avanzini, and Federico Fontana. Use of Personalized Binaural Audio and Interactive Distance Cues in an Auditory Goal-Reaching Task. In *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, pages 73–80, Graz, Austria, July 2015.
- [11] Michele Geronazzo, Federico Avanzini, and Federico Fontana. Auditory navigation with a tubular acoustic model for interactive distance cues and personalized head-related transfer functions. *J Multimodal User Interfaces*, 10(3):273–284, September 2016.
- [12] Michele Geronazzo, Alberto Bedin, Luca Brayda, and Federico Avanzini. Multimodal Exploration of Virtual Objects with a Spatialized Anchor Sound. In *Proc. 55th Int. Conf. Audio Eng. Society, Spatial Audio*, pages 1–8, Helsinki, Finland, August 2014.
- [13] Michele Geronazzo, Alberto Bedin, Luca Brayda, Claudio Campus, and Federico Avanzini. Interactive spatial sonification for non-visual exploration of virtual maps. *International Journal of Human-Computer Studies*, 85:4–15, January 2016.
- [14] Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. Acoustic Selfies for Extraction of External Ear Features in Mobile Audio Augmented Reality. In *Proc. 22nd ACM Symposium on Virtual Reality Software and Technology (VRST 2016)*, pages 23–26, Munich, Germany, November 2016. ACM.
- [15] Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. The Selfear Project: a Mobile Application for Low-Cost Pinna-Related Transfer Function Acquisition. In *Proc. 13th Int. Conf. Sound and Music Computing (SMC 2016)*, pages 164–171, Hamburg, Germany, September 2016.
- [16] Michele Geronazzo, Jari Kleimola, and Piotr Majdak. Personalization support for binaural headphone reproduction in web browsers. In *Proc. 1st Web Audio Conference*, 2015.
- [17] Michele Geronazzo, Enrico Peruch, Fabio Prandoni, and Federico Avanzini. Improving elevation perception with a tool for image-guided head-related transfer function selection.

- In (*DAFx-17*) proceedings of the 20th International Conference on Digital Audio Effects, Sept. 5-9, September, UK, 2017. Accepted for publication.
- [18] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Estimation and modeling of pinna-related transfer functions. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, pages 6–10, 2010.
- [19] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery. In *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, pages 1–8, Santorini, Greece, July 2013.
- [20] Michele Geronazzo, Simone Spagnol, Alberto Bedin, and Federico Avanzini. Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 4463–4467. IEEE, 2014.
- [21] Yukio Iwaya. Individualization of head-related transfer functions with tournament-style listening test: Listening with other’s ears. *Acoustical science and technology*, 27(6):340–343, 2006.
- [22] Brian FG Katz and Gaëtan Parseihian. Perceptually based head-related transfer function database optimization. *The Journal of the Acoustical Society of America*, 131(2):EL99–EL105, 2012.
- [23] Piotr Majdak, Peter Balazs, and Bernhard Laback. Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. *JAES*, 55(7/8):623–637, July 2007.
- [24] Sebastian Prepelită, Michele Geronazzo, Federico Avanzini, and Lauri Savioja. Influence of Voxelization on Finite Difference Time Domain Simulations of Head-Related Transfer Functions. *J. Acoust. Soc. Am.*, 139(5):2489–2504, May 2016.
- [25] Vikas C Raykar, Ramani Duraiswami, and B Yegnanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *The Journal of the Acoustical Society of America*, 118(1):364–374, 2005.
- [26] Ludivine Sarlat, Olivier Warusfel, and Isabelle Viaud-Delmon. Ventriloquism aftereffects occur in the rear hemisphere. *Neuroscience letters*, 404(3):324–329, 2006.

- [27] Patrick Satarzadeh. *A study of physical and circuit models of the human pinnae*. PhD thesis, UNIVERSITY OF CALIFORNIA DAVIS, 2006.
- [28] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. On the relation between pinna reflection patterns and head-related transfer function features. *IEEE transactions on audio, speech, and language processing*, 21(3):508–519, 2013.
- [29] Elizabeth M Wenzel, Marianne Arruda, Doris J Kistler, and Frederic L Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993.
- [30] Satoshi Yairi, Yukio Iwaya, and Yôiti Suzuki. Individualization feature of head-related transfer functions based on subjective evaluation. In *Proc. of International Conference on Auditory Display (ICAD2008), Paris, 2008*.
- [31] DYN Zotkin, Jane Hwang, R Duraiswaini, and Larry S Davis. Hrtf personalization using anthropometric measurements. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 157–160. Ieee, 2003.