

Università degli Studi di Padova
Facoltà di Ingegneria
Corso di Laurea Triennale in Ingegneria dell'Informazione

tesi di laurea

**Valutazione soggettiva di un
modello parametrico di orecchio
esterno per il rendering
binaurale del suono**

Relatore: Prof. Federico Avanzini
Correlatore: Dott. Simone Spagnol

Laureando: Sorato Giacomo

17 febbraio 2014

Prefazione

Quando ci accingiamo ad ascoltare con le nostre cuffie un brano registrato, prestando le dovute attenzioni ci accorgiamo che non vi sono grandi interazioni tra suono e spazio. Se ascoltassimo la registrazione di un concerto e la relazionassimo all'ascolto live, ci accorgeremmo ancora meglio di quanto questa relazione tra suono e spazio venga assottigliata dagli attuali metodi di registrazione. Vi sono tuttavia delle tecniche che mirano a realizzare una scena sonora coinvolgente, capace di far provare una sensazione di totale immersione nell'ambiente musicale desiderato. A tal fine, la tecnologia complessa dell'audio binaurale è una delle possibilità maggiormente accattivanti. A livello tecnologico il risultato da noi richiesto può essere raggiunto cercando di ricreare le stesse pressioni sonore che i due timpani misurano durante le condizioni di ascolto quotidiano e naturale. Sorge quindi la necessità di modellare il suono nello spazio, e qui entra in gioco questa tesi, il cui scopo consiste nella realizzazione di un esperimento pilota per testare la bontà di un modello di orecchio esterno (apparato dell'orecchio esterno alla testa) progettato precedentemente nei laboratori del Sound and Music Computing Group del Dipartimento di Ingegneria dell'Informazione dell'Università di Padova, che mira a modellare il suono nel piano verticale attraverso dei parametri estratti da una fotografia dell'orecchio di un qualsiasi ascoltatore. L'esperimento consisterà quindi nell'estrazione di tali parametri dalla fotografia dell'orecchio sinistro del soggetto per processare dei suoni ad hoc, ed attraverso una GUI grafica accompagnata da dispositivi di riproduzione testare come vengono percepiti. Le applicazioni di tale studio trovano spazio in tutto ciò che riguarda la realtà virtuale aumentata, cioè dai videogames all'ausilio per la riabilitazione, dall'audio per il cinema all'orientamento per persone non vedenti, caratterizzandosi quindi non solo in campi artistici e ludici ma anche in settori medici.

L'autore.

Sommario

Capitolo 1 - Qui viene presentata una breve introduzione su quanto necessario per comprendere il concetto di rendering spaziale del suono. Viene introdotta una descrizione delle componenti anatomiche che concorrono alla resa spaziale del suono concentrandosi sulla parte relativa all'orecchio. Viene data una descrizione qualitativa dell'audio binaurale e si introducono i concetti di HRTF e di modello strutturale.

Capitolo 2 - In questo capitolo si approfondisce il modello parametrico di HRTF concentrandoci sullo studio dei principali effetti introdotti dalla testa e soprattutto dal padiglione auricolare (pinna). Di quest'ultima viene analizzato il modello a blocchi evidenziando quali filtri sono stati utilizzati. Viene anche esposto il procedimento di estrazione dei parametri dal contorno della pinna, definendo la PRTE. Viene dunque esposto il modello completo con un esempio di sua applicazione.

Capitolo 3 - Il cuore del lavoro effettuato risiede in questo capitolo, viene presentata tutta la descrizione del protocollo sperimentale adottato per l'esperimento pilota, definendo gli strumenti utilizzati, gli accorgimenti apportati e le varie procedure seguite per giungere ai risultati.

Capitolo 4 - L'ultimo capitolo contiene una descrizione di tutti i risultati, evidenziando quelle che sono le conclusioni ottenute dall'esperimento pilota realizzato.

Appendice A - Contiene tutti i grafici di output dell'esperimento.

Appendice B - Contiene le parti maggiormente rilevanti del codice elaborato.

Indice

Prefazione	III
Sommario	V
1 Il suono nello spazio	1
1.1 Descrizione fisica dell'onda sonora	1
1.2 Localizzazione sonora	2
1.2.1 La testa	3
1.2.2 L'orecchio esterno	5
1.3 Rendering Spaziale	7
1.3.1 Virtualizzazione del suono	7
1.3.2 Audio binaurale	7
1.4 Rendering spaziale tramite HRTF	9
1.4.1 Modello strutturale	10
2 Modello strutturale di HRTF	13
2.1 Modello strutturale di PRTF	13
2.1.1 Filtri	15
2.1.2 Estrazione parametri dalla pinna	18
2.1.3 Scelta dei parametri	18
2.2 Modello completo di HRTF	19
3 Protocollo sperimentale	21
3.1 Introduzione	21
3.2 Apparato	21
3.2.1 Matlab	21
3.2.2 Cuffie	21
3.2.3 Scheda audio	22
3.2.4 Camera silente	22
3.3 Soggetti	22
3.3.1 Presentazione dei soggetti	22
3.3.2 Caratterizzazione dei soggetti	24
3.4 Stimoli	24

3.5	Procedura	25
3.5.1	Inserimento dati	25
3.5.2	Acquisizione immagine e tracciamento	26
3.5.3	Test di ascolto	29
4	Risultati	31
4.1	Descrizione dell'output	31
4.1.1	Output grafico	31
4.1.2	Output parametrico	32
4.2	Analisi dettagliata dei risultati grafici	32
4.3	Analisi dettagliata dei risultati parametrici	34
4.4	Conclusioni	37
4.5	Sviluppi futuri	39
5	Appendice A	41
5.1	Output Grafici di tutti i soggetti	41
6	Appendice B	47
6.1	Codice	47
	Bibliografia	58
	Elenco delle tabelle	60
	Elenco delle figure	62

Capitolo 1

Il suono nello spazio

1.1 Descrizione fisica dell'onda sonora

Il suono è definito mediante il concetto di *onda*, ovvero una qualsiasi perturbazione di una grandezza fisica, impulsiva o periodica, che si propaga con una velocità ben definita. L'onda sonora viene generata da movimenti vibratori che inducono oscillazioni di particelle intorno alla loro posizione di riposo e lungo la direzione di propagazione dell'onda. Sono gli urti tra queste particelle a consentire al fenomeno di propagarsi, dunque l'onda sonora, per questa sua caratteristica di interagire meccanicamente con il mezzo, prende il nome di onda elastica longitudinale. Osservando la propagazione si caratterizza quindi una variazione locale della pressione nel mezzo, tale caratteristica fisica è il fondamento della definizione ondulatoria del suono. La pressione è la grandezza fisica la cui perturbazione definisce l'onda sonora.

La *velocità* con cui il suono si propaga è proporzionale alla radice quadrata del rapporto tra l'elasticità del mezzo e la sua densità, quindi ci aspettiamo che in mezzi poco densi la velocità aumenti. In realtà l'elasticità dei mezzi è molto variabile (basti pensare allo sforzo per comprimere una colonna d'aria e di acqua), il che rende la trattazione fortemente dipendente dal valore di entrambi i parametri:

$$v \sim \sqrt{\frac{K}{\rho}}$$

Dove K è l'elasticità e ρ la densità.

Per quanto riguarda lo studio dell'*intensità* delle onde sonore si deve introdurre il concetto di energia trasportata da un'onda, definita dal lavoro prodotto dalle varie forze elastiche che agiscono sulle particelle del mezzo. Quando si produce una deformazione una porzione del mezzo esce dall'equilibrio, quindi una forza elastica la richiama verso la posizione iniziale. Quando tuttavia la porzione del mezzo si trova di nuovo nella posizione iniziale, essa, a causa della sua

inerzia, tenderà a continuare il suo moto, trovandosi presto fuori dall'equilibrio in verso opposto alla deformazione iniziale. Il ciclo si ripeterà per tutta la durata della propagazione, e le varie forze elastiche di richiamo e di allontanamento dall'equilibrio concorreranno alla produzione di lavoro, che si traduce in energia. Definiamo dunque come intensità dell'onda sonora il valore medio dell'energia che passa attraverso una sezione ortogonale alla direzione di propagazione per unità di tempo e per unità di area:

$$I = \frac{1}{S} \left(\frac{dU}{dt} \right)$$

dove I è l'intensità, S l'area della sezione e U l'energia.

Il concetto di intensità porta allo studio di un'altra proprietà dell'onda, definita dalla sorgente che la ha generata: il *fronte d'onda*. Esso può essere comunemente di tipo sferico, cilindrico o piano ed ha un'influenza notevole sull'intensità di un'onda sonora mediante il parametro S . Per quanto riguarda i fronti d'onda sferici (generati solitamente da sorgenti puntiformi), il flusso di energia emessa (cioè l'energia che passa attraverso una area unitaria a distanza r dalla sorgente) dovrà decrescere come $\frac{1}{r^2}$ dovendo attraversare superfici che invece crescono come il quadrato della distanza. Se la sorgente sonora ha una simmetria cilindrica ci aspettiamo che il flusso di energia emessa decresca come $\frac{1}{r}$ dovendo attraversare superfici la cui area aumenta linearmente con la distanza dalla sorgente. Se la sorgente sonora è invece piana (teoricamente un piano infinito) il fronte d'onda è tale che il flusso di energia attraverso un piano parallelo alla sorgente rimane costante.

Le caratteristiche del suono e quelle spaziali vengono ulteriormente modificate dall'interazione dell'onda sonora con altre onde sonore o con mezzi diversi da quello dove è stata generata. Tali fenomeni prendono i nomi di riflessione e diffrazione (interazione onda-materia) ed interferenza (interazione onda-onda), un esempio pratico di queste si riscontra nel fenomeno del *riverbero* dove si ha una sequenza di piccole riflessioni che arrivano ad intervalli regolari al nostro orecchio, permettendoci di ascoltare un suono meno piatto e più corposo.

1.2 Localizzazione sonora

L'abilità di localizzare la sorgente di un determinato suono riveste un ruolo importante nella vita di tutti i giorni. Alcuni fenomeni ci sono di aiuto in questo compito, il riverbero ad esempio svolge un ruolo fondamentale nel permetterci di individuare alcune caratteristiche di una ipotetica stanza nella quale viene prodotto un suono: dimensioni, tipo di materiale delle pareti e così via. Ci fornisce anche dei suggerimenti sulla distanza e direzione dalla sorgente, e un senso di immersione nel suono. Tuttavia, la localizzazione è possibile anche in ambienti anecoici (ovvero in assenza di qualsivoglia riflessione ambientale dell'onda sonora), grazie al contributo di parti anatomiche quali la testa, il busto e le orecchie.

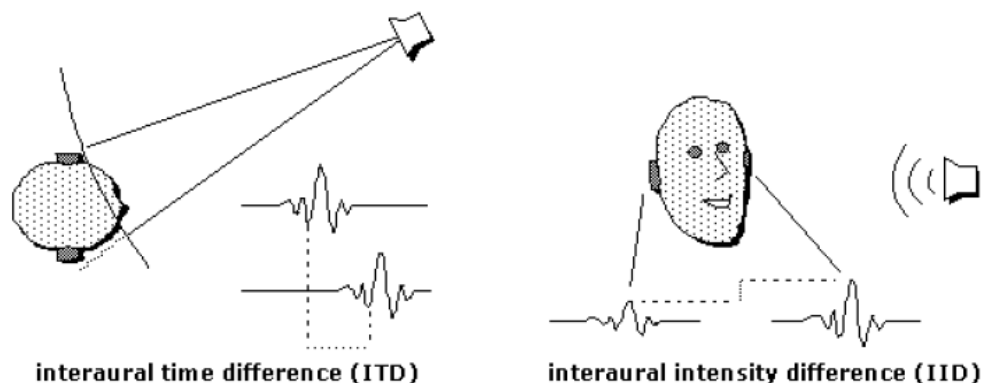


Figura 1.1: Visualizzazione grafica di ITD ed IID

Le applicazioni pratiche dell'uso di tali informazioni sono svariate e negli ultimi anni stanno notevolmente evolvendo, un esempio pratico sono i grandi impianti audio dei cinema, dove la spazializzazione del suono diventa il fulcro cardine dell'immersione nella storia proiettata sullo schermo. Vediamo qual è l'influenza delle componenti anatomiche grazie alle quali è possibile localizzare le sorgenti sonore.

1.2.1 La testa

Di certo gli elementi che concorrono maggiormente alla percezione del suono sono le nostre orecchie. Esse però non sono oggetti isolati nello spazio, sono posizionate alla stessa altezza ai due lati della testa, che ha un'alta impedenza acustica. Questa costituisce un ostacolo per la libera propagazione del suono e ha due principali effetti sulla percezione sonora, mostrati schematicamente in Figura 1.1:

- ITD (Interaural Time Difference), causato dal fatto che le onde sonore devono percorrere una maggiore distanza per raggiungere l'orecchio più lontano. Uno studio approssimato ma abbastanza preciso di questo fenomeno si può effettuare facendo delle ipotesi semplificative: si suppongono la sorgente lontana, la testa sferica, e le orecchie poste agli antipodi. La prima assunzione implica che le onde sonore che colpiscono la testa saranno onde piane; l'ulteriore distanza Δx che un raggio sonoro dovrà percorrere per raggiungere l'orecchio più lontano, viene stimata da considerazioni geometriche elementari (vedi Figura 1.2):

$$ITD \sim \frac{a}{c}(\theta + \sin\theta)$$

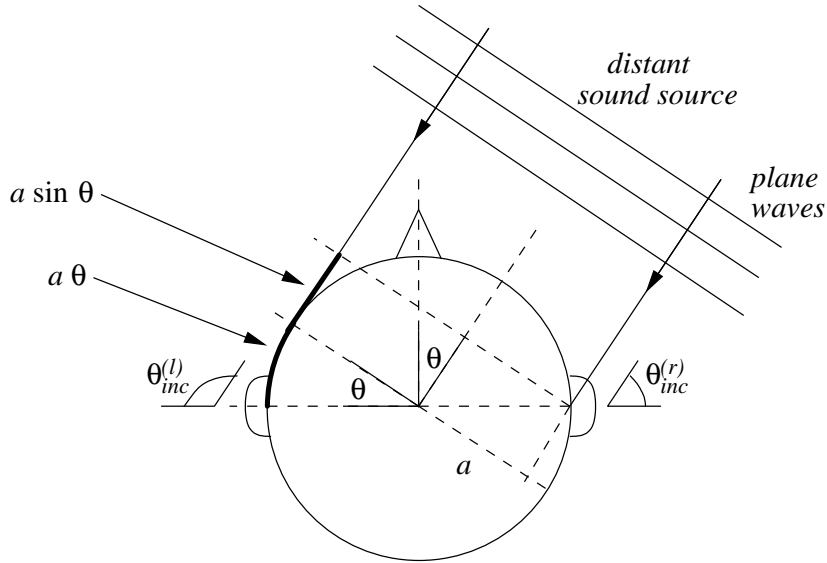


Figura 1.2: Stima di ITD nel caso di sorgente lontana (onde piane) e testa sferica

dove c è la velocità del suono, a rappresenta il raggio della testa e θ l'angolo di azimut che definisce la direzione della sorgente sonora sul piano orizzontale. La formula mostra che il minimo ITD si ha quando la sorgente è frontale ($\theta = 0$), mentre è massimo quando la sorgente è ai lati della testa ($\theta = \pi/2$).

- ILD (Interaural Level Difference), causato dal fatto che all'orecchio più lontano giunge un suono di minore intensità causato dall'introduzione da parte della testa di un'attenuazione. Mentre è accettabile supporre l'ITD come un parametro indipendente dalla frequenza, l'ILD è invece fortemente legato alla frequenza: a basse frequenze (per lunghezze d'onda dell'ordine del diametro della testa) non vi è quasi alcuna differenza, mentre ad alte frequenze diventa particolarmente significativo. Come già fatto in precedenza, l'ILD può essere analizzato nel caso ideale di testa perfettamente sferica (di raggio a) e con la sorgente sonora situata ad una distanza $r > a$ dal centro della sfera. È solito utilizzare costanti normalizzate: $\mu = \frac{\omega a}{c}$ (frequenza normalizzata) e $\rho = \frac{r}{a}$ (distanza normalizzata). Se si considera un punto sulla sfera allora la diffrazione di un'onda acustica causata dalla sfera nel punto scelto si può esprimere con la funzione di trasferimento:

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

dove P_m e h_m sono, rispettivamente, i polinomi di Legendre di ordine m -esimo e la funzione sferica di Hankel; θ_{inc} rappresenta l'angolo di incidenza. È noto che la funzione di Hankel $h_m(x)$ ammette un'approssimazione asintotica quando l'argomento x tende all'infinito. Dal grafico in Figura 1.3 si nota che a bassa frequenza la funzione di trasferimento non è dipendente dalla direzione e il suo modulo $|H_{sphere}(\rho, \theta_{inc}, \mu)|$ è pressoché costante ed uguale a 1 qualunque sia l'angolo. Se $\mu > 1$ la dipendenza dall'angolo di incidenza diventa molto sensibile. È da notare inoltre che la risposta minima non si presenta per $\theta_{inc} = \pi$: questo è dovuto al fatto che tutte le onde sonore che si propagano attorno alla sfera arrivano in quel punto in fase; a frequenze elevate questo fenomeno diventa molto meno presente e la parte posteriore della sfera risulta essere in una zona di ombra sonora. La differenza tra i moduli di questa funzione di trasferimento per l'orecchio destro e l'orecchio sinistro definisce l'ILD.

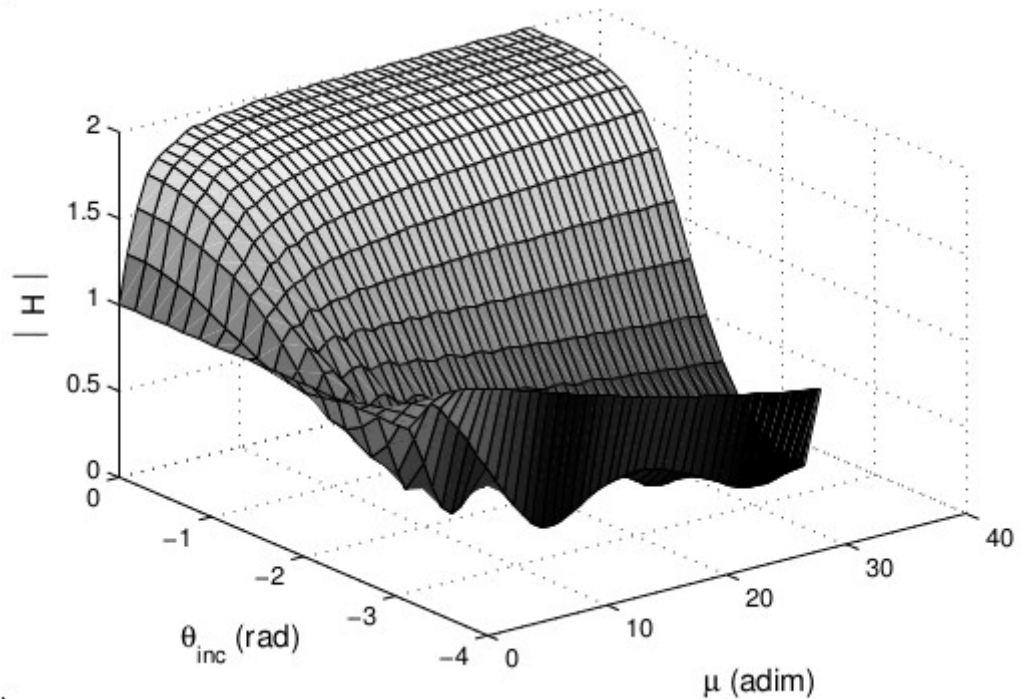


Figura 1.3: Modulo di una sfera $H_{sphere}(\infty, \theta_{inc}, \mu)$ nel caso di una sorgente a distanza infinita

1.2.2 L'orecchio esterno

Quando ci si riferisce all'orecchio esterno si sta sostanzialmente parlando sia della pinna, che rappresenta la parte effettivamente visibile, sia del condotto uditivo

che conduce al timpano. Diversi studi hanno accertato che la pinna modifica significativamente il suono in arrivo, specialmente alle alte frequenze. Tutto ciò concorre ovviamente a influire sulla nostra capacità di localizzare i suoni.

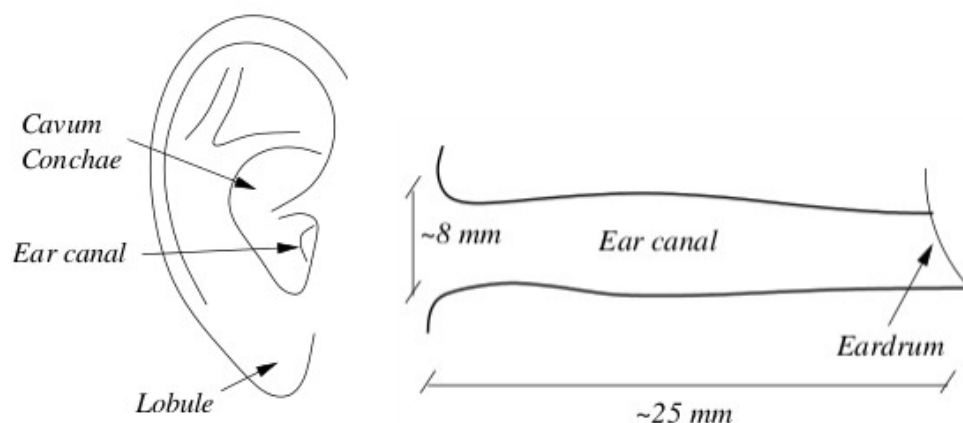


Figura 1.4: Orecchio esterno: pinna e condotto uditivo

Analizzando cosa accade quando un suono giunge al nostro orecchio si osserva che il suono, dopo aver attraversato i condotti uditivi raggiunge i timpani provocandone la vibrazione. La figura 1.4 mostra la parte esterna di un orecchio. Si osserva che la pinna presenta dei bassorilievi visibili ed è direttamente connessa al condotto uditivo. Quest'ultimo può essere descritto in prima approssimazione come un tubo di larghezza costante caratterizzato da pareti ad alta impedenza acustica: mentre la pinna svolge la funzione di antenna acustica, esso si comporta come un risonatore monodimensionale. Le cavità della pinna, che possono essere molto differenti da persona a persona, hanno la capacità di amplificare alcune frequenze, mentre la loro particolare geometria introduce interferenze che ne attenuano altre.

L'intensità percepita alle due orecchie dipende da come è posizionata la testa (banalmente si ha che con l'orecchio diretto verso la sorgente si percepisce un'intensità maggiore che non con l'orecchio posto in altre posizioni). Si ha quindi un effetto di filtraggio da parte della pinna che può essere effettivamente pensata come un filtro la cui risposta in frequenza dipende in genere dalla distanza e dalla direzione del suono che giunge all'orecchio. Tale effetto può essere interpretato studiando le alterazioni dei raggi sonori introdotte dalla pinna, oppure mediante uno studio nel dominio della frequenza.

Supponiamo di avere una sorgente sonora ad altezza variabile di fronte a noi. In questo caso ci potrebbero non essere differenze di intensità percepita dalle due orecchie. Ciononostante, un ascoltatore dovrebbe essere comunque in grado di percepire la provenienza del suono, resa possibile dalla presenza della pinna, i cui effetti agiscono soprattutto nella direzione verticale.

In conclusione si può quindi dire che la pinna e il canale uditivo formano una sorta di risonatore acustico, le cui frequenze di risonanza dipendono dalla direzione e dalla distanza della sorgente.

1.3 Rendering Spaziale

1.3.1 Virtualizzazione del suono

Un sistema stereo a due canali, o anche a quattro canali, non riesce a riprodurre perfettamente il suono che uno o più sorgenti generano in un ambiente chiuso. Nel sistema stereo, per esempio, variando la potenza del suono da un altoparlante all'altro, si riesce a creare l'impressione che la sorgente sonora si muova continuamente tra le due casse. Tuttavia, il suono percepito con questa tecnica non uscirà mai dalla linea tra le due casse (per esempio non si riuscirà a metterlo dietro l'ascoltatore). Molti studi sono aperti su come, con l'uso delle sole cuffie, riprodurre quel senso di immersione che si prova quando si ascolta, per esempio, della musica in una stanza o si assiste ad un concerto. Lo scopo di questa tesi si focalizza su uno degli aspetti più complessi del rendering spaziale: la percezione del suono nel piano mediano. Prima di procedere con l'affrontare l'argomento nello specifico sorge spontaneamente una domanda. Si riesce davvero a riprodurre la stessa esatta pressione acustica prodotta dalle sorgenti che si presenterebbe alle orecchie dell'ascoltatore se questi fosse in una sala concerti o, in altre parole, si riesce a virtualizzare (rendere tridimensionale) il suono con l'utilizzo delle sole cuffie?

1.3.2 Audio binaurale

Il problema della realizzazione di una buona riproduzione sonora con effetto di immersione nello spazio acustico è stato affrontato nello studio di Schroeder, Gottlob e Siebrasse (1974) con costruzione di una testa di manichino dotato di orecchio esterno (pinna e canale uditivo) e di microfoni posizionati alla fine del canale uditivo [6]. Collegando l'uscita dei microfoni alle cuffie di un ascoltatore umano (uscita del microfono sinistro collegato alla cuffia sinistra e quello destro alla cuffia destra) si riuscivano ad approssimare i segnali nei canali uditivi dell'ascoltatore come se egli fosse effettivamente presente nel luogo dove è stata effettuata la registrazione.

Questa particolare tecnica dove due segnali diversi vengono inviati alle due orecchie prende il nome di audio binaurale.

Questo approccio però è valido soggettivamente, la registrazione effettuata per un soggetto non è detto sia compatibile con un secondo ascoltatore. Il motivo principale è che l'orecchio esterno, nonché la testa, variano in forma e caratteristiche da individuo a individuo, quindi l'effetto sarebbe migliore se si riuscisse ad effettuare la registrazione direttamente con la propria testa piuttosto che quella di un manichino.

Con rendering spaziale del suono si intende quindi l'abilità di trasmettere all'ascoltatore e riprodurre con sistema audio (ad esempio cuffie o surround) quella sensazione di immersione che l'ascoltatore avrebbe se fosse effettivamente in presenza della sorgente del suono. Si cerca quindi una virtualizzazione del suono che riesca a riprodurre le stesse condizioni (la stessa pressione acustica) e gli stessi segnali ai due timpani, in modo da riprodurre fedelmente un suono reale posizionato in un particolare punto nello spazio.

Ci sono diverse tecniche di rendering spaziale, la più nota è senza dubbio il sistema surround che, con l'utilizzo di diverse sorgenti opportunamente posizionate, riesce a ricreare il senso di immersione nel suono. Questo modo di procedere, se pur non perfetto, ha un grado di errore accettabile, tanto che è la tecnica di gran lunga più diffusa. I problemi più comuni sono la difficoltà nel posizionamento delle sorgenti e l'impossibilità di rendere le orecchie mutuamente esclusive (eliminando il cross-talk).

La seconda tecnica, che è quella che d'ora in poi andremo ad analizzare, prevede l'utilizzo delle cuffie auricolari. È una tecnica piuttosto complicata poiché la forma della stanza, il materiale delle pareti, il corpo dell'ascoltatore (testa, torso, pinna ecc.) modificano in modo significativo la percezione del suono, e tutte queste differenze devono essere opportunamente modellate nella creazione del segnale binaurale. Il modello che ne deriva prende il nome di *modello strutturale*.

Vediamo ora che nel caso di un sistema binaurale, basato sull'utilizzo di cuffie, si possono presentare due tipi di localizzazione:

- *Inside-the-head localization* (IHL, o lateralizzazione): questo termine è tipicamente utilizzato per indicare quando la sorgente sonora è percepita all'interno della testa, e principalmente lungo l'asse interaurale (vedi asse x in Figura 1.5); si può indurre questa percezione manipolando opportunamente ITD ed ILD in un sistema binaurale. Questo fenomeno illustra un esempio fondamentale di posizionamento virtuale della sorgente nello spazio; infatti se vengono trasmessi due segnali monoaurali perfettamente identici la sensazione sarà che la sorgente sia collocata al centro della testa. Variando ITD ed ILD si avrà la sensazione che la sorgente si sposti lungo l'asse interaurale da un'orecchio all'altro sempre all'interno della testa.
- *Esternalizzazione*: è un tipo di localizzazione molto difficile da riprodurre in un sistema binaurale basato su cuffie, dove risulta molto più semplice l'IHL; infatti non è del tutto chiaro quali siano i parametri supplementari più efficaci per rendere possibile questo tipo di localizzazione, tuttavia si è notato come l'aggiunta di un riverbero, naturale o artificiale, possa migliorarla sensibilmente.

1.4 Rendering spaziale tramite HRTF

Le HRTFs (Head-Related Transfer Functions) catturano le trasformazioni subite da un'onda sonora nel suo tragitto verso i timpani. Tali trasformazioni sono tipicamente dovute alla diffrazione e alle riflessioni su torso, pinna, spalle e testa dell'ascoltatore. Una conoscenza profonda dei parametri di processo permette il posizionamento virtuale delle sorgenti sonore nello spazio: coerentemente con la posizione relativa della testa dell'ascoltatore, il segnale viene filtrato con il corrispondente paio di HRTFs creando così i segnali binaurali da 'consegnare' agli auricolari. In questo modo, un ambiente tridimensionale in cui si percepisce un elevato senso di immersione può essere simulato.

Prima di procedere è bene definire la sorgente in coordinate sferiche. I parametri caratterizzanti sono: azimuth θ , elevazione ϕ e distanza r . La Figura 1.5 illustra questo sistema di riferimento nel piano verticale.

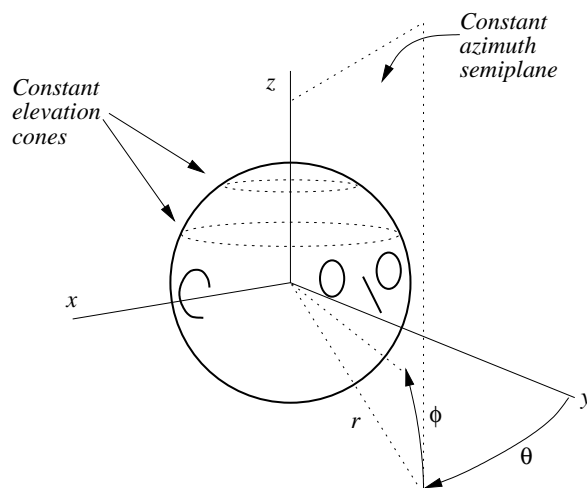


Figura 1.5: Coordinate sferiche usate nella definizione della HRTF: vertical-polar coordinate system

L'idea di base dietro al rendering 3D tramite le HRTF è quella di usare HRIR (Head-Related Impulse Response) e HRTF misurate. Dato un segnale anecoico e una posizione virtuale individuata da $(\theta$ e ϕ), un segnale destro ed uno sinistro vengono sintetizzati tramite le due seguenti operazioni:

1. ritardando il segnale anecoico di una quantità prefissata si riesce a introdurre l'ITD desiderato
2. facendo la convoluzione del segnale anecoico con le due HRIR (sinistra e destra).

Uno schema sintetico è dato in Figura 1.6.

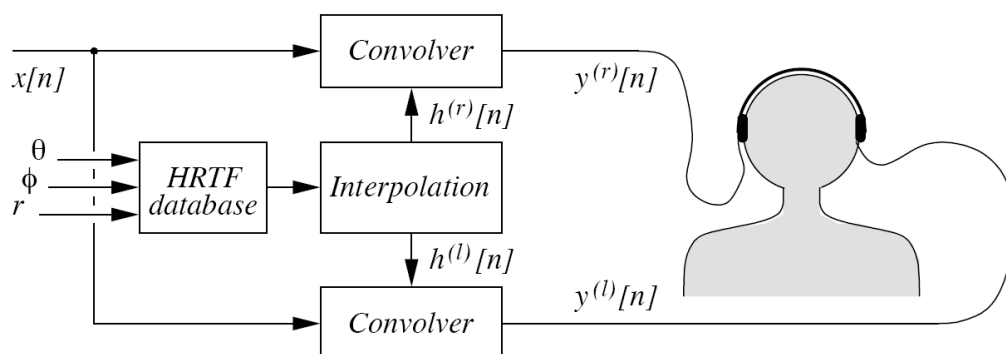


Figura 1.6: Schema a blocchi di un sistema di rendering spaziale del suono basato sulle HRTF

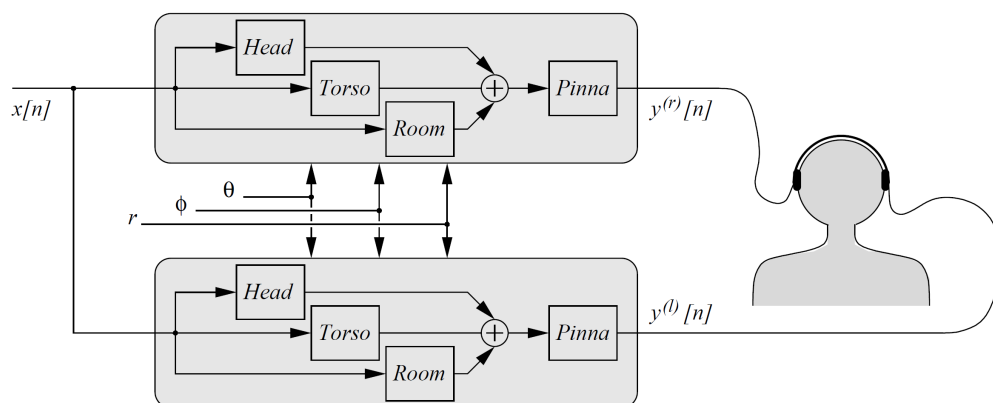


Figura 1.7: Schema a blocchi di un sistema di rendering spaziale del suono basato su modello strutturale

1.4.1 Modello strutturale

Al contrario del metodo di valutazione delle HRTF discusso poco sopra, il modello strutturale presentato in questo paragrafo si basa sull'astrazione degli effetti delle diverse parti del corpo che, una volta combinati, danno origine alla HRTF. Quest'ultima viene quindi modellata come una combinazione di blocchi (filtri), ognuno per ogni contributo di parti anatomiche diverse. I parametri di ciascun filtro possono essere legati a misure antropometriche (es. la distanza interaurale, o il diametro della cavum conchae), con il vantaggio che un generico modello strutturale per la HRTF può essere adattato ad un specifico ascoltatore, e può tenere in considerazione anche le variazioni di posizione dell'ascoltatore. Un altro vantaggio è che anche gli effetti dovuti alla stanza possono essere considerati nel modello, in particolare le prime riflessioni possono essere elaborate tramite il

modello della pinna. Risulta chiaro che separare i contributi delle varie strutture anatomiche in blocchi perfettamente indipendenti è un'approssimazione euristica che non tiene in considerazione le varie interazioni dovute alle onde riflesse da una struttura all'altra. Tuttavia, le ricerche mostrano che i modelli strutturali sono in grado di fornire una buona approssimazione di HRTF reali. La Figura 1.7 mostra un generico modello strutturale.

Da studi condotti [2], si può dedurre, in prima approssimazione, che gli effetti più rilevanti della pinna (ultimo blocco della Figura 1.8) che devono essere presi in considerazione sono quelli dovuti alle riflessioni. Questo significa che la pinna è modellata come una sequenza di filtri, in cui ciascun filtro determina una traccia nello spettro. Per realizzare un modello per la pinna, il tutto si riduce a stimare i ritardi dovuti alle riflessioni e alla loro dipendenza da ϕ e θ , tramite misurazioni di HRIRs/HRTFs, o tramite simulazioni numeriche.

Per l'approfondimento sugli altri blocchi si rimanda a [2].

Capitolo 2

Modello strutturale di HRTF

Nella modellazione strutturale l'HRTF viene scomposta in parti tra loro indipendenti: testa, busto, spalle e pinna.

Possiamo quindi considerare la pinna come un oggetto indipendente dal resto e modellarla separatamente. Indichiamo quindi la risposta della pinna con PRTF (Pinna Related Transfer Function). Una volta fissata la direzione della sorgente relativa all'ascoltatore, le differenze più marcate tra le HRTF sono dovute alle caratteristiche della pinna, quest'ultima diversa da persona a persona. La pinna ha un ruolo fondamentale nel determinare il contenuto in frequenza delle HRTF grazie a due fenomeni acustici principali già accennati:

1. riflessioni lungo i bordi della pinna (Figura 2.1 (a)). Secondo Batteau [6], le onde sonore vengono tipicamente riflesse dall'orecchio esterno finché la loro lunghezza d'onda è abbastanza piccola rispetto alle dimensioni della pinna stessa, e l'interferenza causata dalle onde provenienti direttamente dalla sorgente e quelle riflesse fa sì che, nella parte delle alte frequenze, compaiano dei notch acuti nello spettro del segnale ricevuto con periodicità inversamente proporzionale al tempo di ritardo di ogni riflessione.
2. risonanze nelle condotte della pinna (Figura 2.1 (b)). Poiché la conca si comporta come un risonatore, alcune bande di frequenze, sia delle onde dirette che di quelle riflesse, vengono amplificate in modo significativo in base all'elevazione della sorgente.

Di conseguenza, la parte della HRTF dovuta al contributo della pinna presenta nella sua ampiezza una sequenza di *peak* e *notch*.

2.1 Modello strutturale di PRTF

Partendo da studi precedenti, in [3] gli autori hanno realizzato un algoritmo che permette di scomporre la PRTF in due componenti: componente risonante e componente riflettente.

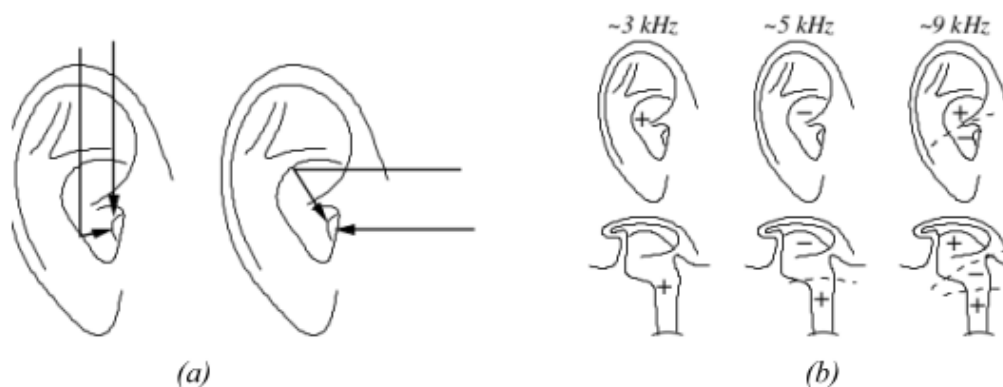


Figura 2.1: Effetti della pinna: (a) riflessioni dipendenti dalla direzione, (b) risonanze

L'algoritmo produce una decomposizione realistica, almeno nell'intervallo di frequenza 3-18 kHz. Per mezzo di un sistema di identificazione basato su un modello ARMA del sesto ordine, gli autori sono riusciti a ricavare le frequenze centrali delle risonanze evidenziando due aree con forte guadagno (due risonanze). La prima, intorno ai 4 kHz, che si denota rimanere costante per soggetti diversi lungo tutte le elevazioni. La seconda, intorno ai 7 kHz dovuta ad una seconda risonanza che interferisce all'interno del range di frequenze in cui opera l'algoritmo. Le risonanze a più alte frequenze sono irrilevanti, poiché al limite dell'intervallo di frequenze udibili.

Per quanto riguarda la componente riflettente si osserva che, rispetto alla componente risonante, le frequenze dei notch sono dipendenti dall'elevazione e dalla forma della pinna. Per sorgenti sopra la testa dell'ascoltatore non si sono notati notch nella PRTF, mentre, al diminuire dell'elevazione, vi è la comparsa di notch con profondità e larghezza di banda dipendente proprio da quest'ultima.

Partendo da queste osservazioni, gli autori hanno realizzato un modello strutturale di pinna costituito da due blocchi principali:

- Blocco risonante
- Blocco riflettente

Il blocco risonante è costituito da due filtri peak del secondo ordine in parallelo con all'ingresso lo stesso segnale. L'uscita dei due filtri viene sommata per poi essere inviata all'ingresso del blocco riflettente, costituito dalla cascata di 3 filtri notch del secondo ordine. In Figura 2.2 è mostrato lo schema del modello strutturale.

La regione di validità del modello è mostrata in Figura 2.3. Essa si estende nel piano mediano da -45 a $+45$ gradi e, previa correzione dell'ITD/ILD, per

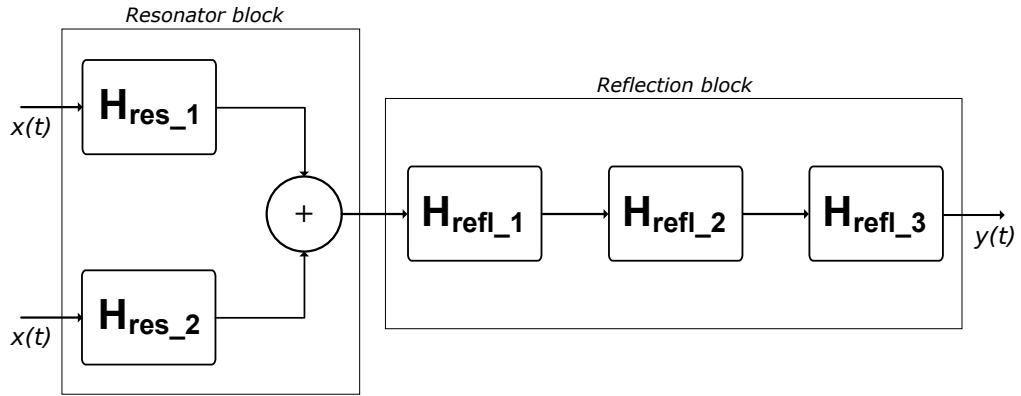


Figura 2.2: Modello per la pinna

un azimut che va da -30 a $+30$ gradi. Si è infatti notato che in questo range le frequenze dei notch rimangono stabili rispetto all'elevazione della sorgente.

2.1.1 Filtri

Blocco risonante

Come già detto, il blocco risonante è realizzato mediante due filtri del secondo ordine. Per modellare la prima risonanza si utilizza il filtro:

$$H_{res}(z) = \frac{1 + \frac{H_0}{2} + d(1-k)z^{-1} + (-k - \frac{H_0}{2}(1+k))z^{-2}}{1 + d(1-k)z^{-1} - kz^{-2}} \quad (2.1)$$

$$d = -\cos\left(2\pi\frac{f_C}{f_s}\right) \quad (2.2)$$

$$V_0 = 10^{\frac{G}{20}} \quad (2.3)$$

$$H_0 = V_0 - 1 \quad (2.4)$$

$$k = \frac{\tan\left(\pi\frac{f_B}{f_d}\right) - 1}{\tan\left(\pi\frac{f_B}{f_d}\right) + 1} \quad (2.5)$$

La scelta di tale filtro deriva dal fatto che esso non modifica l'intervallo al di fuori della frequenza di banda. Tuttavia, per fare in modo che la prima risonanza non interferisse con la seconda, è stato necessario aggiungere un filtro passa basso. Per la seconda risonanza è stato utilizzato un filtro IIR del tipo:

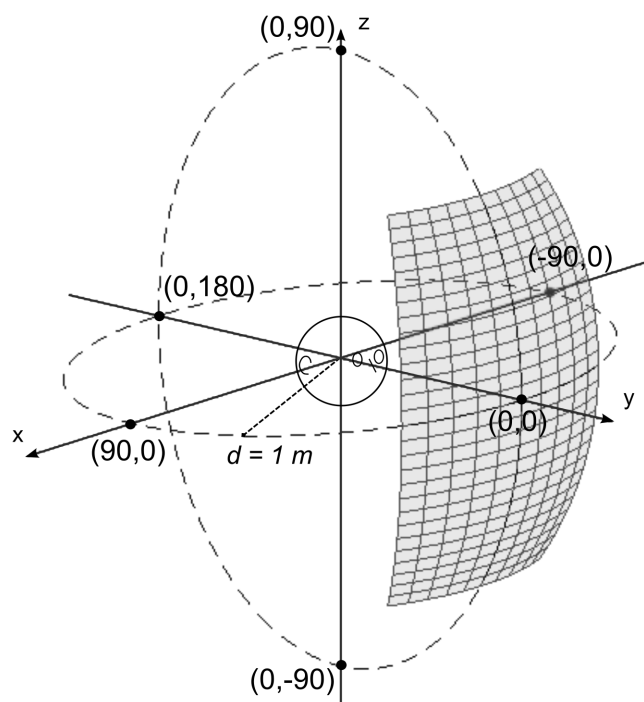


Figura 2.3: Regione di validità del modello

$$H_{res}(z) = \frac{V_0(1-h)(1-z^{-2})}{1+2dhz^{-1}-(2h-1)z^{-2}} \quad (2.6)$$

$$d = -\cos\left(2\pi\frac{f_C}{f_s}\right) \quad (2.7)$$

$$V_0 = 10^{\frac{G}{20}} \quad (2.8)$$

$$h = \frac{1}{1 + \tan\left(\pi\frac{f_B}{f_s}\right)} \quad (2.9)$$

Alla seconda risonanza è quindi stato applicato un filtro passa alto in modo da non modificare la risposta in frequenza tra i due peak.

I parametri di rilievo in questo caso sono f_C , frequenza centrale, G , guadagno, f_B , larghezza di banda del peak e f_s , frequenza di campionamento. L'approssimazione è buona per frequenze fino a 14 kHz, diventando scarsa per frequenze superiori, ma, dal momento che le frequenze interessanti ai fini psicoacustici sono molto basse, tale discrepanza tra le PRTF non influisce.



Figura 2.4: Tracciamento con-
torni della pinna

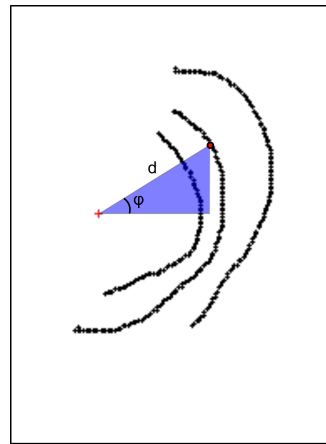


Figura 2.5: Punti di riflessione
in coordinate polari

Blocco riflettente

Per quanto riguarda la parte riflettente, sono state utilizzate, per la creazione dei notch, funzioni disponibili in Matlab (`fdesign.parameq`), che permettono di creare filtri equalizzatori parametrici con la possibilità di settare frequenza centrale, profondità e larghezza di banda del filtro. La funzione utilizzata segue le seguenti specifiche come parametri d'ingresso:

- N : ordine del filtro
- F_0 : frequenza centrale
- BW : larghezza di banda
- G_{ref} : profondità di riferimento (dB)
- G_0 : profondità alla frequenza centrale (dB)
- GBW : profondità alla quale la larghezza di banda è misurata (dB)

Per i parametri N , G_{ref} , GBW i valori scelti sono, rispettivamente, 2, 0 dB, -3 dB. I valori di frequenza centrale dipendono dalla pinna del soggetto, mentre profondità e larghezza di banda derivano da uno studio precedentemente condotto in [4].

Dall'unione del blocco risonante e dal blocco riflettente si ottiene il modello di PRTF.

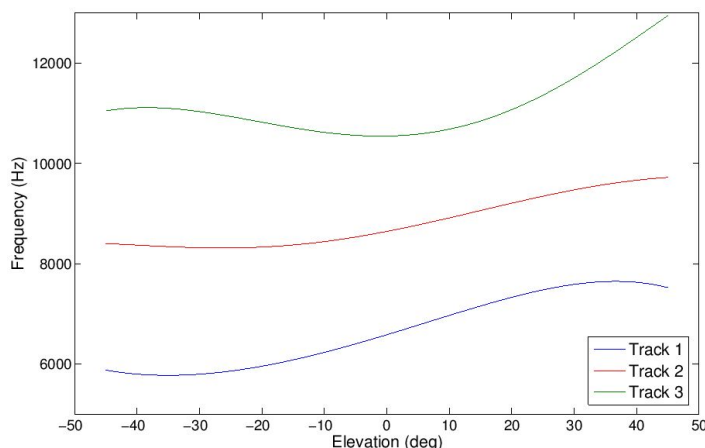


Figura 2.6: Estrazione delle frequenze di notch da una foto della pinna: Estrazione e approssimazione traccia notch

2.1.2 Estrazione parametri dalla pinna

In [1] gli autori hanno realizzato una procedura che a partire dall'immagine della pinna produce le frequenze centrali dei notch. Per far questo vengono tracciati dall'immagine della pinna i tre contorni più rilevanti, come mostrato in Figura 2.4. Questi vengono poi salvati come un sequenza di pixel e, attraverso calcoli trigonometrici, trasformati in una coppia di coordinate polari (d, ϕ) rispetto al canale uditivo, come è mostrato in Figura 2.5. Quindi mediante la formula:

$$f_0(\phi) = \frac{c}{2d(\phi)} \quad (2.10)$$

con c velocità del suono, ϕ l'elevazione e d la distanza del punto di riflessione, si ricavano le frequenze centrali dei notch ad ogni elevazione e per ognuna delle tre tracce (vedi Figura 2.6).

2.1.3 Scelta dei parametri

Come già citato, uno studio precedente a questo è stato sviluppato per produrre il miglior set di parametri utili all'esperimento. La sperimentazione eseguita da [4] per estrarre questi parametri prevedeva un set di 12 condizioni dove variavano G , guadagno o profondità del notch, BW , larghezza di banda del notch calcolata a -3 dB SPL, ed infine il numero dei notch. Al termine della sperimentazione è risultato che la condizione espressa in Tabella 2.1 era quella che permetteva una migliore percezione verticale del suono. In questo test pilota, per quanto riguarda i parametri G e BW , sono quindi stati utilizzati i valori riportati in Tabella 2.1.

Depth (dB)	Bandwidth	Nr.Notch
-30	2000	3

Tabella 2.1: Condizioni sperimentali

L'unico parametro associato alle misurazioni antropometriche del soggetto è f_c , la frequenza centrale del notch. Tale valore viene ricavato per mezzo di una GUI sviluppata in Matlab con cui si effettua il tracciamento dei contorni sulla pinna e, da tali contorni, si ricavano le frequenze centrali dei notch associati ai punti di riflessione.

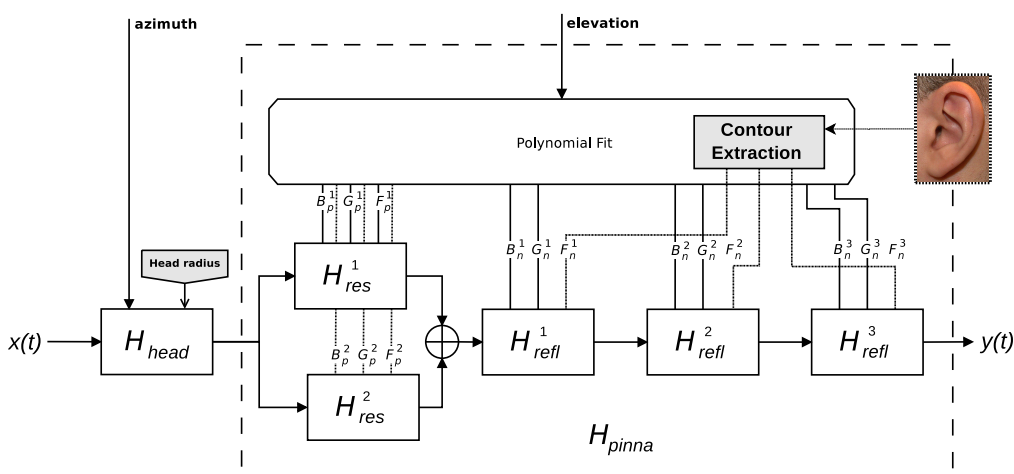


Figura 2.7: Modello completo di HRTF

2.2 Modello completo di HRTF

Mediante il modello di PRTF precedentemente studiato si è realizzato il modello completo di HRTF mostrato in Figura 2.7 che comprende anche un blocco relativo al modello di testa sferica che aggiunge ITD e ILD. Come si vede, l'unico parametro indipendente usato dal blocco relativo alla pinna è l'elevazione, grazie al quale è possibile calcolare le funzioni polinomiali associate alla frequenza centrale, alla larghezza di banda e profondità dei notch. Le frequenze centrali dei notch sono legate alla forma della pinna, quindi la funzione polinomiale ad esse associata può essere calcolata offline. Come già visto, la parte risonante è realizzata con due filtri peak del secondo ordine in parallelo, mentre la parte riflettente è costituita dalla cascata di tre filtri notch del secondo ordine. In figura 2.8 è mostrato un confronto tra PRTF originali e ricostruite dal modello per il soggetto 048 di un database di dominio pubblico contenente HRIR misurate

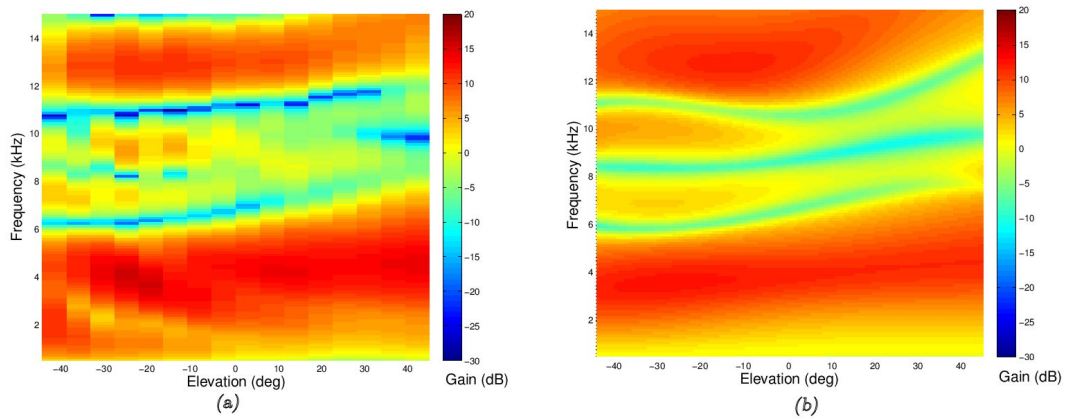


Figura 2.8: PRTF a confronto: (a) originale, (b) ricostruita dal modello

per 1250 direzioni spaziali (database CIPIC). La focalizzazione di questo esperimento sarà solo sul piano mediano, quindi ad azimuth 0, dove il contributo della testa (H_{head}) è trascurabile. Il modello dai noi utilizzato quindi non si curerà del blocco relativo alla testa ma si concentrerà piuttosto sulla sola PRTF.

Capitolo 3

Protocollo sperimentale

3.1 Introduzione

Le sperimentazioni di seguito documentate sono state eseguite con l'intenzione di verificare il protocollo sperimentale adottato, per poterlo applicare ad una sperimentazione più ampia ed approfondita. Quello realizzato in questa tesi è dunque un test pilota e come tale mira ad ottenere dei risultati la cui analisi possa permettere osservazioni ed eventuali modifiche al protocollo.

3.2 Apparato

3.2.1 Matlab

Matlab è il software utilizzato per la realizzazione dell'intero esperimento. Si tratta di un linguaggio ad alto livello utilizzato in diversi ambiti, tra cui l'elaborazione dei segnali. Matlab consente di analizzare dati, sviluppare algoritmi e creare modelli integrando calcolo, visualizzazione e programmazione. Inoltre, mette a disposizione un ambiente grafico, chiamato GUIDE (GUI Design Environment) utile per la realizzazione di semplici ed immediate interfacce grafiche. Di grande utilità risultano i toolbox integrati in Matlab: in particolare il toolbox statistico, quello relativo alla manipolazione delle immagini e dell'audio. Tale linguaggio è risultato altresì adatto per l'analisi e la rappresentazione dei dati. Tutti i grafici riportati in questa tesi sono stati ottenuti mediante questo software.

3.2.2 Cuffie

Le cuffie adottate sono le Sennheiser HDA 200. Si tratta di cuffie circumaurali, di tipo dinamico e chiuse, con i cuscinetti che circondano completamente l'orecchio. Sono in grado di isolare quasi completamente l'orecchio da rumori esterni e hanno una risposta in frequenza che va da 20 a 20000 Hz, con un'impedenza di 40 Ohm.

Un problema affrontato deriva dal fatto che la riproduzione tramite cuffie introduce distorsioni di varia natura. In particolare le Sennheiser HDA 200 presentano un comportamento in frequenza assimilabile con un filtro passa banda che si discosta da una risposta a guadagno costante ideale, con frequenze di taglio attorno ai 150 Hz e 9 kHz. Inoltre, associata ai riposizionamenti della cuffia sulla testa, si rileva anche una variabilità della HpTF (Headphones Transfer Function). Essa è più pronunciata per le frequenze superiori a 7 kHz. Ovviamente tutto questo si traduce in un'errata riproduzione del segnale binaurale. Si è reso quindi necessario un intervento di compensazione delle cuffie. Da sottolineare che la risposta in frequenza delle cuffie Sennheiser HDA 200 è stata calcolata su di un manichino KEMAR¹ senza pinna. In realtà un'esatta riproduzione binaurale è possibile solo per il binomio soggetto-cuffia su cui sono state acquisite le HpIR. Di conseguenza la compensazione delle cuffie valutata sul KEMAR è solo una compensazione media su frequenze sotto i 10 kHz.

3.2.3 Scheda audio

La scheda audio esterna utilizzata è una Roland Edirol AudioCapture UA-101. Si tratta di un'interfaccia audio Hi-Speed con connessione USB 2.0. Essa è in grado di gestire contemporaneamente 10 ingressi e 10 uscite alla qualità di 24-bit 96kHz. Grazie alla tecnologia Hi-Speed si garantiscono livelli di latenza estremamente bassi. La scheda è in grado di lavorare a 44.1, 48, 88.2, 96, 192 kHz. Tutti i canali sono bilanciati sia in ingresso che in uscita.

3.2.4 Camera silente

L'esperimento si è svolto all'interno di una cabina insonorizzata. Si tratta della Sound Station Pro 45 by Puma s.r.l di Figura 3.1. In figura 3.2 si possono vedere le caratteristiche di abbattimento sonoro e di riverbero. All'interno della cabina è stato posizionato un computer collegato alla scheda audio collegata a sua volta alle cuffie.

3.3 Soggetti

3.3.1 Presentazione dei soggetti

I soggetti testati, tutti di sesso maschile, non sono stati scelti secondo criteri particolari, in modo da validare i risultati in caso generico. L'età dei soggetti varia tra i 21 e i 41 anni e tutti, prima di iniziare l'esperimento, sono stati sottoposti ad una breve sessione di training. In tabella 3.1 vi è la lista dei soggetti che hanno partecipato all'esperimento.

¹Manichino con caratteristiche antropometriche medie utilizzato per la registrazione di segnali binaurali



Figura 3.1: Cabina silente: Sound Station Pro 45

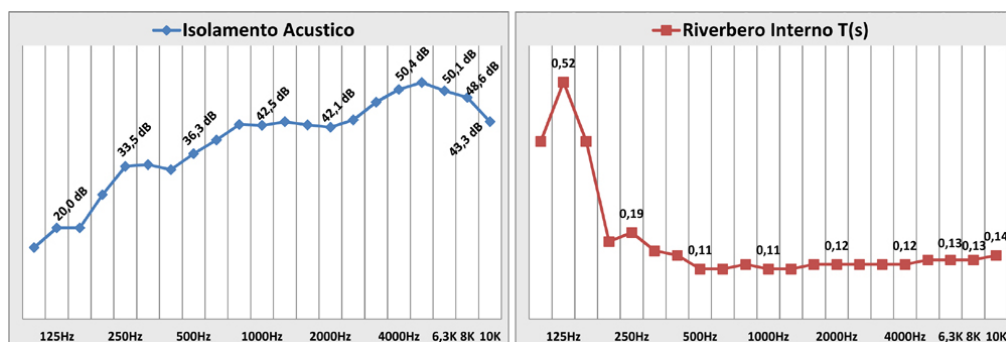


Figura 3.2: Caratteristiche della Sound Station Pro 45

ID	Sesso	Età
20	M	33
21	M	28
22	M	29
30	M	41
35	M	23
36	M	23
37	M	23
40	M	23
41	M	23
42	M	23
43	M	23
44	M	21

Tabella 3.1: Soggetti

La sessione di training è stata realizzata mediante un piccolo contenitore pieno di piccoli oggetti, messo in vibrazione a varie altezze davanti e dietro il soggetto che per l'occasione teneva gli occhi chiusi. In questa fase di training si sono evidenziate delle difficoltà nel riconoscere l'elevazione del suono prodotto dal contenitore per i soggetti 40 e 42.

3.3.2 Caratterizzazione dei soggetti

Un elemento rilevante al fine di comprendere i risultati risiede nella suddivisione di tutti i soggetti in due distinte categorie. La prima categoria comprende i soggetti 20, 21, 22, 30 e 35, che avevano precedentemente all'esperimento acquisito una esperienza di suono spazializzato. La seconda categoria comprende tutti i soggetti rimanenti, considerati inesperti, che invece non hanno mai avuto esperienze nel campo.

3.4 Stimoli

Ogni stimolo è costituito da un treno di 3 impulsi di rumore gaussiano, ognuno dei quali della durata di 300 ms, separato dal successivo da una pausa di 250 ms. Tali stimoli sono presentati ad un livello di 60 dbA misurati a livello della pinna sulle cuffe. Questi rumori di breve durata permettono di prevenire, durante l'esperimento, movimenti della testa, quindi sono appropriati considerando esperimenti di tipo statico.

Gli stimoli riprodotti sono suddivisi in tre condizioni:

- A. Stimoli processati a partire dalla pinna del soggetto
- B. Stimoli processati a partire dalla pinna del manichino KEMAR

C. Stimoli direttamente registrati dal manichino KEMAR

e ogni condizione viene presentata in due distinti blocchi seguendo una sequenza ben precisa, decisa casualmente prima dell'esperimento. Le sequenze sono pensate in modo che le due ripetizioni di ogni blocco non fossero contigue. A tal proposito sono state adottate le 3 sequenze di Tabella 3.2.

Nr. Sequenza	Sequenza
1	A B C B C A
2	B C A C A B
3	C A B A B C

Tabella 3.2: Sequenze di riproduzione

Quindi ogni sequenza è composta da sei blocchi in ognuno dei quali una singola condizione viene presentata.

Le elevazioni testate sono 9 e vengono mostrate in tabella 3.3, esse si trovano tutte sul piano mediano frontale, espresse utilizzando il sistema di coordinate verticale polare.

Elevazioni
-45 -33.75 -22.5 -11.25 0 11.25 22.5 33.75 45

Tabella 3.3: Elevazioni considerate

Ogni stimolo associato ad una posizione viene ripetuto tre volte per blocco, quindi il numero totale di stimoli è:

$$9 \text{ elevazioni} * 3 \text{ ripetizioni} * 6 \text{ blocchi} = 162 \text{ stimoli},$$

presentati in ordine casuale all'interno di ogni blocco. Lo stimolo finale, presentato al soggetto, è ottenuto tramite convoluzione con le nostre PRTF sintetiche e il filtraggio per la compensazione delle cuffie utilizzate per l'esperimento.

3.5 Procedura

L'esperimento ha inizio digitando *StartMain* dal prompt di Matlab. Tale comando lancia il file `startMain.m` che a sua volta richiama `subject.m` che visualizza l'interfaccia GUI di figura 3.3. Essa è utilizzata per la gestione del database dei soggetti e ha una serie di bottoni per le varie funzioni dell'esperimento.

3.5.1 Inserimento dati

Banalmente per ogni soggetto che effettua l'esperimento vengono riempiti i campi *Last Name*, *First Name*, *Age* e *Gender*. Per inserire il soggetto nel database

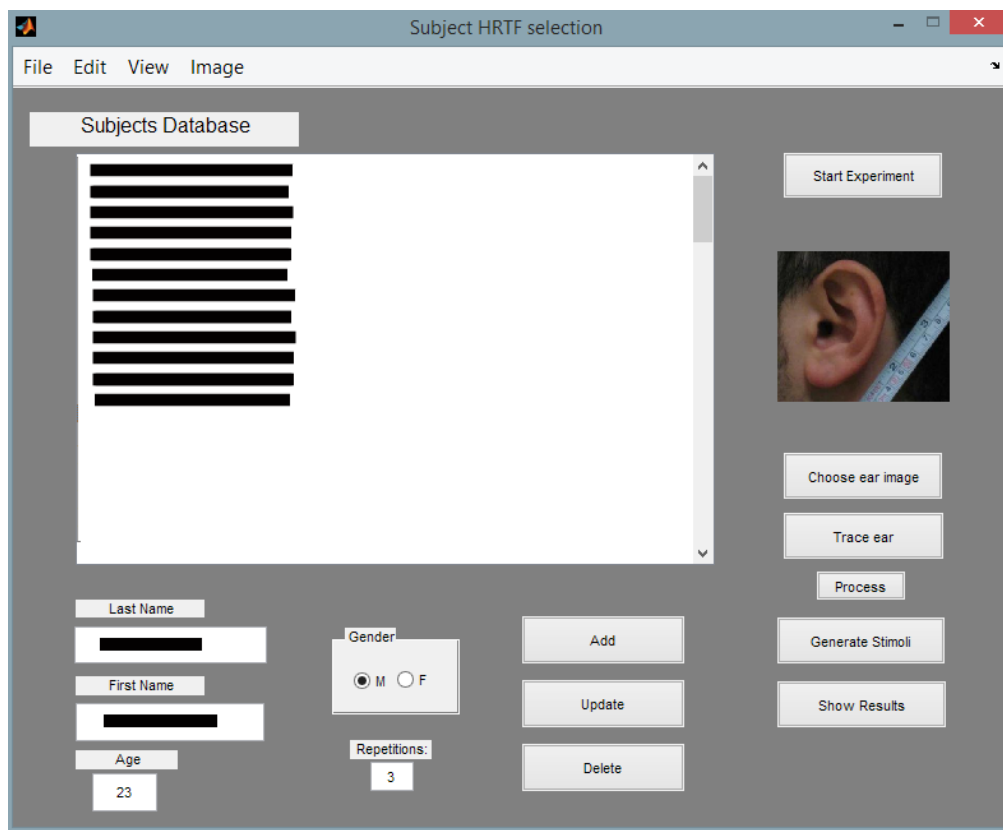


Figura 3.3: GUI per l'inserimento dati e tracciamento contorni

bisogna premere il pulsante *Add*, per modificare un soggetto bisogna premere il pulsante *Update* dopo averlo selezionato dal database e modificato nei campi errati. Per cancellare un soggetto bisogna premere il pulsante *Delete*. Ogni modifica viene salvata alla chiusura della GUI.

3.5.2 Acquisizione immagine e tracciamento

I soggetti che partecipano all'esperimento vengono fatti sedere all'interno di una stanza dove è stato creato un setup al fine di acquisire l'immagine della pinna. Il setup è mostrato in Figura 3.4. Al soggetto viene chiesto di centrare la propria testa su uno specchio posto davanti a sé, sul quale è stata posta una linea verticale dritta. Al soggetto viene quindi chiesto alternativamente di chiudere un occhio e muovere la testa in modo da sovrapporre l'occhio chiuso alla linea dritta. In questo modo si evitano movimenti della testa a destra e sinistra. I restanti gradi di libertà vengono fissati per mezzo di una livella posta sulla fotocamera e di un lavoro di post-produzione sull'immagine acquisita. Quindi per mezzo di una fotocamera posta su di un treppiede acquisiamo l'immagine della pinna. La

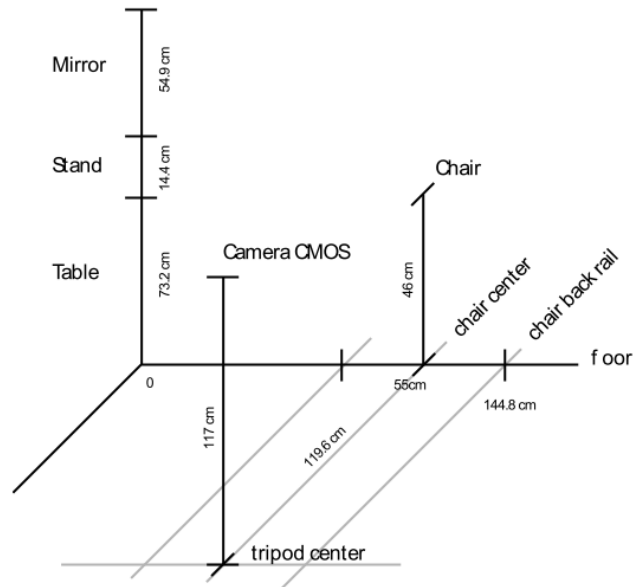


Figura 3.4: Setup per l'acquisizione dell'immagine della pinna

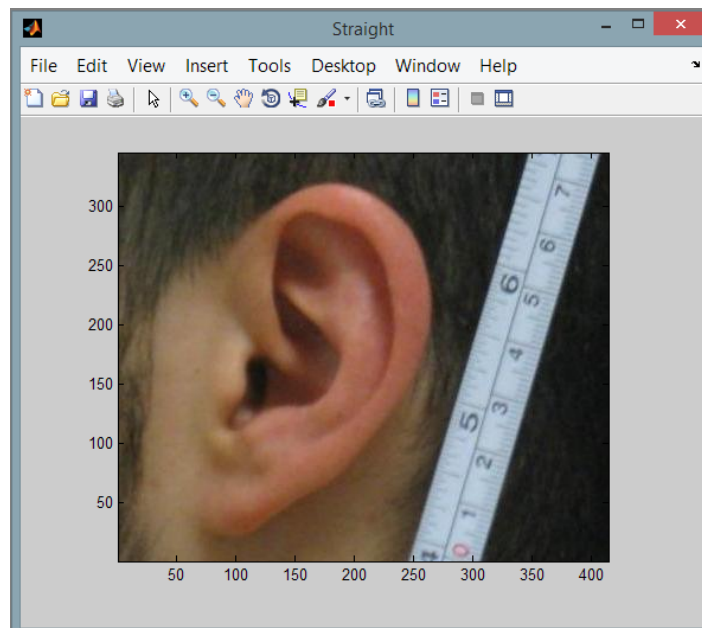


Figura 3.5: Immagine di pinna fotografata

fotocamera utilizzata è una Canon Digital IXUS 85 IS con lunghezza focale 6.2–

18.6 mm (equivalente a 35 – 105 mm in formato 35 mm). Il sensore CMOS della fotocamera è perpendicolare alla pinna del soggetto. Come è possibile vedere in un'immagine di pinna fotografata (Figura 3.5), è stato posto un righello vicino all'orecchio del soggetto in modo da aver un fattore di conversione pixel-metri.

L'immagine così acquisita viene manipolata in Matlab per mezzo di una GUI richiamata dalla GUI principale di figura 3.3. Per mezzo di questa seconda GUI è possibile scalare l'immagine, ruotarla, ed applicare il fattore di conversione pixel-metri. Tale GUI viene richiamata dalle varie funzioni inserite nel menù *Image*. A questo punto l'immagine della pinna è pronta per poterne tracciare i contorni per mezzo di una terza GUI realizzata sempre in Matlab. A tal fine ci siamo serviti di una tavoletta grafica Wacom larga abbastanza da facilitare il tracciamento della pinna. Vengono tracciati i 3 contorni su cui avvengono le riflessioni dell'onda sonora come indicato in Figura 3.6. Si parte con il tracciamento più esterno effettuato sul bordo dell'elice, per poi continuare con il tracciamento dell'antielice e della conca. Infine indichiamo la posizione del canale uditivo (ear canal).

Premendo successivamente il pulsante *Process* vengono calcolate le distanze tra il canale uditivo e i contorni di riflessione. Tali dati vengono salvati all'interno di un file .mat associato all'immagine della pinna. Tale file contiene:

- Il vettore F_0 di dimensioni 3x9, con le frequenze estratte dai 3 contorni, relativamente alle 9 elevazioni testate. Ogni riga del vettore F_0 è associata ad un contorno: la prima riga rappresenta le frequenze relative al contorno più esterno, la seconda rappresenta le frequenze relative al contorno

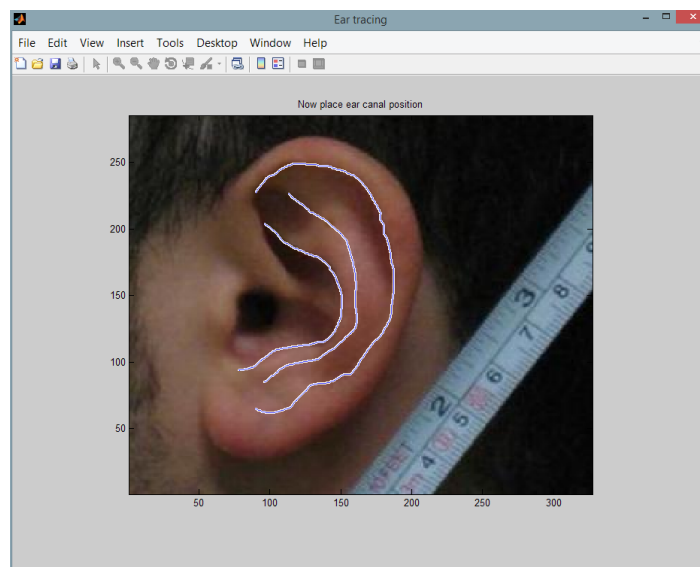


Figura 3.6: Contorni tracciati su pinna

centrale, mentre l'ultima riga è associata al contorno più interno. Ogni colonna è invece associata ad una delle 9 elevazioni testate.

- I vettori C_x e C_y contenenti i punti tracciati sui 3 contorni.
- *Earcanal* contenente la posizione del canale uditivo.
- *Dist*, vettore con le distanze tra l'ear canal e i punti sui contorni.
- *PixelToMeterFactor* contenente il fattore di conversione pixel-metri.

L'immagine della pinna in formato .jpg ed il relativo file .mat, contenente i dati di cui sopra, vengono salvate all'interno della cartella img.

3.5.3 Test di ascolto

Una volta terminata la procedura di acquisizione parametri con l'immagine della pinna siamo pronti a generare gli stimoli attraverso il pulsante *Generate Stimoli* che crea i 9 stimoli (vedi Figura 3.7) del soggetto sulla base delle informazioni acquisite tramite la procedura del paragrafo precedente. I restanti 18 stimoli, rispettivamente della KEMAR processata e della KEMAR registrata sono stati creati precedentemente l'inizio degli esperimenti e sono stati utilizzati invariati per ognuno degli esperimenti con i vari soggetti.

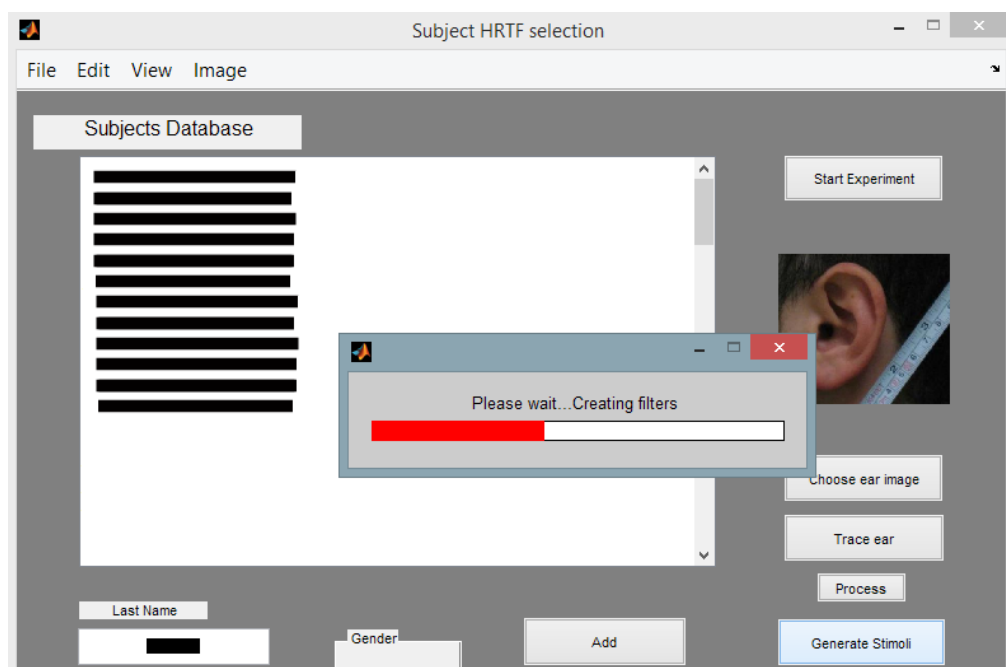


Figura 3.7: Creazione stimoli personali

A questo punto si può cominciare l'esperimento vero e proprio premendo il pulsante *Start Experiment*.

La GUI che si apre è riportata in Figura 3.8. L'elemento più evidente è una grande testa nera con due ventagli, uno davanti e uno dietro, di colore verde. Questi due ventagli rappresentano l'area attiva dove poter cliccare per inserire le informazioni relative all'altezza percepita dei vari stimoli ascoltati. Si hanno anche due campi informativi: *Test* che indica quanti stimoli si sono già ascoltati sul totale di 162; *Subject*, che visualizza le informazioni del soggetto testato.

L'esperimento ha inizio quando si preme sul pulsante *Begin!* (che viene disattivato dopo essere stato premuto). Inizia così la riproduzione degli stimoli secondo l'ordine definito nel paragrafo 3.4. Dopo ogni stimolo viene chiesto di indicarne l'altezza in una delle due aree verdi, a seconda che il suono sia stato percepito davanti o dietro. Tutti gli stimoli sono stati pensati per essere sentiti unicamente davanti, è tuttavia interessante vedere se essi vengano percepiti davanti o dietro. Nel capitolo dei risultati tale aspetto verrà approfondito.

Nel caso si cliccasse fuori dall'area verde verrà visualizzato un messaggio di errore e verrà richiesto di indicare l'elevazione del suono percepito nelle aree verdi. Una volta scelta un'elevazione plausibile vi è un processo di conversione della coordinata cartesiana in coordinata polare (tra -45° e $+45^\circ$ di elevazione), al fine di poter realizzare dei grafici di facile comprensione. Nel capitolo successivo verranno analizzati alcuni dei grafici ottenuti.

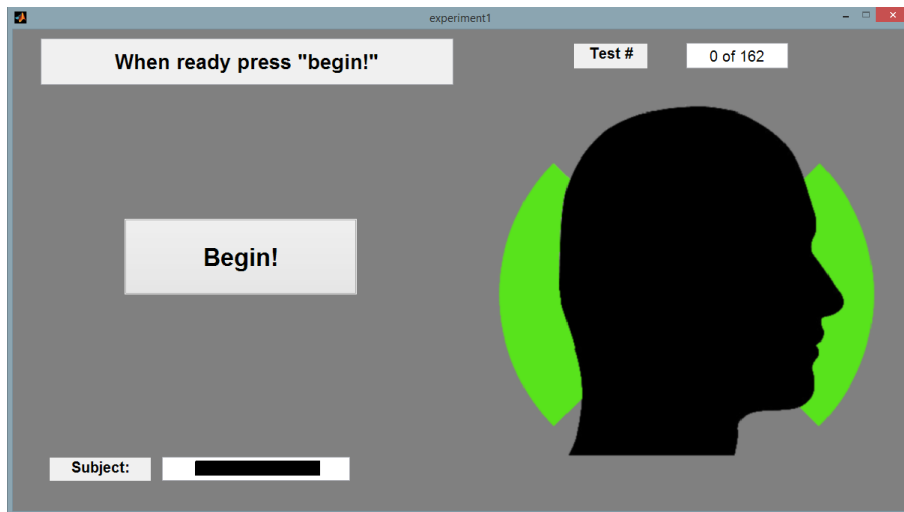


Figura 3.8: GUI esperimento

Dopo ogni blocco di 27 stimoli vi è una pausa di minimo 30 secondi, sarà infatti l'utente a dare l'ok per il proseguimento dell'ascolto. Al termine tutti i risultati vengono salvati. L'intero test di localizzazione dura, considerate tutte le pause, tra i 15 e i 25 minuti.

Capitolo 4

Risultati

4.1 Descrizione dell'output

Attraverso gli script `plot_conditions_new.m` e `plot_all_new.m` vengono generati due tipi diversi di output. Il primo genera una sequenza di 3 grafici, uno per ogni categoria di stimoli (cioè un grafico per gli stimoli personali, un grafico per gli stimoli processati dalla KEMAR ed un grafico per gli stimoli registrati sulla KEMAR), mentre il secondo genera una tabella con una serie di dati numerici.

4.1.1 Output grafico

In Figura 4.1 vi è un esempio di output grafico. La linea blu corrisponde alla corretta posizione delle elevazioni mentre la linea nera è una regressione lineare dei punti (di colore rosso) corrispondenti alle varie elevazioni assegnate agli stimoli percepiti. Come si può vedere i punti, che sono distribuiti sulle nove linee verticali corrispondenti alle nove possibili elevazioni, sono 6, ogni stimolo infatti viene ripetuto tre volte per blocco (e ogni blocco viene ripetuto due volte per condizione). Essi sono distribuiti sull'asse delle ordinate a seconda dell'elevazione percepita.

Più i punti sono sparsi, più è confusa la percezione di quella relativa altezza, quando invece i punti si concentrano su un determinato punto vuol dire che lo stimolo viene sempre percepito alla stessa altezza, a prescindere che quell'altezza sia corretta o meno.

Come si può vedere non vi è alcun riferimento alla percezione dello stimolo davanti o dietro in questi grafici.

Più la linea nera è vicina alla linea blu (per posizione e pendenza), migliori sono i risultati, senza dimenticare che è la posizione dei punti rossi a indicare la reale bontà del grafico: più sono raccolti nelle vicinanze della linea blu, più la percezione del suono registrata nell'esperimento corrisponde a quella reale dello stimolo.

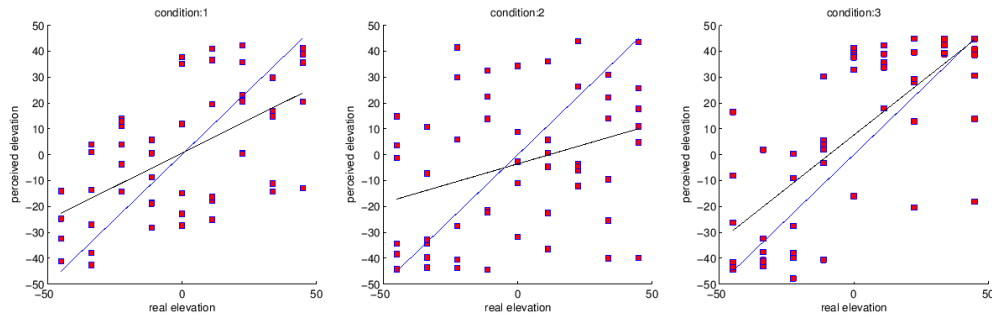


Figura 4.1: Grafici del soggetto ID 30

4.1.2 Output parametrico

L'output parametrico consiste in una struct di tre tabelle, una per ogni condizione, ciascuna delle quali contenente 5 parametri fondamentali:

- **Mean angle error**, ovvero la media delle differenze tra angolo reale e angolo percepito su tutti gli stimoli di quella condizione. Più questo valore è basso, migliore è da considerarsi il risultato.
- **Slope**, che visualizza la pendenza della retta di regressione (retta nera) dell'output grafico.
- **Goodness of fit**, che indica quanto bene i punti siano distribuiti lungo la retta di regressione. Più questo indice si avvicina a 1 migliore è la regressione.
- **Up/down confusion**, che rappresenta la percentuale di stimoli che vengono percepiti sopra il piano orizzontale invece che sotto o viceversa.
- **Front/back confusion**, che rappresenta la percentuale di stimoli che vengono percepiti dietro piuttosto che davanti.

L'analisi di questi parametri ci permetterà di concludere quali sono i benefici apportati dal nostro modello (condizioni 1 e 2) rispetto alle HRTF registrate sul KEMAR (condizione 3) nonché della parametrizzazione sui contorni personali (condizione 1) rispetto a contorni generici (condizione 2), e le eventuali migliorie da apportare per condurre la sperimentazione vera e propria. Essi perciò decideranno l'esito del test pilota realizzato in questa tesi.

4.2 Analisi dettagliata dei risultati grafici

In Appendice A sono riportati tutti i grafici dei soggetti che hanno partecipato all'esperimento. Prendendo come riferimento i grafici del soggetto 21 osserviamo

come, nel primo grafico (relativo alla Condizione 1), egli abbia una distribuzione non sparsa, ma ben distribuita attorno alla retta blu (che indica l'esatta posizione degli stimoli). Vi sono solo alcuni punti che si allontanano dall'area più densa, probabilmente dovuti ad errori di distrazione o stanchezza, che sono sicuramente presenti in tutti i risultati di tutti i soggetti. Nel secondo grafico osserviamo che la distribuzione diminuisce la pendenza, il soggetto infatti sembra avere difficoltà nel riconoscere suoni molto alti e molto bassi. Nel terzo grafico la situazione degenera e vi sono punti molto sparsi, non è più possibile evidenziare visibilmente una distribuzione crescente nonostante la retta di regressione sia quasi identica a quella del secondo grafico. In questa ultima situazione si evince che il soggetto non riesca a riconoscere le altezze dei vari stimoli che, a parità di stimolo, indica in modo sparso o errato.

Osservando invece i grafici del soggetto 30 osserviamo una distribuzione sparsa ma definita nella Condizione 1, è infatti distinguibile un flusso di punti che partono dal basso e vanno verso l'alto all'aumentare della reale elevazione. Si denota della confusione solo nella percezione di stimoli ad altezza media, la percezione di stimoli alti e bassi invece risulta, a meno di pochi punti, buona. Tutt'altra immagine si ha guardando il grafico associato alla Condizione 2, in questo caso infatti non è possibile fare alcuna osservazione se non che tutti i dati sono completamente sparsi, suoni alti vengono percepiti come bassi e viceversa, suoni di media altezza vengono a volte percepiti come bassi e altre come alti. Vi è quindi una totale confusione nella percezione dell'altezza dello stimolo. Nel terzo grafico la situazione migliora, si denota una miglior capacità di distinguere suoni alti o bassi, anche se non vi è alcuna via di mezzo, da un punto in poi infatti tutti i suoi diventano alti e la distribuzione assume un andamento a gradino. Nonostante la retta di regressione molto vicina alla retta ideale per la Condizione 3, queste ultime considerazioni permettono di concludere che anche in questo caso la Condizione 1 risulta essere quella meglio compresa.

Per ultimo analizziamo i grafici di un soggetto non esperto (come invece lo erano i precedenti). Prendiamo a tal proposito in esame i grafici del soggetto 43. Dall'analisi del primo grafico si evidenzia una distribuzione piuttosto sparsa e confusa per gli stimoli di elevazione bassa, con un leggero miglioramento nella collocazione degli stimoli ad elevazione media, ed infine un ulteriore miglioramento nella percezione degli stimoli ad elevazione alta. Nel secondo grafico si evidenzia invece una concentrazione verticale (asse di *perceived elevation*) nel range da -30° a $+20^\circ$ rendendo la retta di regressione quasi piatta. Questo sta ad indicare che i suoni vengono più o meno percepiti tutti ad una elevazione media, trascurando le elevazioni alte e basse. Nell'ultimo grafico si ha un andamento un po' più uniforme, alcuni stimoli a bassa elevazione vengono percepiti male, ma salendo con le elevazioni si osserva che la percezione migliora. Essa però raggiunge un limite oltre il quale iniziano ad esserci degli errori, si tratta di elevazioni alte che vengono invece percepite più basse di quanto in realtà sono.

Questo tipo di analisi può essere fatta per tutti i grafici, noi ci limitiamo a

descrivere nelle conclusioni un sommario sull'analisi di tutti questi riportando i lati positivi ed i lati negativi evidenziati da questo test pilota sul modello da noi adottato. Andiamo adesso ad analizzare alcuni dati parametrici associati a tutti i soggetti.

4.3 Analisi dettagliata dei risultati parametrici

Nella Tabella 4.1 sono riportati tutti i risultati parametrici. Analizziamo per ogni punto tali risultati:

Mean angle error

I valori più bassi di *mean angle error* si registrano per la condizione 1 (soggetti 20 e 21) con un minimo di circa 12°. I valori più bassi per la condizione 2 sono di circa 17° (soggetto 42) mentre i valori più bassi per la condizione 3 sono di circa 14° (soggetto 22).

In 5 casi (soggetto 20, 21, 22, 35 e 44) si ha che la condizione in cui il *mean angle error* è minimo risulta essere la 1, in tutti gli altri casi la condizione che sembra essere migliore è la condizione 3.

Il risultato è verificato dall'istogramma di Figura 4.2 in cui gli ultimi due risultati corrispondono alle medie dei soggetti esperti (Mean1) e dei soggetti non esperti (Mean2). Per i soggetti esperti (soggetti 20, 21, 22, 30 e 35) si evidenzia il miglioramento della condizione 1 di circa il 15% rispetto alla condizione 3. Lo stesso non si può dire per i soggetti non esperti in per i quali la condizione soggettiva è la peggiore.

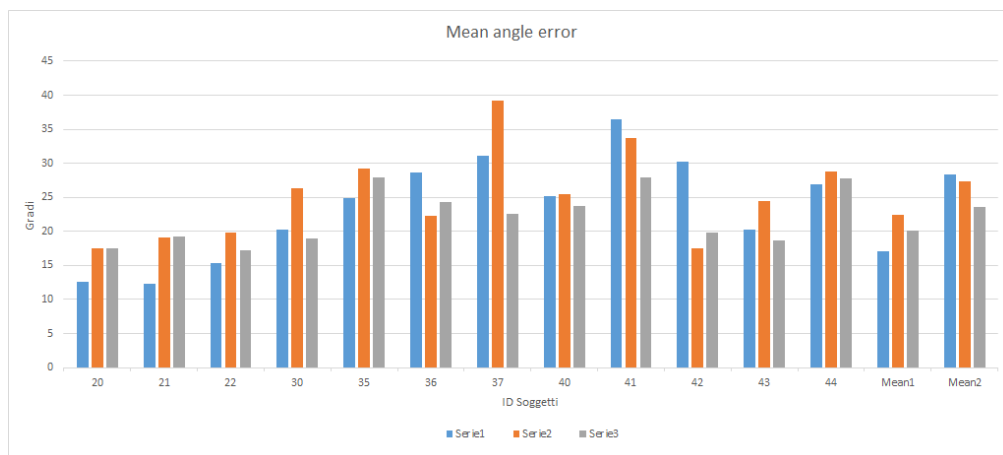


Figura 4.2: Iistogramma relativo al parametro Mean angle error

ID	Cond.	Mean angle error	Slope	Goodness of fit	U/D conf.	F/B conf.
20	1	12.6236	0.7878	0.7225	5.55	0
	2	17.5047	0.6575	0.4687	16.66	0
	3	17.4724	0.5562	0.4092	11.11	0
21	1	12.3699	0.6745	0.7066	0	100
	2	19.1077	0.3262	0.3921	1.85	100
	3	19.2808	0.3309	0.3652	5.55	100
22	1	15.3718	0.6875	0.5072	5.556	77.778
	2	19.8968	0.5261	0.3532	7.407	55.556
	3	17.2785	0.7317	0.5308	5.556	42.593
30	1	20.2942	0.5171	0.3599	5.55	3.704
	2	26.3825	0.3021	0.1042	16.67	90.741
	3	18.9683	0.8197	0.5316	7.407	79.630
35	1	24.8271	0.1993	0.1237	9.259	27.778
	2	29.2784	0.1629	0.0628	22.222	35.185
	3	27.8569	0.2951	0.3351	22.222	48.148
36	1	28.6351	0.0765	0.0099	29.630	81.482
	2	22.2409	0.2524	0.1569	11.11	94.44
	3	24.2675	0.2339	0.1006	11.11	77.778
37	1	31.1596	0.1181	0.0180	27.778	81.482
	2	39.2589	-0.2553	0.0745	42.593	61.111
	3	22.5055	0.4608	0.2844	12.963	62.963
40	1	25.1165	0.2153	0.0891	16.667	53.704
	2	25.4327	0.2662	0.1176	14.815	53.704
	3	23.7545	0.2006	0.1009	9.259	59.259
41	1	36.5264	-0.2671	0.1128	40.741	46.296
	2	33.6920	-0.1732	0.0403	38.889	59.259
	3	27.9421	0.1310	0.0490	20.370	59.259
42	1	30.1941	0.0998	0.0167	27.778	61.111
	2	17.5804	0.4721	0.3785	7.407	90.741
	3	19.8254	0.3919	0.2706	12.963	88.889
43	1	20.2686	0.4387	0.2934	12.963	88.889
	2	24.4825	0.1279	0.0379	12.963	77.778
	3	18.7297	0.5057	0.4141	5.556	81.482
44	1	26.9248	0.1329	0.0828	20.370	94.444
	2	28.7404	0.0880	0.0222	18.519	90.741
	3	27.7863	0.0768	0.0395	18.519	81.482

Tabella 4.1: Risultati di tutti i soggetti

Slope

Guardando la tabella 4.1 si osserva che i valori di *slope* più elevati per la condizione 1 sono di circa 0.79 (soggetto 20), mentre per la condizione 2 sono di

circa 0.66 (soggetto 22), infine per la condizione 3 sono di circa 0.82 (soggetto 30). Quindi sia per la condizione 1 che per la condizione 3 si sono ottenuti valori abbastanza alti. Dall'istogramma di figura 4.3 si osserva come le due categorie, esperti e non esperti, abbiano ottenuto risultati molto diversi, anche qui infatti si ha che i valori migliori li si hanno per i soggetti con esperienza nel campo, negli altri soggetti invece si arriva anche ad avere *slope* negative, indicanti una totale confusione nella collocazione della sorgente del suono. Questo parametro quindi conferma quanto già asserito per il mean angle error.

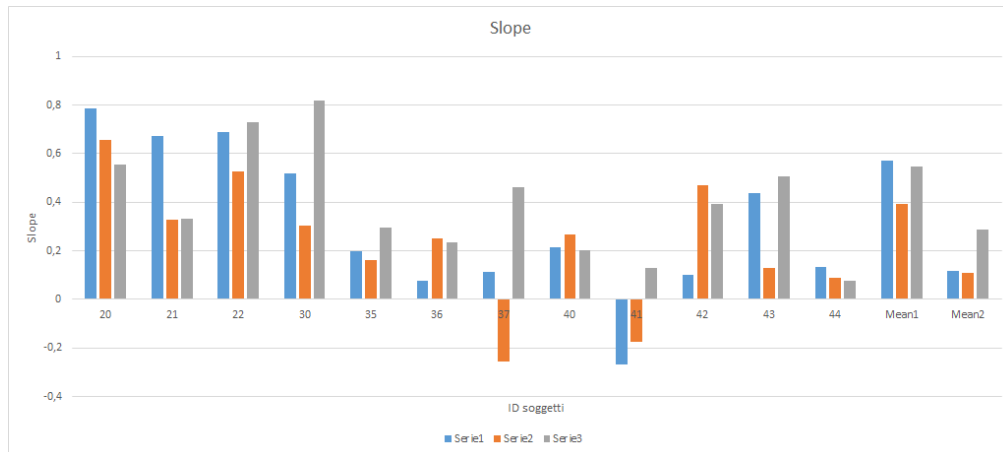


Figura 4.3: Istogramma relativo al parametro Slope

Goodness of fit

Per analizzare questo parametro ci serviamo da subito dell'istogramma di Figura 4.4. Il grafico evidenzia subito quali siano i picchi. Si osservano, come per i parametri precedenti, dei buoni risultati per il gruppo dei soggetti esperti per la condizione 1, anche se la condizione 3 in qualche caso risulta essere migliore. Per quanto riguarda la condizione 3 si nota che vi sono dei picchi anche in altri soggetti, rendendola mediamente anche in questo caso la condizione più facile da riconoscere.

Up/down confusion

Dai valori di Tabella 4.1 si denota subito che questo parametro mostra un risultato molto importante. Per i soggetti esperti (osservando Mean1) si ha un miglioramento del 50% tra la condizione 1 e la condizione 3, infatti il risultato per la KEMAR ha un valore circa il doppio di quello soggettivo. Questo sta ad indicare che con il modello da noi adottato riusciamo a rendere più facile la distinzione tra suono alto e suono basso.

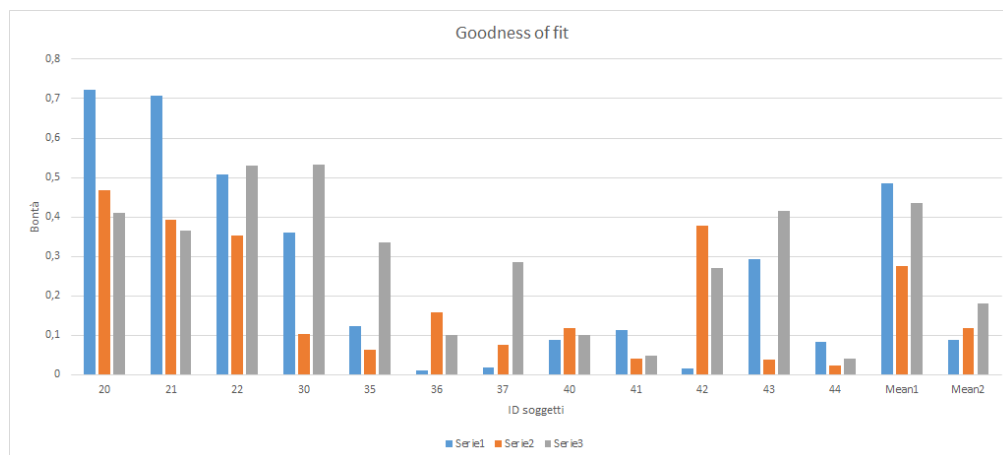


Figura 4.4: Istogramma relativo al parametro Goodness of fit

Per i soggetti non esperti nemmeno in questo caso si ottengono risultati soddisfacenti.

Front/back confusion

Si osserva subito dalla Tabella 4.1 il comportamento anomalo dei primi due soggetti che sentono rispettivamente tutti i suoni di fronte e tutti i suoni dietro. Gli altri soggetti invece presentano un andamento diverso da quello dei parametri precedenti. Sembra infatti che sia più facile riconoscere la posizione, davanti o dietro, per gli stimoli della Condizione 1. Osservando le due medie si osserva che sono entrambe migliori per la condizione 1 (seppur di pochissimo nei soggetti non esperti). Questo è il primo parametro che denota un miglioramento per entrambi i gruppi di soggetti.

4.4 Conclusioni

Dall'analisi grafica eseguita si evidenzia per i soggetti con esperienza un andamento abbastanza accurato per la Condizione 1, confuso e sparso per la Condizione 2 e abbastanza accurato anche per la Condizione 3. I soggetti non esperti hanno invece mostrato molto spesso una elevata confusione per tutti e 3 i grafici, evidenziando qualcosa di positivo solo per la Condizione 3. I problemi più frequenti sono stati dati dalla confusione di suoni alti con suoni bassi, ottenendo una pendenza della retta di regressione anche negativa rispetto a quella reale. In altri casi vi è una incompleta capacità di assegnare un'elevazione su di un sistema di riferimento fisso, osservabile dal fatto che ogni elevazione reale aveva punti segnati di elevazione percepita in tutto il range possibile. In alcuni soggetti

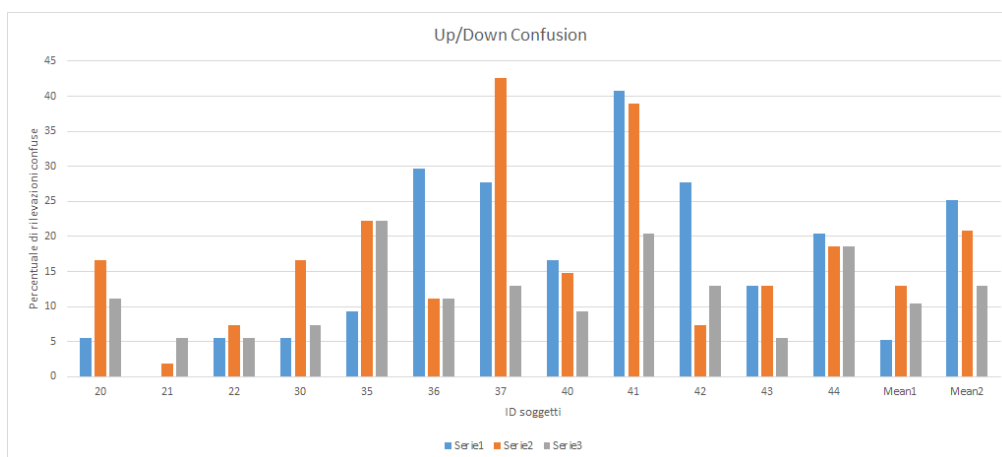


Figura 4.5: Istogramma relativo al parametro Up/Down confusion

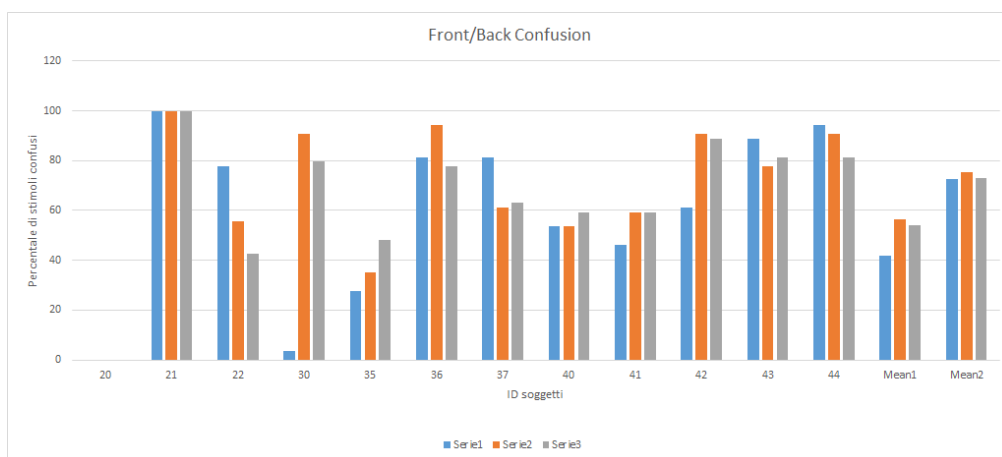


Figura 4.6: Istogramma relativo al parametro Front/Back confusion

si è osservato che vi è molta più difficoltà nel percepire suoni alti che suoni bassi, evidenziati soprattutto nelle Condizioni 1 e 2.

Dall'analisi dei parametri si conferma quanto detto per l'analisi grafica, i dati evidenziano che chi ha sostenuto l'esperimento con alle spalle conoscenze nel campo di applicazione dello stesso, è riuscito ad ottenere risultati nettamente migliori di chi invece si è approcciato per la prima volta a questo tipo di argomento. Il risultato più rilevante ed importante di questo test pilota deriva dall'osservazione che attraverso il modello da noi adottato si è resa una più facile comprensione dello stimolo nelle posizioni davanti/dietro e sopra/sotto l'ascoltatore.

Una giustificazione parziale plausibile dell'andamento non sempre migliore

della condizione soggettiva rispetto alla condizione 3 relativa agli stimoli registrati dalla KEMAR, può pervenire dal fatto che il nostro modello è privo del contributo delle spalle e del torso, che per le elevazioni da noi considerate modificano la HRTF nella banda di frequenze tra 1 e 2 kHz circa e riescono comunque a dare, anche senza il contributo dell'orecchio, una debole indicazione di elevazione della sorgente. Tale contributo di spalle e torso è invece presente nelle risposte del manichino KEMAR.

Un altro fattore che potrebbe aver influito negativamente sta nel fatto che il modello è stato parametrizzato solo sull'orecchio sinistro e quindi non prendiamo in considerazione eventuali differenze interaurali per stimoli sonori nel piano mediano.

4.5 Sviluppi futuri

Da quanto detto nelle conclusioni risulta dunque necessario rivedere il protocollo sperimentale prima di eseguire il test completo, possibilmente inserendo nel nostro modello anche il contributo di spalle e torso. Le altre opzioni plausibili evidenziate da questo test pilota prevedono una fase di training per i soggetti, in modo da dar loro quella componente di competenza che ha permesso ai soggetti considerati nella fascia degli esperti di ottenere buoni risultati.

La valutazione soggettiva è quindi ancora da ritenersi incompleta anche se in via di sviluppo.

Con invece il risultato sulla buona comprensione delle posizioni davanti-dietro e sopra-sotto dello stimolo personale, si aprono le porte a studi su come rendere più marcate queste caratteristiche e su quali siano i motivi che le rendono migliori rispetto agli altri parametri. Queste potrebbero infatti essere le caratteristiche su cui l'esperimento successivo potrebbe convergere.

Capitolo 5

Appendice A

5.1 Output Grafici di tutti i soggetti

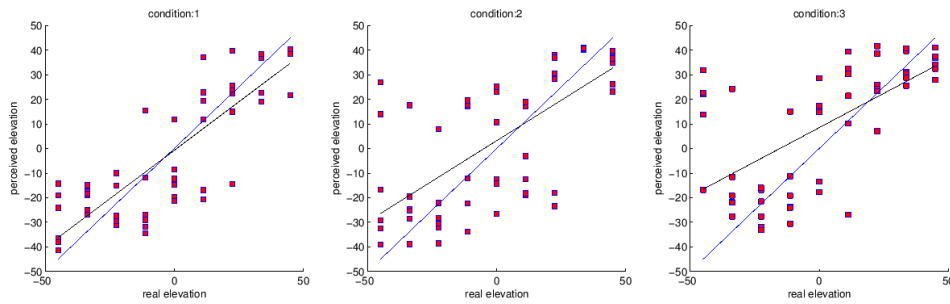


Figura 5.1: Grafici relativi al soggetto 20

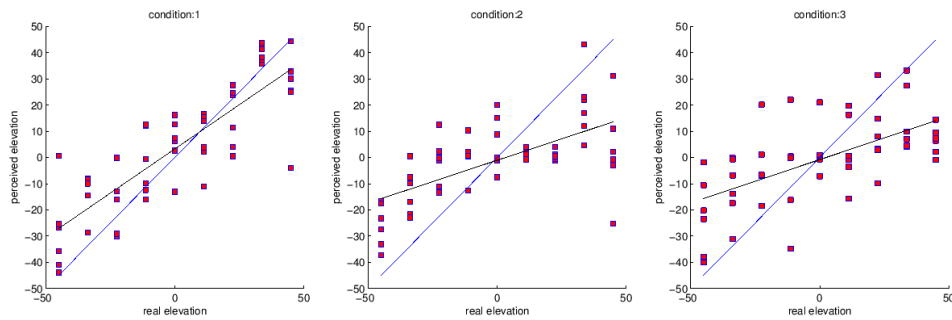


Figura 5.2: Grafici relativi al soggetto 21

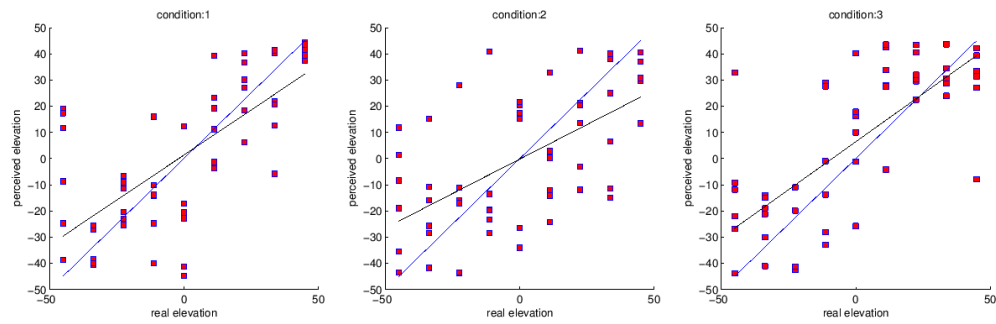


Figura 5.3: Grafici relativi al soggetto 22

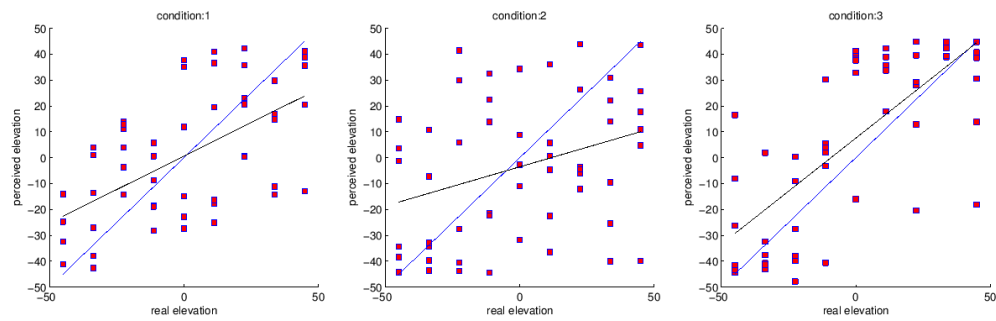


Figura 5.4: Grafici relativi al soggetto 30

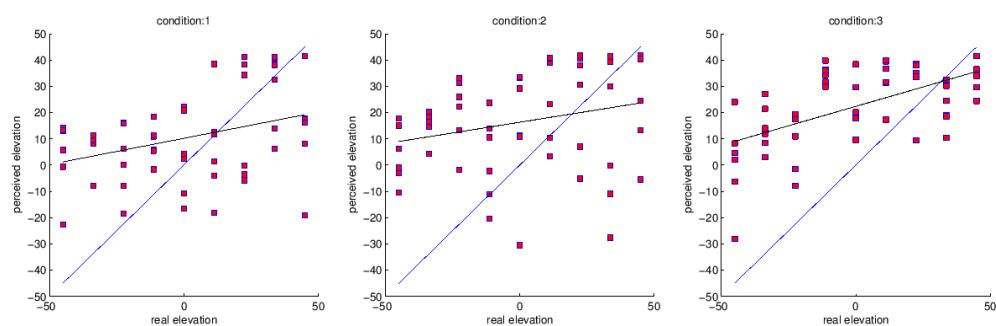


Figura 5.5: Grafici relativi al soggetto 35

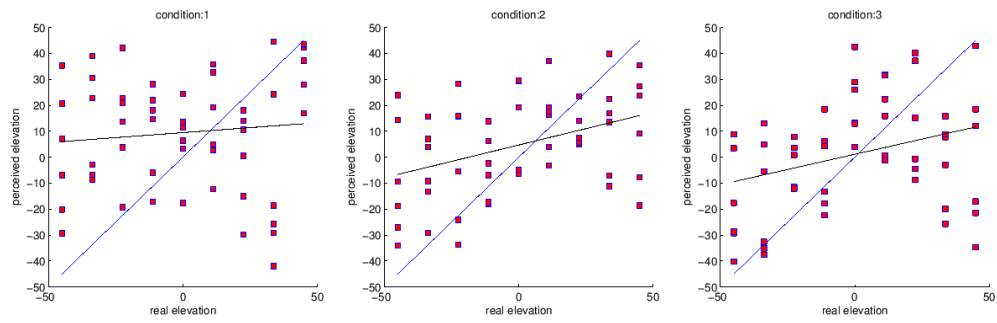


Figura 5.6: Grafici relativi al soggetto 36

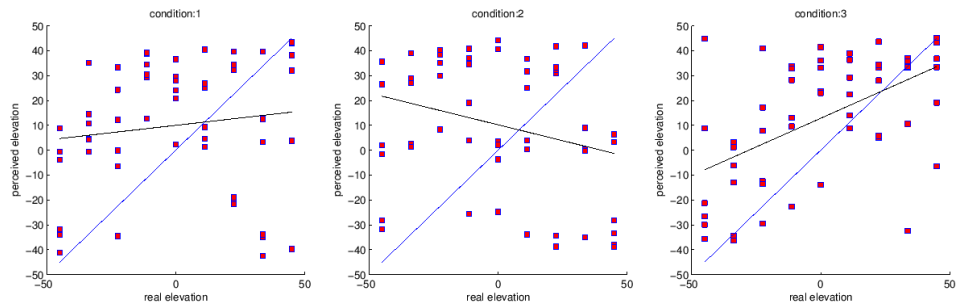


Figura 5.7: Grafici relativi al soggetto 37

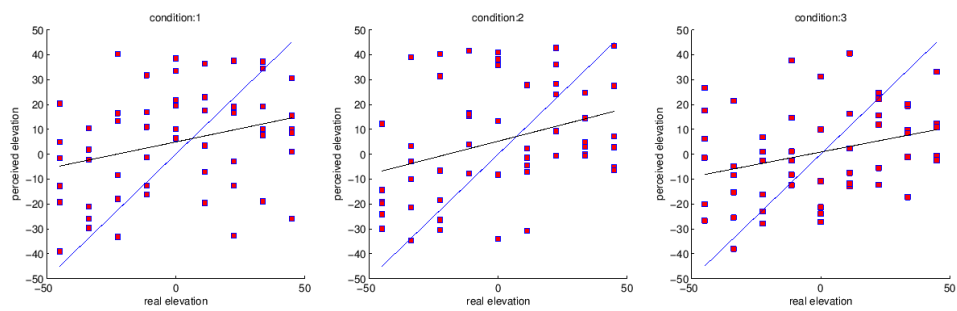


Figura 5.8: Grafici relativi al soggetto 40

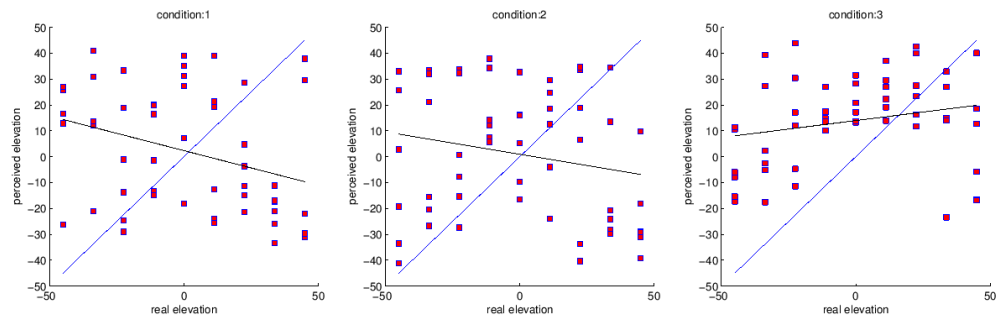


Figura 5.9: Grafici relativi al soggetto 41

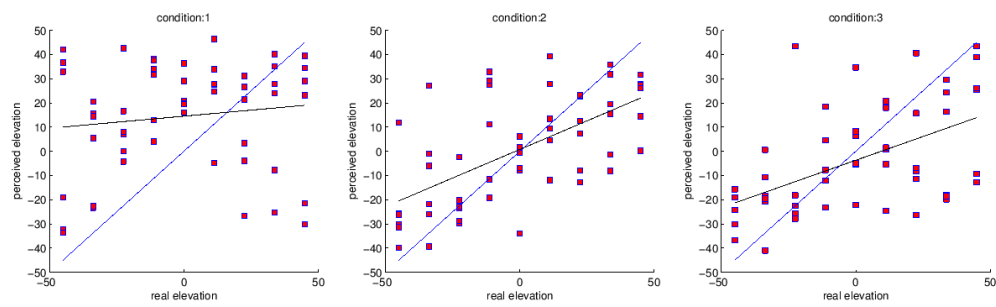


Figura 5.10: Grafici relativi al soggetto 42

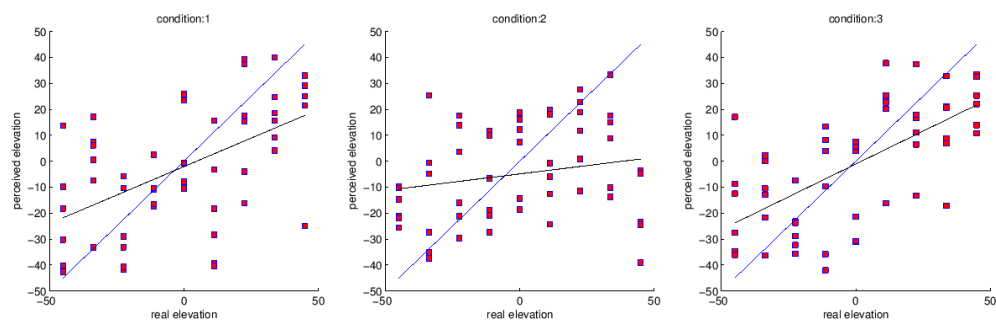


Figura 5.11: Grafici relativi al soggetto 43

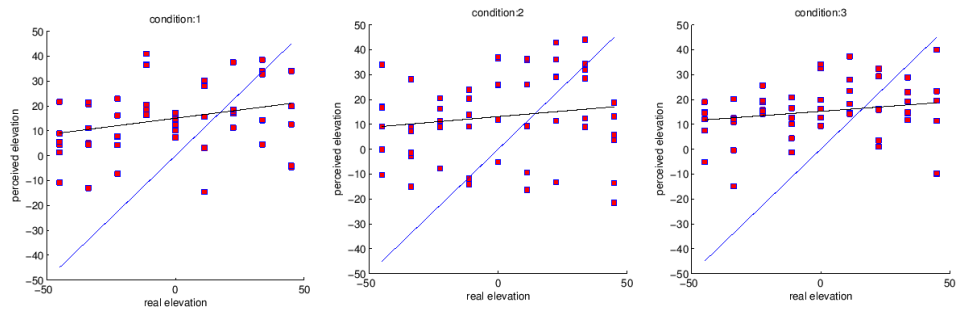


Figura 5.12: Grafici relativi al soggetto 44

Capitolo 6

Appendice B

6.1 Codice

Listing 6.1: Codice per il tracciamento dei contorni della pinna

```
1 function [Cx, Cy] = contour_trace(imageName, num_contours)
3 f = figure;
  set(f,'name','Ear tracing','numbertitle','off');
5
  %Set plural on the title if necessary
7 if ( num_contours > 1 )
    plural = 's';
9 else
    plural = '';
11 end
13
  %read the img
  img = imread(imageName);
15 %create the matrix-image using a flipped version of the original img
  image(flipdim(img,1));
17 %set property values
  set(gca,'YDir','normal');
19 %adjust display size of img
  truesize(f);
21
  % Contours-----
23
  title(['Draw' int2str(num_contours)' line' plural' starting from the most external one']);
25
  %Matrix definition
27 Cx = zeros(2000,num_contours);
  Cy = zeros(2000,num_contours);
29
  % Process to calculate the contours starting from the most external.
31 % Quantized data are stored on Cx and Cy
33 for k = 1:num_contours
    try
35         fhand = imfreehand('Closed', false);
           if ( ~isempty(fhand) )
```

```

37         c = fhand.getPosition;
38         [rows,cols] = size(c);
39         Cx(1:rows,k) = c(:,1); %dimension of Cx and Cy grow as a matrix
40         Cy(1:rows,k) = c(:,2);
41     else
42         return;
43     end
44     catch ME
45         uiwait(msgbox('Tracing aborted.', ...
46             'Abort','none'));
47         return;
48     end
49 end

51 % Focus Point-----
52
53 title(['Now place ear canal position']);
54
55 try
56     [xChosen,yChosen] = ginput(1); %input from mouse cursor
57 catch ME
58     uiwait(msgbox('Tracing aborted.', ...
59         'Abort','none'));
60     return;
61 end
62 earCanal = [xChosen yChosen];
63
64 hold on;
65 plot(xChosen, yChosen, '*');
66
67 % Ear height-----
68
69 prompt = {'Enter image scale (pixel to meters factor):'};
70 dlg_title = 'Pixel to meter';
71 num_lines = 1;
72 def = {'0.00026458333'};
73 pixelToMeterFactor = inputdlg(prompt,dlg_title,num_lines,def);
74 if (~isempty(pixelToMeterFactor))
75     pixelToMeterFactor = str2num(pixelToMeterFactor{1});
76 end
77 if (isempty(pixelToMeterFactor)||~isnumeric(pixelToMeterFactor)||pixelToMeterFactor<=0)
78     uiwait(msgbox('Invalid scale factor. Tracing aborted.', ...
79         'Abort','none'));
80     close();
81     return;
82 end
83
84 % Process earDataFile and calculate F0.
85 % Contours are stored in earDataFile.Cx and Cy as double arrays.
86 % C(1,:) is the first contour (C1/C2), C(2,:) the second one (C4)
87 % while C(3,;) is the third one (C3).
88 % Focus point is stored in earDataFile.earCanal
89
90 % Define conversion factor from pixels to meters according to image
91 % scale
92 % Default: pixelToMeterFactor = 0.00026458333;
93
94
95 dist = zeros(3,9);
96 F0 = zeros(3,9);
97
98 [path, name, ~] = fileparts(imageName);
99 sep = filesep;

```



```

99     save([path sep name '.mat'],'Cx','Cy', 'earCanal', 'F0', 'dist', 'pixelToMeterFactor');
100     uiwait(msgbox('Traces saved successfully.', ...
101                 'Saved','none'));
102     close();
103 end

```

Listing 6.2: Codice per il calcolo di F_0 :

```

1 function pushbutton7_Callback(hObject, eventdata, handles)
2 % hObject    handle to pushbutton7 (see GCBO)
3 % eventdata  reserved - to be defined in a future version of MATLAB
4 % handles    structure with handles and user data (see GUIDATA)
5 selected = get(handles.subjectsListBox, 'Value');
6 cd img;
7 earImage = which(handles.subjects{selected,5});
8 [~, name, ~] = fileparts(earImage);
9 earImage=[name '.mat'];
10 load(earImage);
11 cd ..;
12
13 dist = zeros(3,9);
14 F0 = zeros(3,9);
15
16 for i = 1:3 % cycle contours
17     phiIndex = 1;
18     for phi = -45:11.25:45 % cycle elevations
19         ...11.25
20         minDiff = 360;
21         j = 1;
22         while ( Cx(j,i) ~= 0 ) % cycle points
23             angle = radtodeg(atan((Cy(j,i)-earCanal(2))/(Cx(j,i)-earCanal(1))));
24             if ( angle >= -45 && angle <= 45 )
25                 angleDiff = abs(phi - angle);
26                 if ( angleDiff < 5 )
27                     if ( angleDiff < minDiff )
28                         xChosen = Cx(j,i);
29                         yChosen = Cy(j,i);
30                         minDiff = angleDiff;
31                     end
32                 end
33             end
34             j = j+1;
35         end
36         if ( minDiff ~= 360 )
37             dist(i,phiIndex) = pdist([earCanal(1), earCanal(2); ...
38                                     xChosen, yChosen]) * pixelToMeterFactor;
39             F0(i,phiIndex) = 343.2 / (2*dist(i,phiIndex));
40         else
41             dist(i,phiIndex) = 0;
42             F0(i,phiIndex) = 0;
43         end
44         if F0(i,phiIndex) > 22000
45             F0(i,phiIndex)= 22000;
46         end
47         phiIndex = phiIndex + 1;
48     end
49 end
50 cd img;
51 save(earImage, 'Cx', 'Cy', 'earCanal', 'F0', 'dist', 'pixelToMeterFactor');

```

```

53      uiwait(msgbox('Done.', ...
                    'Saved','none'));
      cd ..;

```

Listing 6.3: Codice per la creazione degli stimoli del soggetto 1

```

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
2  %      Crea Stimoli      %
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
4  % --- Executes on button press in Generate.
function Generate_Callback(hObject, eventdata, handles)
6  % hObject      handle to Generate (see GCBO)
% eventdata      reserved - to be defined in a future version of MATLAB
8  % handles      structure with handles and user data (see GUIDATA)

10 %Find current selected ID-----
    selected = get(handles.subjectsListBox,'Value');
12     id = handles.subjects{selected,6};

14 %Check repetitions number-----
    repetitions = get(handles.repetitionsText, 'String');
16     repetitions = str2double(repetitions);

18     cd subjects

20     if ( (exist(['subject_' int2str(id) '/personal_stimoli_wav'],'dir')~=0)&&...
           (exist(['subject_' int2str(id) '/personal_filters'],'dir')~=0)&&...
           (exist(['subject_' int2str(id) '/data.mat'],'file')~=0))
22         choice = questdlg('Stimoli già presenti per il soggetto, sovrascrivere?', ...
24             'Confirm changes', ...
                'Yes', 'No', 'No');
26         switch choice
                case 'Yes'
28             case 'No'
                    cd ..;
30             return;
                end
32     end
    cd ..

34
36 % Load subject list and extract data
    subjectsStruct = load('subjects');
    subjects = subjectsStruct.subjects;
38

40     if ( isempty(subjects(:,1)) ),
        disp('Database empty. Abort. ');
        close();
42     end

44     numOfRecords = length(subjects(:,1));
    for i = 1:numOfRecords,
46         if ( subjects{i,6} == id )
                lastName = subjects{i,1};
48                 firstName = subjects{i,2};
                gender = subjects{i,3};
50                 age = subjects{i,4};
                [~, earDataFile, ~] = fileparts(subjects{i,5});
52                 earDataFile = [earDataFile '.mat'];
                cd img;
54                 earDataFile = load(earDataFile);

```

```

        cd ..;
56         break;
        end
58     end

60     data.id=id;
        data.lastName=lastName;
62     data.firstName=firstName;
        data.gender=gender;
64     data.age=age;

66     % Extract stored F0 value from ear trace.
        F0 = earDataFile.F0;
68

70     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
71     %      Creazione blockTest      %
72     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
73     % Prepare arrays with test data, calculate test number and random
74     % shuffle test order.

76     load('positions.mat');
        [numElevation, ~] = size(positions);
78     handles.numOfTests = 6*numElevation * repetitions;
        handles.condTest = zeros(2,repetitions*numElevation);
80     handles.blockTest = zeros(repetitions*2,numElevation*repetitions);

82     %%% SCELTA OPZIONE %%%%%%%%%%%%%%%

84     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
85     % A = 1 personal %
86     % B = 2 KemarP %
87     % C = 3 kemarNP %
88     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

90     choice = questdlg('Express type of option', ...
        'Confirm changes', ...
92     '1', '2', '3', 'Cancel');
        switch choice
94         case '1'
                block=[1 2 3 2 3 1];
96                 data.optionChoise=1;
            case '2'
                block=[2 3 1 3 1 2];
98                 data.optionChoise=2;
            case '3'
                block=[3 1 2 1 2 3];
100                data.optionChoise=3;
            case 'Cancel'
102                return;
        end
104     end

106     %%% Generazione blocktest %%%%%%%%%%%%%%%

108     for k=1:6
110
111         pos = 1;
112
113         for i = 1:repetitions
114
115             for j = 1:numElevation
116

```

```

118         handles.condTest(1,pos) = block(k);
        handles.condTest(2,pos)=positions(j,1);
        pos = pos + 1;
120
        end
122
        end
124
        %RANDOMIZZAZIONE%
126        randomOrder = randperm(repetitions*numElevation);
        handles.condTest = handles.condTest(:,randomOrder);
128        blockTest([k*2-1,k*2],:) = handles.condTest([1,2],:);
130
        end
132        %SALVATAGGIO_BLOCKTEST% *contiene tutte le sequenze
        data.blockTest = blockTest(:,:);
134
        %%%%%%%%%CALL%%%%%%%%
        create_subject(F0,id); %
136        %%%%%%%%%%%%%%%%%
138        save(['subjects\' ['subject_' int2str(id) '\'] ['data' '.mat'] ],'data');

```

Listing 6.4: Codice per la creazione degli stimoli del soggetto 2

```

function []=create_subject(F0,id)
2 %IN:
%
4 % F0 : vettore frequenze centrali notch
% id : subject id
6
h = waitbar(0,'Please wait...Creating filters');
8
%%%%%%%%
10 % Load and Definitions %
%%%%%%%%
12
load('peak_first_interp.mat'); % load matrice con i dati del primo peak
14 load('peak_second_interp.mat'); % load matrice con i dati del secondo peak
load('equalization_filter.mat'); % load filtro di equalizzazione headphone
16 ...s=which('pinna050.mat');
...load(s,'F0');
18
strId=int2str(id); %per velocizzare il codice..
20
%%%%%%%%
22 % Creazione Cartelle utente %
%%%%%%%%
24 cd subjects;
26
%se la cartella già esiste eliminatala
28 if (exist(['subject_' strId])~=0)
    rmdir(['subject_' strId],'s');
end
30
%generazione nuove cartelle personal
32 % creo directory subject_id e filters contenente i filtri peak+notch
mkdir(['subject_' strId ],'personal_filters');
34 % creo directory stimoli_wav contenente stimoli già filtrati da inviare in cuffia
mkdir(['subject_' strId ],'personal_stimoli_wav');

```

```

36     cd ...;
38     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
39     % Personal Stimoli Generation %
40     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
41     for i=1:9
42         c=-45+(i-1)*(11.25);
43
44         BW=0.15*F0(:,9-(i-1))';
45         fr_centrali=F0(:,9-(i-1))';
46         depth=[-30,-30,-30];
47
48         [f]=prtf([new_peak(i,1),new_peak_second(i,1)],[new_peak(i,2),...
49                 new_peak_second(i,2)],[new_peak(i,3),new_peak_second(i,3)],...
50                 fr_centrali,depth,BW); % calcolo della prtf
51
52         [l]=stimulus3(0.3);
53         y=filter(f,l); % filtraggio dello stimolo con prtf
54         max_val = 1.05*max(max(abs(y)));
55         y = y/max_val;
56         y=fftfilt(data_IR,[y';zeros(length(data_IR)-1,1)]);
57
58         wavwrite(y,48000,32,['subjects\' ['subject_' strId] ...
59                 '\personal_stimoli_wav\' ['cond_11_elev_' int2str(c)]]
60
61         save(['subjects\' ['subject_' strId] '\personal_filters\' ...
62                 ['cond_11_elev_' int2str(c),'.mat']], 'f');
63
64         waitbar(((i-1)*12+11)/84);
65     end
66     close(h);

```

Listing 6.5: Codice della sintesi della PRTF

```

% IN:
2 % fc_peak : vettore di frequenze centrali -> peak
3 % a_peak  : vettore di ampiezze corrispondenza vettore fc_peak
4 %         fc_peak(i) -> a_peak(i)
5 % fb_peak : vettore di bande passanti con corrispondenza vettore fc_peak
6 %         fc_peak(i) -> fb_peak(i)
7 % fc_notch: vettore di frequenze centrali -> notch
8 % a_notch : vettore di ampiezze corrispondenza vettore fc_notch
9 %         fc_notch(i) -> a_notch(i)
10 % fb_notch: vettore di bande passanti con corrispondenza vettore fc_notch
11 %         fc_notch(i) -> fb_notch(i)
12
13 % OUT:
14 % Prtf    : modello prtf dato dalla cascata del modello di risonanza
15 %         e di quello riflessivo
16
17
18 function [ Prtf ] = prtf(fc_peak,a_peak,fb_peak,fc_notch,a_notch,fb_notch)
19
20 % costruisce modello di risonanza:
21 [H1,A1,b,a]=exp1_model_resonance(fc_peak,a_peak,fb_peak);
22 % trasforma output modello risonanza in filter object:
23 Pr=dfilt.df2(b,a);
24 % costruisce modello di riflessione:
25 [H,A,Nr]=exp1_model_reflection(fc_notch,a_notch,fb_notch);
26 % costruisce modello prtf dalla cascata dei due modelli creati:

```

```

Prtf=cascade(Pr,Nr);
28
end

```

Listing 6.6: Codice per la riproduzione degli stimoli e l'acquisizione dei dati

```

1 function listenButton_Callback(hObject, eventdata, handles)
2 % hObject    handle to listenButton (see GCBO)
3 % eventdata  reserved - to be defined in a future version of MATLAB
4 % handles    structure with handles and user data (see GUIDATA)
5
6     if ( handles.currentTest ~= 0 )
7         handles.endTime = round(etime(clock,handles.startTime) * 1000);
8         handles.experimentResults{handles.currentTest, 1}= ...
9             handles.chosenElevation;
10        handles.experimentResults{handles.currentTest, 2}= ...
11            handles.blockTest(2*handles.nr_pause,handles.counter);
12        handles.experimentResults{handles.currentTest, 3}= ...
13            handles.blockTest(2*handles.nr_pause-1,handles.counter);
14        ...handles.experimentResults{handles.currentTest, 3} = ...
15            handles.endTime;
16
17        end
18
19        handles.currentTest = handles.currentTest + 1;
20        handles.counter=handles.counter+1;
21        set(handles.currentTestText, 'String', [int2str(handles.currentTest) ...
22            ' of ' int2str(handles.numOfTests)]);
23        guidata(hObject, handles);
24
25
26        if ( handles.currentTest > handles.numOfTests )
27            cd results;
28            results = handles.experimentResults;
29            save([int2str(handles.id) '.mat'],'results');
30            cd ..;
31            uiwait(msgbox('Experiment completed and saved.', ...
32                'Completed','none'));
33            guidata(hObject, handles);
34            close();
35            return;
36        end
37
38        if ( handles.currentTest - 1 ~= 0 )
39            set(handles.listenButton,'Enable','off');
40            delete(handles.elevationPlotHandler);
41        else
42            set(handles.listenButton,'Enable','off');
43        end
44
45 % Pause and saving partial results
46     if ( handles.currentTest-1 ~= 0 && mod(handles.currentTest-1, handles.pauseInterval) == 0 )
47         cd results;
48         results = handles.experimentResults;
49         save([int2str(handles.id) '_autosave.mat'],'results');
50         cd ..;
51         set(handles.statusText, 'String', 'Take a short break before continuing..');
52         pause(1);
53         countdown(0,30);
54         pause(32);

```

```

55     box = msgbox('Click ok when ready: the next trial will start.', ...
56               'Take a break!', 'help');
57     uiwait(box);
58     handles.nr_pause=handles.nr_pause+1;
59     handles.counter=1;
60     end
61
62     handles.startTime = clock;
63
64     set(handles.listenButton, 'Enable', 'off');
65     set(handles.continueButton, 'Enable', 'off');
66
67     pause(0.5);
68
69     ...axes(handles.elevationAxes);
70     ...img = imread('img/newhead_yellow.jpg');
71     ...imagesc([0, 1], [0, 1], flipdim(img,1));
72     ...set(gca, 'YDir', 'normal');
73     ...axis image off;
74
75     set(handles.statusText, 'BackgroundColor', 'yellow');
76     set(handles.statusText, 'String', 'Sound playing ...');
77     pause(0.3);
78
79     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
80     %   RIPRODUZIONE STIMOLI
81     %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
82
83     switch handles.blockTest(2*handles.nr_pause-1, handles.counter)
84
85     case 1 % STIMOLO PERSONALE
86
87         y=wavread(['subjects\' ['subject_' int2str(handles.id)] '\personal_stimoli_wav\'...
88                 ['cond_11_elev_' int2str(handles.blockTest(2*handles.nr_pause, handles.counter))]);
89
90     case 2 %STIMOLO KEMARP
91
92         y=wavread(['subject_KemarP\KemarP_stimoli_wav\'...
93                 ['cond_11_elev_' int2str(handles.blockTest(2*handles.nr_pause, handles.counter))]);
94
95     case 3 %STIMOLO KEMARNP
96
97         y=wavread(['subject_KemarNP\KemarNP_stimoli_wav\'...
98                 ['cond_11_elev_' int2str(handles.blockTest(2*handles.nr_pause, handles.counter))]);
99     end
100
101     %PLAY STIMOLO
102     wavplay(y, 48000);
103
104     clear y;
105
106     ...axes(handles.elevationAxes);
107     ...img = imread('img/newhead_green.jpg');
108     ...imagesc([0, 1], [0, 1], flipdim(img,1));
109     ...set(gca, 'YDir', 'normal');
110     ...axis image off;
111     ...pause(0.1);
112     set(handles.statusText, 'BackgroundColor', 'green');
113     set(handles.statusText, 'String', 'Choose elevation inside the green area. ');
114     pause(0.2);
115     alreadyChosenElevation = false;

```

```

117 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
118 %      ACQUISIZIONE INPUT ELEVAZIONE      %
119 %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

121 while ( true )
122     try
123         [xChosen,yChosen] = ginput(1);
124     catch ME
125         return
126     end

127     if ( xChosen < 0 || xChosen > 1 || yChosen < 0 || yChosen > 1 )
128         uiwait(msgbox('Choose a point inside the green area.', 'Error','warn'));
129         continue;
130     end

131

132     if ( gca == handles.elevationAxes )
133         xElevationChosen = xChosen;
134         yElevationChosen = yChosen;
135         hold on;
136         if ( alreadyChosenElevation == true )
137             delete(elevationPlotHandler);
138             elevationPlotHandler = plot(xElevationChosen, yElevationChosen, '*');
139         else
140             elevationPlotHandler = plot(xElevationChosen, yElevationChosen, '*');
141             alreadyChosenElevation= true;
142         end
143     end
144     if ( alreadyChosenElevation )
145         handles.chosenElevation = getAngle(0.5, 0.5, 1, 0.5, xElevationChosen, ...
146             yElevationChosen);
147         if ( xElevationChosen > 0.5 && yElevationChosen < 0.5 )
148             handles.chosenElevation = -handles.chosenElevation;
149         end
150         if ( xElevationChosen < 0.5 && yElevationChosen < 0.5 )
151             if ( handles.chosenElevation ~= 180 )
152                 handles.chosenElevation = 360 - handles.chosenElevation;
153             end
154         end
155     end

156     if ((handles.chosenElevation > 48 && handles.chosenElevation < 132) ...
157         || (handles.chosenElevation > 228 || handles.chosenElevation < -48)==1)
158         uiwait(msgbox('Choose a point inside the green area.', 'Error','warn'));
159         continue;
160     else
161         break
162     end
163 end

164 end

165

166 handles.elevationPlotHandler = elevationPlotHandler;

167

168

169

170 ...axes(handles.elevationAxes);
171 ...img = imread('img/newhead2.jpg');
172 ...imagesc([0, 1], [0, 1], flipdim(img,1));
173 ...set(gca,'YDir','normal');
174 ...axis image off;
175 ...pause(0.1);
176 set(handles.statusText, 'BackgroundColor', 'red');
177 set(handles.statusText, 'String', 'Loading next sound');

```

```
179     ...set(handles.statusText, 'String', 'When ready press "Listen now".');
180     set(handles.listenButton, 'Enable', 'off');
181     guidata(hObject, handles);
182     pause(0.5);
183     listenButton_Callback(hObject, eventdata, handles);
```

Bibliografia

- [1] S. Spagnol, M. Geronazzo, and F. Avanzini, *On the relation between Pinna Reflection Patterns and Head-Related Transfer Function features*. IEEE transactions on audio, speech, and language processing, vol 21, March 2013.
- [2] F. Avanzini and G. D. Poli, “Algorithms for sound and music computing.” <http://smc.dei.unipd.it/education.html>, 2010.
- [3] M. Geronazzo, S. Spagnol, and F. Avanzini, “Customized 3d sound for innovative interaction design.” CHIItaly, 2011.
- [4] S. Scaiella, “Valutazione parametrica di un modello strutturale di orecchio esterno per il rendering spaziale del suono.” Master’s thesis, University of Padova, 2014.
- [5] M. Geronazzo, S. Spagnol, and F. Avanzini, “Estimation and modeling of pinna-related transfer functions.” In Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10), 2010.
- [6] J. Blauert, “Spatial hearing: Pyschophysics of human sound localization..” MIT Press, 2nd edition, 1996.

Elenco delle tabelle

2.1	Condizioni sperimentali	19
3.1	Soggetti	24
3.2	Sequenze di riproduzione	25
3.3	Elevazioni considerate	25
4.1	Risultati di tutti i soggetti	35

Elenco delle figure

1.1	Visualizzazione grafica di ITD ed ILD	3
1.2	Stima ITD semplificata	4
1.3	Modulo sfera	5
1.4	Orecchio esterno	6
1.5	Coordinate sferiche	9
1.6	Schema a blocchi HRTF	10
1.7	Schema a blocchi modello strutturale	10
2.1	Effetti della pinna	14
2.2	Modello per la pinna	15
2.3	Regione di validità del modello	16
2.4	Tracciamento contorni della pinna	17
2.5	Punti di riflessione in coordinate polari	17
2.6	Estrazione frequenze notch	18
2.7	Modello completo di HRTF	19
2.8	PRTF a confronto: (a) originale, (b) ricostruita dal modello	20
3.1	Cabina silente: Sound Station Pro 45	23
3.2	Caratteristiche della Sound Station Pro 45	23
3.3	GUI per l'inserimento dati e tracciamento contorni	26
3.4	Setup acquisizione pinna	27
3.5	Immagine di pinna fotografata	27
3.6	Contorni tracciati su pinna	28
3.7	Creazione stimoli personali	29
3.8	GUI esperimento	30
4.1	Grafici del soggetto ID 30	32
4.2	Istogramma relativo al parametro Mean angle error	34
4.3	Istogramma relativo al parametro Slope	36
4.4	Istogramma relativo al parametro Goodness of fit	37
4.5	Istogramma relativo al parametro Up/Down confusion	38
4.6	Istogramma relativo al parametro Front/Back confusion	38
5.1	Grafici relativi al soggetto 20	41

5.2	Grafici relativi al soggetto 21	41
5.3	Grafici relativi al soggetto 22	42
5.4	Grafici relativi al soggetto 30	42
5.5	Grafici relativi al soggetto 35	42
5.6	Grafici relativi al soggetto 36	43
5.7	Grafici relativi al soggetto 37	43
5.8	Grafici relativi al soggetto 40	43
5.9	Grafici relativi al soggetto 41	44
5.10	Grafici relativi al soggetto 42	44
5.11	Grafici relativi al soggetto 43	44
5.12	Grafici relativi al soggetto 44	45