

UNIVERSITÀ DEGLI STUDI DI PADOVA  
DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE  
CORSO DI LAUREA IN INGEGNERIA DELL'INFORMAZIONE

*TESI DI LAUREA TRIENNALE*

# Valutazione soggettiva di un modello di campo vicino per la percezione binaurale della distanza relativa tra sorgenti sonore virtuali

**RELATORE:** Prof. Federico Avanzini

**CORRELATORE:** Dott. Simone Spagnol

**LAUREANDA:** Erica Tavazzi

Padova, 22 luglio 2014

A.A. 2013-2014



# Abstract

Nel contesto dello studio di modelli matematici che permettono il rendering 3D di suoni, si propone un modello a basso costo di sorgente sonora virtuale nel campo vicino. Ne viene fatta l'analisi dal punto di vista teorico e si presentano i risultati dell'esperimento di psicoacustica condotto ai fini di (1) paragonare la performance del modello alla controparte teorica; (2) testare l'efficacia del modello nel rendere la distanza relativa tra sorgenti sonore virtuali, paragonandolo a condizioni di controllo presenti in letteratura.

I dati raccolti hanno mostrato buoni risultati per quanto riguarda il primo obiettivo, mentre per raggiungere il secondo saranno necessari ulteriori studi.

È inoltre emerso un risultato interessante per quanto riguarda l'ordine di presentazione degli stimoli, che potrebbe aprire nuovi orizzonti di ricerca.



# Indice

<b>Abstract</b>	<b>i</b>
<b>Sommario</b>	<b>1</b>
<b>1 Il suono nello spazio</b>	<b>3</b>
1.1 Il sistema di riferimento . . . . .	3
1.1.1 Sistema di coordinate verticale-polare . . . . .	4
1.1.2 Sistema di coordinate interaurale-polare . . . . .	4
1.2 La testa . . . . .	4
1.2.1 ITD . . . . .	5
1.2.2 ILD . . . . .	6
1.3 L'orecchio esterno . . . . .	6
1.3.1 La pinna . . . . .	6
1.3.2 Il canale uditivo . . . . .	6
1.3.3 L'orecchio esterno come riflettore sonoro . . . . .	7
1.4 Il busto e le spalle . . . . .	7
1.5 Head-related transfer function . . . . .	8
1.6 Percezione della posizione di una sorgente sonora . . . . .	9
1.6.1 Internalizzazione ed esternalizzazione . . . . .	9
1.6.2 Percezione azimutale . . . . .	9
1.6.3 Percezione dell'elevazione . . . . .	10
1.6.4 Percezione della distanza . . . . .	11
1.7 Algoritmi per il rendering 3-D del suono . . . . .	12
1.7.1 Rendering spaziale del suono . . . . .	12
1.7.2 Tecnologie . . . . .	13
1.7.3 Rendering basato sulle HRTF . . . . .	13
<b>2 Un modello di HRTF nel campo vicino</b>	<b>19</b>
2.1 Far-field e Near-field . . . . .	19
2.2 Sintesi dell'HRTF . . . . .	20
2.3 Validazione teorica della NFTF . . . . .	22
2.3.1 Comportamento in continua . . . . .	22
2.3.2 Comportamento in frequenza . . . . .	23

2.3.3	Descrizione del modello e considerazioni . . . . .	26
<b>3</b>	<b>Design dell'esperimento</b>	<b>29</b>
3.1	Apparato sperimentale . . . . .	29
3.1.1	Matlab . . . . .	29
3.1.2	Cuffie . . . . .	29
3.1.3	Scheda audio esterna . . . . .	31
3.1.4	Pulsanti USB . . . . .	31
3.1.5	Cabina Silente . . . . .	32
3.2	Soggetti . . . . .	33
3.3	Stimoli . . . . .	33
3.4	Condizioni . . . . .	34
3.5	Scelte sperimentali . . . . .	34
3.6	Esecuzione dell'esperimento . . . . .	35
<b>4</b>	<b>Risultati dell'esperimento</b>	<b>37</b>
4.1	Risultati dei soggetti . . . . .	37
4.2	Analisi statistica . . . . .	39
4.2.1	Analisi sulle coppie di distanza . . . . .	40
4.2.2	Analisi sugli azimut . . . . .	40
4.2.3	Analisi sulle condizioni . . . . .	41
4.2.4	Analisi sulle coppie di distanza: stimoli in ordine ascendente/di- scendente . . . . .	42
4.2.5	Analisi sugli azimut: stimoli in ordine ascendente/discendente . . .	43
4.2.6	Interazione tra variabili indipendenti . . . . .	45
<b>5</b>	<b>Conclusioni e sviluppi futuri</b>	<b>47</b>
<b>A</b>	<b>Algoritmi MATLAB</b>	<b>49</b>
A.1	stimulus . . . . .	49
A.2	create_hrtfs . . . . .	50
A.3	hrir . . . . .	51
A.4	hrtf . . . . .	52
A.5	theo_nftf . . . . .	53
A.6	shelving . . . . .	54
A.7	approx_nftf . . . . .	55
A.8	random_sort . . . . .	59
A.9	run_exp . . . . .	60
A.10	countdown . . . . .	63
A.11	error_count . . . . .	65
	<b>Bibliografia</b>	<b>69</b>

# Elenco delle figure

1.1	Sistemi di coordinate sferiche: (a) sistema di coordinate verticale-polare, e (b) sistema di coordinate interaurale-polare . . . . .	4
1.2	Stima dell'ITD in caso di sorgente sonora distante (onde piane) e testa sferica . . . . .	5
1.3	Orecchio esterno: (a) pinna, e (b) canale uditivo . . . . .	7
1.4	Effetti del busto: (a) riflessione, e (b) schermatura . . . . .	8
1.5	Cono di confusione . . . . .	11
1.6	Schema a blocchi di un sistema di rendering 3-D del suono tramite cuffie basato sulle HRTF. . . . .	14
1.7	Profilo ed interno di un manichino KEMAR. . . . .	15
1.8	Schema a blocchi di un sistema di rendering 3-D del suono tramite cuffie basato su un modello strutturale. . . . .	16
2.1	STF calcolata per $\rho = 1.25$ ( <i>near-field</i> ) e $\rho \rightarrow +\infty$ ( <i>far-field</i> ). . . . .	21
2.2	Guadagno DC della NFTF. . . . .	22
2.3	Comportamento in frequenza della NFTF normalizzata $\hat{H}_{NF}$ per $\rho = 1.25$ . . . . .	24
2.4	Confronto tra la NFTF normalizzata $\hat{H}_{NF}$ (a sinistra) e la sua approssimazione tramite il shelving filter $H_{sh}$ (a destra) per $\rho = 1.25$ , $\rho = 4$ , $\rho = 16$ . . . . .	25
2.5	Schema a blocchi del modello per gli effetti binaurali nel campo vicino che tiene conto della distanza. . . . .	26
2.6	Distorsione spettrale introdotta dal modello $H_{dist}$ . . . . .	27
3.1	Sennheiser HDA 200. . . . .	30
3.2	Sennheiser HDA 200: risposta in frequenza (in alto) e deviazione standard su 10 riposizionamenti (in basso). . . . .	30
3.3	Roland Edirol UA-101. . . . .	31
3.4	USBButton™. . . . .	31
3.5	Sound Station Pro 45 by Puma s.r.l. . . . .	32
3.6	Caratteristiche di isolamento acustico e riverbero interno della Sound Station Pro 45 by Puma s.r.l. . . . .	32
4.1	Media e deviazione standard dell'errore sulle <b>coppie di distanze</b> . . . . .	41

4.2	Media e deviazione standard dell'errore sugli <b>azimut</b> . . . . .	41
4.3	Media e deviazione standard dell'errore sulle <b>condizioni</b> . . . . .	42
4.4	Stimoli in ordine <b>ascendente</b> : media e deviazione standard dell'errore sulle <b>coppie di distanze</b> . . . . .	43
4.5	Stimoli in ordine <b>discendente</b> : media e deviazione standard dell'errore sulle <b>coppie di distanze</b> . . . . .	43
4.6	Stimoli in ordine <b>ascendente</b> : media e deviazione standard dell'errore sugli <b>azimut</b> . . . . .	44
4.7	Stimoli in ordine <b>discendente</b> : media e deviazione standard dell'errore sugli <b>azimut</b> . . . . .	44

# Elenco delle tabelle

2.1	Coefficienti per l'equazione di $\tilde{G}_0$ (2.5) ed errore RMS. . . . .	23
2.2	Coefficienti per l'equazione di $\tilde{G}_\infty$ (2.9) ed errore RMS. . . . .	25
	Tabella 2.3: Coefficienti per l'equazione di $\tilde{f}_c$ (2.10) ed errore RMS.	
3.1	Soggetti dell'esperimento. . . . .	33
4.1	Tasso d'errore (%) per <b>coppie di distanze</b> . . . . .	38
4.2	Tasso d'errore (%) per <b>azimut</b> . . . . .	38
4.3	Tasso d'errore (%) per <b>condizione</b> . . . . .	38
4.4	Stimoli in ordine <b>ascendente</b> : tasso d'errore (%) per <b>coppie di distanze</b> . . . . .	39
4.5	Stimoli in ordine <b>discendente</b> : tasso d'errore (%) per <b>coppie di distanze</b> . . . . .	39
4.6	Stimoli in ordine <b>ascendente</b> : tasso d'errore (%) per <b>azimut</b> . . . . .	39
4.7	Stimoli in ordine <b>discendente</b> : tasso d'errore (%) per <b>azimut</b> . . . . .	40



# Sommario

Il veloce sviluppo tecnologico degli ultimi anni ha permesso la diffusione sempre più capillare di dispositivi che prevedono la riproduzione audio tridimensionale. Si è reso quindi necessario indagare con accuratezza il comportamento di sorgenti sonore reali, e lo sviluppo di tecniche per permetterne la virtualizzazione.

Si presenta, in questa tesi, un modello di sorgente virtuale a basso costo nel campo vicino sviluppato da alcuni componenti del *Sound and Music Computing Group* di Padova. Per testarne la validità pratica, è stato condotto un esperimento psicoacustico, che viene di seguito descritto: in particolare, si è testata l'efficacia del modello nel rendere la distanza relativa tra due sorgenti virtuali.

Il **capitolo 1** descrive come un'onda sonora che si propaga nello spazio venga modificata dall'interazione con le diverse parti del corpo umano, nel tragitto dalla sorgente fino alla membrana del timpano. Il suono che raggiunge il ricevitore è descritto univocamente dalla Head Related Transfer Function (HRTF), funzione misurabile e riproducibile tramite adeguati modelli matematici.

Il **capitolo 2** presenta gli studi condotti nel corso degli ultimi anni sulle HRTF nel campo vicino: si confrontano i modelli matematici presenti in letteratura e viene descritto il modello di sorgente sonora virtuale nel campo vicino che ci si è proposti di analizzare, validandolo dal punto di vista teorico.

Nel **capitolo 3** si descrivono l'apparato sperimentale utilizzato, i soggetti coinvolti e le scelte effettuate nel design dell'esperimento.

Nel **capitolo 4** vengono descritti ed analizzati i risultati dell'esperimento, soffermandosi su ogni singola variabile in gioco e sull'interazione tra variabili.

Il **capitolo 5**, infine, trae le conclusioni sul lavoro svolto e presenta le possibili direzioni di ricerca futura.

Nell'**appendice** viene riportato il codice MATLAB delle funzioni utilizzate.



# Capitolo 1

## Il suono nello spazio

Nel modello classico della teoria della comunicazione *sorgente-mezzo-ricevitore*, sono molti gli aspetti che vanno considerati al fine di permettere uno studio accurato della realtà ed una sua successiva simulazione precisa. Ci si propone, in questa sezione, di esaminare come l'onda sonora propagata nello spazio venga modificata dalle diverse parti anatomiche del ricevitore umano (testa, orecchio esterno e canale uditivo, busto e spalle). L'obiettivo è quello di ottenere una visione d'insieme chiara sui parametri percettivi coinvolti nello *spatial hearing*, per analizzare le tecniche di processing sonoro in 3-D, focalizzandoci in questo particolarmente sulle *tecniche binaurali*, che prevedono l'immissione nelle due orecchie di due distinti segnali sonori indipendenti. Viene fatta l'assunzione che suddetti segnali acustici contengano tutte le informazioni spaziali necessarie all'ascoltatore per elaborarne la percezione uditiva.

Le proprietà spaziali del campo sonoro, dopo l'azione di filtraggio subita dalle varie parti del corpo, vengono codificate come attributi temporali e spettrali della pressione acustica all'altezza del timpano. Si definisce innanzitutto il sistema di riferimento che useremo in questa descrizione.

### 1.1 Il sistema di riferimento

Per descrivere la posizione della sorgente rispetto all'ascoltatore, si individuano nello spazio tre piani, ortogonali tra loro:

- il piano *orizzontale xy*: è il piano su cui giacciono le due orecchie, determina la divisione sopra-sotto;
- il piano *frontale xz*: incide la testa nel senso del torso, definendo fronte-retro;
- il piano *mediano yz*: taglia in verticale all'altezza del naso, in modo equidistante dalle due orecchie, separando destra-sinistra.

Nel contesto di questa tesi, si utilizzerà il sistema di coordinate *verticale-polare*, preferito rispetto al sistema *interaurale-polare*. Li si descrive di seguito entrambi per completezza.

In entrambi i casi si definiscono la coordinata angolare verticale *elevazione* ( $\phi$ ), quella orizzontale *azimut* ( $\theta$ ) e la radiale *raggio* ( $r$ ). Quest'ultima rappresenta, in entrambi i sistemi di riferimento, la distanza euclidea tra la sorgente sonora e l'origine degli assi (i.e. centro della testa)[5].

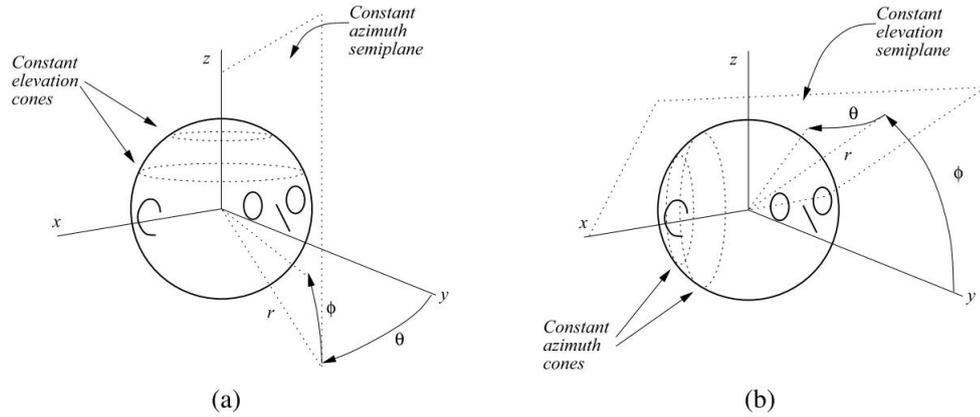


Figura 1.1: Sistemi di coordinate sferiche: (a) sistema di coordinate verticale-polare, e (b) sistema di coordinate interaurale-polare

### 1.1.1 Sistema di coordinate verticale-polare

Nel caso di sistema di coordinate *verticale-polare* (Figura 1.1(a)), l'azimut corrisponde all'angolo tra il piano mediano e il semipiano avente inizio sull'asse verticale  $z$  e contenente la sorgente sonora ( $\theta \in [-\pi; +\pi]$ ). Su questo stesso semipiano, l'elevazione corrisponde all'angolo tra il piano orizzontale ed il segmento che congiunge l'origine degli assi e la sorgente sonora ( $\phi \in [-\pi/2; +\pi/2]$ ).

### 1.1.2 Sistema di coordinate interaurale-polare

Nel sistema di riferimento *interaurale-polare* (Figura 1.1(b)), l'elevazione viene misurata come l'angolo tra il piano orizzontale ed il semipiano avente origine nell'asse interaurale  $x$  e contenente la sorgente sonora ( $\phi \in [-\pi; +\pi]$ ); l'azimut corrisponde all'angolo tra il piano mediano  $yz$  ed il segmento che congiunge origine degli assi e sorgente sonora ( $\theta \in [-\pi/2; +\pi/2]$ ).

## 1.2 La testa

La testa, ai lati della quale sono poste alla stessa altezza le due orecchie, costituisce il primo ostacolo alla libera propagazione del suono.

Essa produce infatti due effetti:

- introduce una differenza temporale tra i tempi d'arrivo del suono alle due orecchie, poichè l'onda sonora impiega più tempo a raggiungere l'orecchio più distante dalla sorgente (ITD, *interaural time difference*);
- causa una differenza di volume tra i suoni percepiti alle due orecchie, dovuto all'effetto di schermatura della testa stessa (ILD, *interaural level difference*).

### 1.2.1 ITD

Per lo studio dell'ITD, vengono fatte due ipotesi semplificative:

- la sorgente viene posta distante dalle orecchie che ricevono il suono, così che le onde che raggiungono la testa possano essere considerate piane;
- la testa viene approssimata ad una sfera e i canali uditivi sono posti in punti diametralmente opposti della sua superficie.

Queste assunzioni facilitano la descrizione geometrica della distanza extra  $\Delta x$  tra la sorgente e l'orecchio più distante.

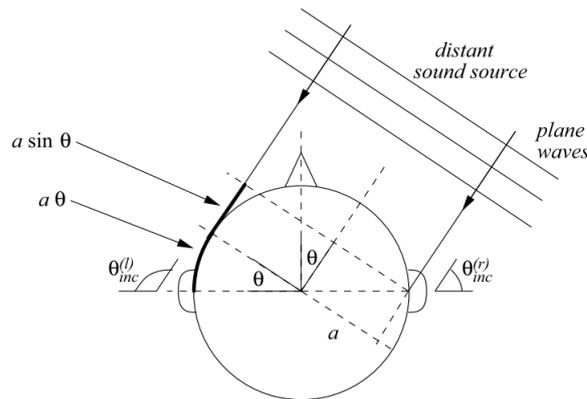


Figura 1.2: Stima dell'ITD in caso di sorgente sonora distante (onde piane) e testa sferica

Definiti  $a$  il raggio della testa,  $c$  la velocità del suono ( $\sim 343 \text{ m/s}$ ),  $\theta$  l'*angolo azimutale* che definisce la direzione dell'onda incidente sul piano orizzontale (Par. 1.1), si ottiene la formula semplice ma accurata per l'ITD:

$$ITD \sim \sin \frac{a}{c} (\theta + \sin \theta) \quad (1.1)$$

da cui è evidente che l'ITD è zero quando la sorgente è posta sul piano mediano ( $\theta = 0$ ) ed è invece massima se essa è posta ad un angolo  $\theta = \pm\pi/2$ , ovvero in corrispondenza di un orecchio.

Poichè la correlazione è debole, è lecito ammettere che l'ITD è indipendente dal parametro di frequenza dell'onda sonora.

### 1.2.2 ILD

Al contrario dell'ITD, l'ILD è fortemente dipendente dal parametro di frequenza: a basse frequenze, ovvero per lunghezze d'onda confrontabili col diametro della testa ( $\lambda \sim 2a$ ), non si nota quasi alcuna differenza tra la pressione sonora alle due orecchie, che diviene invece significativa ad alte frequenze.

Considerando le stesse ipotesi semplificative di cui sopra, la funzione di trasferimento di una testa sferica ideale di raggio  $a$ , alla cui distanza  $r$  è posta una sorgente sonora puntiforme (con  $r > a$ ), risulta [12]:

$$H_{sphere}(\rho, \mu, \theta_{inc}) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{+\infty} (2m+1) P_m(\cos \theta_{inc}) \frac{h_m(\mu\rho)}{h'_m(\mu)} \quad (1.2)$$

dove  $\theta_{inc}$  è l'angolo di incidenza dell'onda sonora sulla testa,  $\mu = \omega a/c$  la frequenza normalizzata,  $\rho = r/a$  la distanza normalizzata,  $P_m$  e  $h_m$  sono rispettivamente il polinomio di Legendre e la funzione sferica di Hankel di ordine  $m$ -simo.

Tale formula, che rappresenta la diffrazione di un'onda acustica su una sfera vista da un determinato punto sulla sfera stessa, è dipendente dalla frequenza. In particolare, per basse frequenze essa non è direzionalmente dipendente e la sua ampiezza  $|H_{sphere}|$  è identicamente unitaria  $\forall \theta_{inc}$ . Per alte frequenze (ovvero con  $\mu > 1$ ), la dipendenza da  $\theta_{inc}$  diventa evidente: nel caso di incidenza normale ( $\theta_{inc} = 0$ ) si misura un guadagno di 6 dB, consistente al caso di un'onda piana che incide perpendicolarmente una superficie rigida. All'aumentare di  $\theta_{inc}$  tale guadagno diminuisce, fino ad annullarsi intorno a  $\theta_{inc} = 170^\circ$  (e non a  $\theta_{inc} = \pi$  come ci si aspetterebbe, a causa del cosiddetto effetto *bright spot*, dovuto alla somma costruttiva delle onde propagate attorno alla sfera).

## 1.3 L'orecchio esterno

L'orecchio esterno è costituito dalla pinna (ovvero il padiglione auricolare) e dal canale uditivo, che si estende fino al timpano.

### 1.3.1 La pinna

La pinna, dalla forma caratteristica a basso-rilievo (Figura 1.3(a)), presenta molte differenze tra un individuo e l'altro. Essa costituisce l'ingresso del canale uditivo e la sua funzione è quella di antenna acustica: le sue cavità amplificano o attenuano la frequenza delle onde sonore che la colpiscono.

Dal punto di vista acustico, si comporta quindi come un filtro, la cui funzione di trasferimento è dipendente sia dalla distanza che dalla direzione relative tra la sorgente e l'orecchio.

### 1.3.2 Il canale uditivo

Il canale uditivo (Figura 1.3(b)) può, in prima approssimazione, essere visto come un tubo cilindrico di sezione costante. Come tale, esso funziona da risonatore unidimensio-

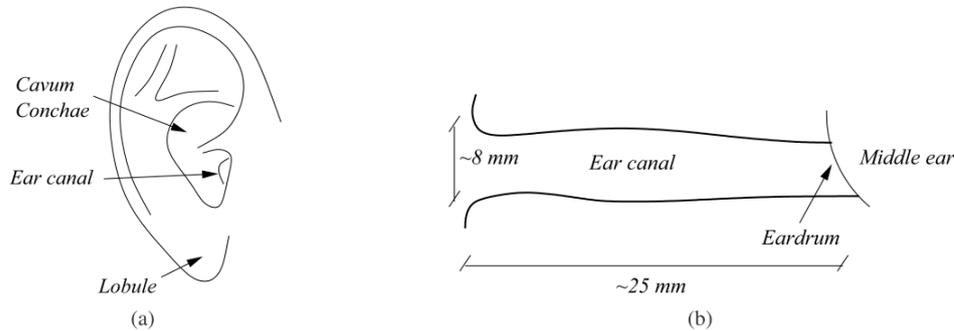


Figura 1.3: Orecchio esterno: (a) pinna, e (b) canale uditivo

nale, con pareti ad alta impedenza acustica, così da riflettere su di esso le onde sonore, limitandone l'assorbimento.

### 1.3.3 L'orecchio esterno come riflettore sonoro

Nel momento in cui un suono viene emesso dalla sorgente, esso arriva alla fine del canale uditivo, ovvero al timpano, come sovrapposizione dell'onda diretta e di onde riflesse. Nel suo tragitto, infatti, il suono si scontra con le pieghe della pinna e viene da esse deviato. Quando onda diretta e onda riflessa giungono in fase, l'intensità sonora aumenta a causa della risonanza (interferenza costruttiva). Nel caso in cui i segnali risultino invece fuori fase, l'intensità sonora diminuisce (interferenza distruttiva). L'intensità risulta minima quando la differenza tra i percorsi è pari a metà della lunghezza dell'onda: a questo fenomeno viene dato il nome di *pinna notch*, e corrisponde ad un filtro elimina banda. Elevazione ed azimut contribuiscono a determinare direzione e lunghezza dei segnali diretto e riflesso e pertanto sono parametri determinanti per lo studio del comportamento dell'orecchio esterno.

## 1.4 Il busto e le spalle

Il busto e le spalle contribuiscono ulteriormente alla colorazione del suono finale che arriva al timpano. Per semplicità descrittiva, si fa uso del modello *snowman* [3], costituito da un busto ellissoidale su cui poggia una testa sferica.

Busto e spalle hanno un duplice effetto di filtraggio, fungendo da ostacolo rigido:

- producono onde riflesse che si sommano al suono diretto (Figura 1.4(a))
- schermano i raggi sonori provenienti dal basso (Figura 1.4(b))

Tali effetti non sono della stessa entità di quelli introdotti dall'orecchio esterno. Essi vengono tuttavia tenuti in considerazione specialmente a basse frequenze, ovvero quando i segnali sonori presentano i valori massimi di energia e la risposta della pinna è essenzialmente piatta: per questo, in termini di frequenza, gli effetti di filtraggio di pinna e busto vengono considerati complementari.

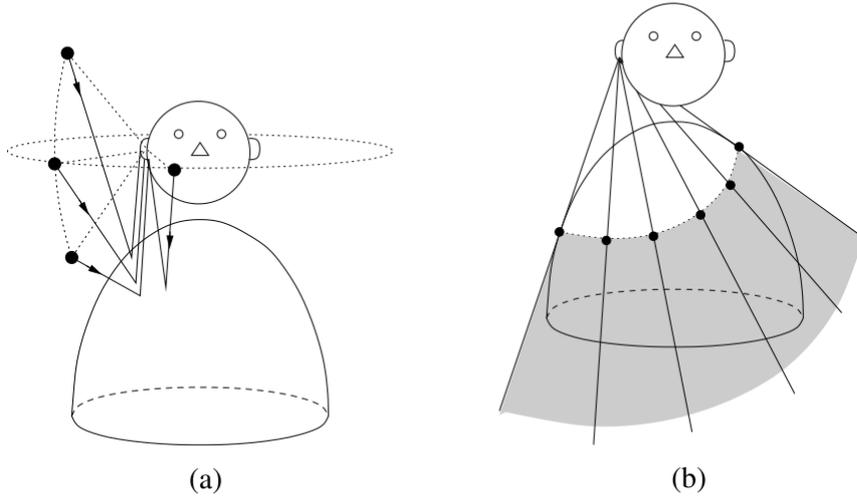


Figura 1.4: Effetti del busto: (a) riflessione, e (b) schermatura

## 1.5 Head-related transfer function

Sono stati descritti nelle sezioni precedenti gli effetti prodotti dalle diverse parti del corpo sull'onda sonora incidente. Essi sono caratterizzati dal fatto di essere tutti lineari, e questo permette di descriverli attraverso funzioni di trasferimento e di sovrapporli.

Nel dominio del tempo, siamo in grado di descrivere univocamente la pressione sul timpano di un suono emesso da una sorgente attraverso l'*Head Related Impulse Response (HRIR)*. Il suo corrispondente nel dominio della frequenza, ricavato attraverso la trasformata di Fourier, è la *Head Related Transfer Function (HRTF)*. Essa comprende tutti gli effetti a cui si è fatto riferimento sopra.

In generale, tale funzione di trasferimento si indica con  $H^{(l),(r)}(r, \theta, \phi, \omega)$  (dove l'apice indica l'orecchio che si sta prendendo in considerazione) e si può usare nella sua definizione il sistema di riferimento verticale-polare, o quello interaurale-polare (Paragrafo 1.1).

Nella maggior parte dei casi, si lavora per sorgenti nel *campo lontano* ( $r \rightarrow +\infty$ , ovvero  $r \gg 1$ ) ed è quindi lecito svincolare l'HRTF dalla dipendenza da  $r$  stesso e scrivere  $H^{(l),(r)}(\theta, \phi, \omega)$ .

Nell'ipotesi di una geometria perfettamente simmetrica, è inoltre possibile indicare la generica  $H(\theta, \phi, \omega)$ , dove  $H^{(r)}(\theta, \phi, \omega) = H(\theta, \phi, \omega)$  e  $H^{(l)}(\theta, \phi, \omega) = H(-\theta, \phi, \omega)$ .

Si definisce quindi la HRTF ad un orecchio come il rapporto, dipendente dal parametro della frequenza  $\omega$ , tra il livello di pressione del suono (*SPL*, *Sound Pressure Level*)  $\Phi^{(l),(r)}(\theta, \phi, \omega)$  sul timpano ed il corrispondente SPL  $\Phi_f(\omega)$  al centro della testa in campo libero, ovvero come se l'ascoltatore fosse assente.

Da cui:

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

A causa della diversa conformazione anatomica, la HRTF è una funzione propria di ogni ricevitore. Se ne può considerare una “media” che approssimi quella individuale con buona precisione.

## 1.6 Percezione della posizione di una sorgente sonora

Lo scopo di questa sezione è di individuare quali siano i parametri percettivi che creano nell'ascoltatore l'idea di dove si posizioni la sorgente sonora nello spazio.

Una volta chiariti, sarà possibile cercare un metodo per simulare la posizione della sorgente in uno spazio virtuale 3-D.

### 1.6.1 Internalizzazione ed esternalizzazione

Quando si analizza la posizione percepita di un suono riprodotto in stereo tramite delle cuffie, si può incorrere nel fenomeno di *internalizzazione*: la sensazione è che la sorgente si trovi all'interno della testa e che si sposti lungo l'asse interaurale (asse  $x$ ), sembrando più vicina ad un orecchio oppure all'altro (*lateralizzazione*).

In particolare, la sorgente verrà percepita come al centro della testa se vengono emessi due identici suoni monoaurali da entrambe le cuffie. Il suono viene invece percepito lateralizzato se si mutano i parametri di ITD ed ILD (Sezioni 1.2.1, 1.2.2), fino ad avere l'impressione che la sorgente si trovi in prossimità di un orecchio, sempre dentro la testa. Tale effetto è denominato IHL (*inside-the-head localization*).

Rimuovendo l'effetto IHL, si riesce a far percepire la sorgente come esterna all'ascoltatore (*esternalizzazione*) e il processo di individuazione della posizione è detto *localizzazione*. Si è notato che l'aggiunta di un riverbero, sia esso naturale o artificiale, migliora decisamente l'esternalizzazione.

### 1.6.2 Percezione azimutale

Gli effetti prodotti dalla presenza della testa nello spazio (Sezione 1.2), ovvero l'ITD e l'ILD, sono i principali responsabili della percezione sul piano orizzontale.

Per studiare questi fenomeni, la *Teoria Duplex* divide lo spettro delle frequenze in intervalli, rendendo complementari i contributi di ITD ed ILD[30]:

- per frequenze  $\omega < 800$  Hz, si considerano solamente gli effetti temporali interaurali (ITD), dal momento che la funzione di trasferimento dell'ILD diventa piatta.

In questo intervallo, che corrisponde alla relazione  $\lambda/2 > 2a$  (con  $\lambda$ =lunghezza dell'onda sonora,  $a$ =raggio della testa), il sistema uditivo è in grado di determinare il ritardo di fase tra le due orecchie senza ambiguità.

Per frequenze inferiori agli 80 Hz circa, tuttavia, non si è più in grado di ricostruire l'informazione azimutale, risultando inutile anche il contributo dell'ITD.

- per frequenze  $\omega > 1600$  Hz, prevalgono le differenze di livello interaurale (ILD): non si è più in grado, infatti, di rilevare quale sia il fronte d'onda principale dai suoi tempi di ritardo.

Esiste tuttavia, se l'involuppo dell'onda non ha ancora una frequenza troppo elevata, una stima dell'ITD come misura dello sfasamento tra l'involuppo delle due onde: in questo caso si parla più opportunamente di *Interaural Envelope Difference (IED)*.

Si introduce qui, infine, un nuovo fenomeno percettivo basato sulla capacità del sistema uditivo di estrarre l'informazione azimutale nell'intervallo di tempo che trascorre tra l'arrivo dell'onda diretta e quello delle onde riflesse. Tale effetto "di precedenza" è particolarmente presente in ambienti molto riverberanti.

- le frequenze intermedie costituiscono una zona di transizione, dove si osservano contributi da entrambi i fenomeni.

La Teoria Duplex è formalizzata per ambienti anecoici, ovvero privi di eco. Questo, in aggiunta alle ipotesi semplificative sulla testa (la testa umana non è in realtà perfettamente sferica, esistono asimmetrie e componenti facciali aggiuntivi, le orecchie non giacciono esattamente sull'asse  $x$ ), non permette di descrivere esattamente la situazione reale.

Ciò che si ottiene è, tuttavia, una buona approssimazione. Pur non godendo della situazione ideale, infatti, si osserva che i valori di ITD ed ILD per coppie di sorgenti poste rispettivamente a  $+\theta$  e  $-\theta$  sono pressochè identici. Tale fenomeno è significativo e causa, nell'ascoltatore, la cosiddetta "*front/back confusion*" portando ad una totale inversione della percezione azimutale.

### 1.6.3 Percezione dell'elevazione

La definizione dell'elevazione di una sorgente rispetto al piano orizzontale risulta più difficile, se confrontata con la percezione azimutale.

Una sorgente sonora posta in qualsiasi punto di una superficie conica avente per vertice l'orecchio di una testa sferica produce, infatti, identici valori di ITD ed ILD, come mostrato in figura 1.5. Tale superficie, detta *cono di confusione*, estende il fenomeno di front/back confusion descritto sopra.

La pinna, grazie alle pieghe tipiche della sua conformazione, è fondamentale per migliorare la localizzazione verticale, producendo picchi e tagliando bande di frequenza diversi a seconda della posizione della sorgente.

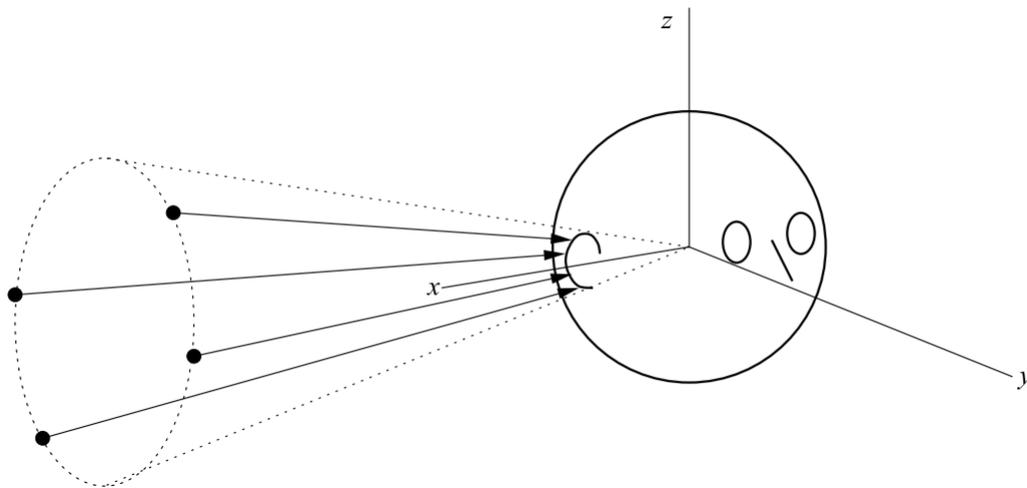


Figura 1.5: Cono di confusione

Si osserva che, semplicemente ripiegando od occludendo i padiglioni, è molto più difficile stabilire l'altezza da cui proviene il suono [18].

#### 1.6.4 Percezione della distanza

Ancor più complicata è l'esatta localizzazione della sorgente dal punto di vista della distanza. In tale processo, infatti, intervengono molti fattori ed il contributo di alcuni di essi può venir annullato dalla somma di altri.

Esaminiamone di seguito i principali:

- esiste innanzitutto un'ovvia correlazione tra la distanza e l'*intensità*. In condizioni prive di eco, il guadagno di intensità sonora  $g$  è inversamente proporzionale alla distanza relativa  $d$  sorgente-ricevitore:

$$g = \frac{1}{d^2}$$

e, per una sorgente omnidirezionale, il guadagno diminuisce di circa 6dB al raddoppiare della distanza.

Si definisce *loudness* l'intensità percepita dal sistema uditivo, che segue pertanto una scala diversa dall'intensità stessa.

- la *familiarità* dell'ascoltatore col suono percepito contribuisce a costruirne l'immagine spaziale. Infatti, se la sorgente che emette quel determinato suono è abitualmente associata ad una certa distanza nota al ricevente, sarà più facile localizzarla.
- nel caso in cui il suono sia arricchito da *riverbero*, il rapporto di proporzionalità inversa sorgente-ricevitore visto per l'onda diretta non si applica alle onde riflesse,

rendendo impossibile basarsi unicamente su di esso per distinguere la distanza relativa di due sorgenti.

Si introduce il rapporto tra energia riflessa ed energia diretta (*R/D ratio*), che sembra invece contribuire nel fornire una diversa impressione tra due sorgenti a differenti distanze dall'ascoltatore.

Si osserva inoltre che, in totale mancanza di riverbero, si è portati a sottostimare la distanza apparente della sorgente[16].

- esistono inoltre determinati *effetti spettrali*, su cui ci si basa per definire la distanza della sorgente. Nel campo vicino (ovvero per distanze intorno ad 1.5 m dall'ascoltatore), ad esempio, a mano a mano che la sorgente si avvicina, il suono si arricchisce alle basse frequenze, creando la sensazione di un tono più scuro. Vanno inoltre considerate, seppur più deboli, le condizioni atmosferiche ed il coefficiente di assorbimento dell'aria, che a loro volta influenzano la percezione della distanza.
- sembra esistere inoltre una qualche correlazione tra l'azimut a cui è posta la sorgente e la capacità del ricevitore umano di poterne ricostruire la distanza: esperimenti dimostrerebbero, infatti, che è più facile riconoscere la distanza di un suono proveniente da una sorgente laterale ( $\theta \sim \pm\pi/2$ ), piuttosto che da una posta sul piano mediano (*auditory parallax*). La questione è tuttavia ancora in fase di studio[17].

## 1.7 Algoritmi per il rendering 3-D del suono

Nello studio portato avanti fino ad ora, è stato fatto cenno all'audio binaurale ed alla riproduzione di un suono tramite delle cuffie.

In questa sezione, illustriamo le basi del rendering spaziale ed approfondiamo questi concetti.

### 1.7.1 Rendering spaziale del suono

Con rendering spaziale si intende la capacità di riprodurre un suono con un sistema di altoparlanti o cuffie, affinché l'ascoltatore abbia impressioni percettive quanto più verosimili possibile. Esso consiste quindi nel riuscire a trasmettere ai timpani dell'ascoltatore un segnale di pressione compatibile a quello realmente prodotto dalla sorgente.

Un suono emesso da una sorgente, come si è visto, arriva in maniera diversa al timpano destro e a quello sinistro. Si crea allora la necessità di costruire almeno due distinti segnali sonori, ciascuno dedicato ad un orecchio (*audio binaurale*), che contengano tutte le informazioni necessarie a stabilire la posizione nello spazio della sorgente simulata. Tale processo è detto *virtualizzazione*.

### 1.7.2 Tecnologie

Per creare il segnale sonoro ottimo, è necessario conoscere il mezzo attraverso cui esso verrà trasmesso alle orecchie.

Le tecnologie utilizzate sono principalmente:

- impianti di tipo *stereo*, che prevedono il posizionamento di due casse ai lati dell'ascoltatore. Se i suoni da esse trasmessi sono in fase, si percepisce la sorgente come posta all'interno del segmento avente per estremi le due casse.
- *sistemi multicanale*, nel quale uno o più canali audio (detti canali surround) vengono riprodotti da diffusori acustici posizionati intorno all'ascoltatore. A ciascun canale è associata una direzione del suono, eccetto una cassa (subwoofer), dedicata invece alla diffusione delle basse frequenze.
- strumenti quali *cuffie* o *auricolari* sfruttano le tecniche dell'*audio binaurale* diffondendo i suoni direttamente nell'orecchio, permettendo facilmente di associarvi il segnale dedicato (destra/sinistra). Si ha inoltre il vantaggio dell'eliminazione del riverbero ambientale, impensabile altrimenti coi sistemi che sfruttano l'uso di casse. Oltretutto, i sistemi con altoparlanti soffrono del fenomeno *cross-talk*, che implica che un suono emesso da una cassa sola raggiunga comunque entrambe le orecchie. Esso è eliminabile con apposite tecniche dette di *cross-talk cancellation*, che intervengono a livello di produzione.

L'uso delle cuffie, tuttavia, presenta anche alcuni svantaggi, come per esempio la perdita della conduzione, da parte del corpo (sistema osseo, cassa toracica, scatola cranica), delle sensazioni legate alle basse frequenze. Inoltre, cuffie ed auricolari hanno una propria risposta in frequenza, che richiede apposite tecniche di compensazione per non influenzare la riproduzione, e possono risultare scomode da indossare per periodi prolungati.

Ciò che risulta determinante nell'uso di questi strumenti è la capacità di porre la sorgente virtuale praticamente ovunque nello spazio (superando così i limiti degli impianti stereo), consentendo anche la simulazione di un'elevazione della sorgente stessa.

### 1.7.3 Rendering basato sulle HRTF

Come visto nella sezione 1.5, la Head Related Transfer Function comprende tutti i parametri di riflessione, diffrazione, risonanza e assorbimento che riguardano il suono emesso da una sorgente nel campo libero e diretto verso il timpano.

Un'accurata conoscenza di tali parametri permette, attraverso la loro manipolazione, il corretto posizionamento di una sorgente virtuale nello spazio e la precisa simulazione di un ambiente sonoro.

È infatti sufficiente ricreare la corretta pressione sonora all'altezza del timpano per rendere uno scenario sonoro sintetico indistinguibile da uno reale[33]. L'idea di partenza è di avere a disposizione un numero sufficiente di misurazioni di HRTF o di risultati

derivanti da modelli della stessa, così da poter aggiungere al suono simulato gli effetti di filtraggio tipici della situazione reale.

Quando l'uso di queste HRTF viene combinato con dispositivi di *head-tracking*, si rende possibile la creazione di uno scenario audio real-time interattivo, in grado di simulare sorgenti sonore in qualsiasi posizione rispetto all'ascoltatore. Questi display audio virtuali trovano importanti applicazioni nell'interfaccia uomo-macchina e nella creazione di ambienti virtuali realistici[10].

### Uso di database di HRTF

Nella figura 1.6 è illustrato, tramite uno schema a blocchi, il procedimento basato sull'uso di database di HRTF, che approfondiamo di seguito.

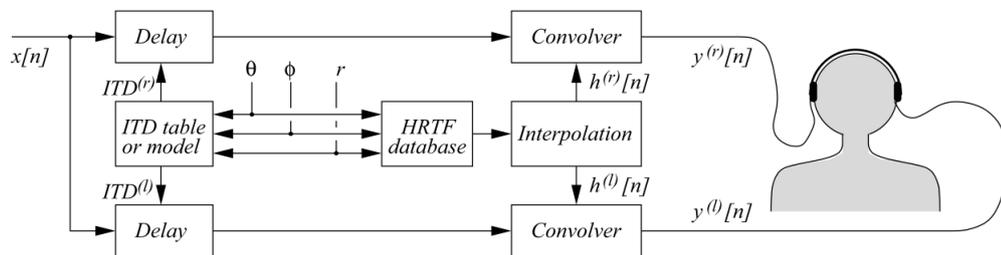


Figura 1.6: Schema a blocchi di un sistema di rendering 3-D del suono tramite cuffie basato sulle HRTF.

Nel caso in cui si intendano utilizzare *misurazioni reali* di HRTF, il procedimento classico prevede di porre all'interno di una camera anecoica un ascoltatore, in posizione fissata, intorno al quale è posto un sistema di altoparlanti. Nelle orecchie dell'ascoltatore vengono posti dei microfoni, all'altezza del timpano oppure all'inizio del canale uditivo, così da registrare come l'impulso emesso dagli amplificatori raggiunga il ricevitore umano. Le casse sono inoltre posizionate a distanze regolari su una struttura di cui l'ascoltatore è il centro, e per ciascuna è quindi nota distanza  $r$ , elevazione  $\phi$  ed azimut  $\theta$ .

Le risposte ai microfoni dell'impulso emesso dalla sorgente sonora costituiscono il database di HRTF (rispettivamente destro e sinistro) per quel singolo ascoltatore.

Il processo sopra descritto, tuttavia, è evidentemente laborioso ed impossibile da effettuarsi per ciascun potenziale soggetto (si pensi, ad esempio, alle necessità del settore commerciale di diffusione audio di fornire buoni risultati a tutti i possibili clienti). È necessario, quindi, costruire un database di HRTF che approssimi bene le caratteristiche dell'ascoltatore tipico (*HRTF generalizzata*).

Una prima soluzione è quella di fare la media tra le risposte ai microfoni di un certo numero di individui.

Un'alternativa spesso utilizzata è, altrimenti, quella che prevede la misurazione dell'HRTF su un manichino antropomorfo (*dummy head*). Quello usato più frequentemente a tal fine è il *KEMAR* (*Knowles Electronic Manikin for Auditory Research*, figura 1.7), che contiene nella sua struttura i microfoni necessari alla misurazione. Le sue forme sono tali da simulare quelle di una testa umana “media”, e per questo i suoi risultati sono più o meno verosimili a seconda delle misure dell'ascoltatore finale. L'uso del KEMAR ha sicuramente il vantaggio di rendere facilmente riproducibili i risultati e di evitare lo scomodo posizionamento dei microfoni nelle orecchie, processo in cui ogni minimo spostamento può essere fonte d'errore per le misurazioni.



Figura 1.7: Profilo ed interno di un manichino KEMAR.

Si ottiene in questo modo una prima idea di database delle risposte alle due orecchie, che viene poi affinata attraverso definiti passi di post-processing.

Dato un segnale anecoico e la posizione desiderata della sorgente virtuale  $(r, \phi, \theta)$ , si definiscono i segnali destro e sinistro:

1. introducendo tra essi un ritardo, così da creare l'ITD desiderato,
2. convolvendo i segnali ottenuti con le risposte all'impulso per orecchio destro/sinistro contenute nel database di HRTF, procedendo eventualmente ad un'*interpolazione* nel caso in cui la posizione desiderata non sia tra quelle presenti nel database.

I suoni così ottenuti vengono infine trasmessi ai canali corrispondenti tramite cuffie, fornendo all'ascoltatore l'idea spaziale desiderata.

Tale procedimento ha, tuttavia, un costo elevato in termini di tempi e carichi computazionali, specialmente in condizioni di rendering dinamiche.

Un'alternativa possibile è considerare la risposta di *modelli matematici di HRTF*, ottenendo così, oltre a migliori prestazioni, un segnale privo di rumore. Studi condotti dimostrerebbero infatti come, nel caso di HRTF misurate, il *Signal-to-Noise Ratio* (SNR) sia funzione della posizione relativa ascoltatore-sorgente[20].

### Modello strutturale

Un'alternativa all'uso delle HRTF è la combinazione degli effetti di filtraggio delle singole parti del corpo (busto, testa, pinna), modellate separatamente e poi unite a creare una HRTF equivalente.

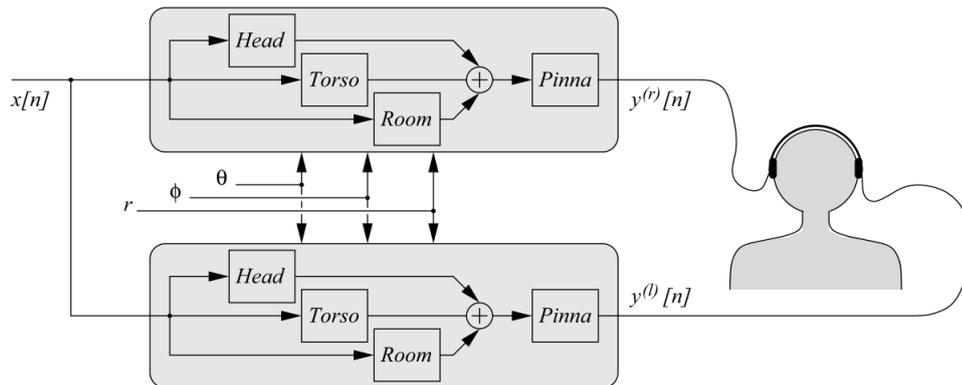


Figura 1.8: Schema a blocchi di un sistema di rendering 3-D del suono tramite cuffie basato su un modello strutturale.

In figura 1.8 viene schematizzato, attraverso uno schema a blocchi, il processo base di elaborazione del segnale finale: il segnale di partenza viene innanzitutto filtrato attraverso i modelli di testa, busto e, eventualmente, stanza. I segnali così ottenuti vengono sommati e quindi nuovamente filtrati attraverso il modello della pinna.

Per la testa si utilizza un filtro  $\tilde{H}_{sphere}(\rho, \theta_{inc}, \mu)$  che approssima la funzione di trasferimento della testa sferica  $H_{sphere}(\rho, \theta_{inc}, \mu)$  di cui a 1.2.2. Tale modello è affinabile considerando una testa ellissoidale e riconoscendo alle orecchie la loro posizione naturale, leggermente posteriore ed abbassata rispetto al diametro della sfera.

Per busto e pinna si considerano principalmente i contributi dati dalle riflessioni da essi prodotte, facendo per il busto attenzione alla presenza di eventuali effetti di schermatura (Sezione 1.4) e tendendo a descrivere le riflessioni multiple della pinna attraverso un set di filtri complementari.

Nello schema di figura 1.8, il filtraggio della pinna è posto come ultimo passo di processo a causa del fatto che i suoni che raggiungono l'orecchio possono aver seguito due percorsi diversi e per questo anteposti: la diffrazione attorno alla testa o la riflessione sul busto. La posizione desiderata della sorgente viene fissata durante tutti questi passi, agendo opportunamente sui parametri  $r, \phi, \theta$ .

Svolgendo questo procedimento sia per il lato destro che per il sinistro, si ottengono i due segnali da trasmettere in cuffia.

### Metodi a confronto

Confrontando i due approcci, si evidenzia come, rispetto all'uso di una HRTF generalizzata (che risulta il metodo più economico), il modello strutturale sia, compiendone una corretta parametrizzazione, più precisamente adattabile all'ascoltatore[7]. Si è verificato che l'uso di HRTF non individuali, infatti, aumenta il rischio di front/back confusion e fornisce risultati meno precisi sul piano mediano in condizioni statiche, aumentando quindi la probabilità di errore sulla corretta posizione della sorgente sonora ([29, 6]).

Il metodo basato sul modello strutturale ha, inoltre, il vantaggio di permettere un efficiente *rendering real-time*, ovvero nel caso in cui il soggetto, munito di cuffie, sia libero di muoversi nello spazio, mantenendo fissa la sorgente virtuale nel punto scelto. I metodi matematici utilizzati per approssimare l'HRTF (funzioni razionali polo/zero o espansioni in serie di componenti principali o di armoniche[15]) sono infatti, in confronto, troppo dispendiosi per poter essere velocemente adattati alla situazione.

L'unico modo comunque per validare un modello, sia esso basato sull'HRTF misurata o di cui siano stati ricostruiti gli effetti, è attraverso accurati test psicoacustici, come vedremo nel Capitolo 3.



## Capitolo 2

# Un modello di HRTF nel campo vicino

Lo scopo di questo capitolo è presentare il modello di sorgente sonora virtuale nel campo vicino di cui si andranno a valutare, nei capitoli seguenti, le prestazioni dal punto di vista della distanza relativa tra due punti.

Nello scenario di *Virtual Auditory Space* (VAS), si presentano nell'ordine il concetto di campo vicino (accennato nel paragrafo 1.6.4), la necessità di creare un modello *ad hoc* e la validazione di quest'ultimo dal punto di vista teorico.

### 2.1 Far-field e Near-field

La maggior parte degli studi condotti finora sulla virtualizzazione si è concentrata sul posizionamento di sorgenti sonore relativamente distanti, per lo più nel cosiddetto campo lontano (*far-field*, o regione distale), ovvero ad oltre 1.5 m dal centro della testa.

Nel caso di una sorgente sonora puntiforme, in questo spazio l'HRTF è considerata sostanzialmente indipendente dalla distanza, a meno di un fattore di intensità costante che decresce secondo la legge quadratica  $1/d^2$ , con  $d$ =distanza relativa sorgente-ricevitore. Essa varia invece sensibilmente per sorgenti poste nel campo vicino (*near-field*, o regione prossimale).

Gli studi di HRTF nel campo vicino condotti da Brungart e Rabinowitz [8, 9], tramite misurazioni acustiche su un manichino antropomorfo, si distinguono dai precedenti [27, 28], che utilizzavano invece un modello di testa a sfera rigida. Quest'ultimo si rivela essere, infatti, un'approssimazione non sufficientemente precisa ai fini di comprendere cosa accade nel campo vicino, non includendo le irregolarità della testa che diventano significative in quest'analisi.

Gli esperimenti psicoacustici di sound localization di Brungart dimostrano come un parametro significativo per la determinazione della distanza di una sorgente sonora posta nel campo vicino sia l'ILD (Interaural Level Difference), che aumenta al diminuire della distanza della sorgente dalla testa. Tale tesi è avvalorata da risultati positivi, per quanto

poco accurati, sulla capacità di discriminare tra due suoni posti a distanza diversa, nella cui virtualizzazione sia stata esclusa l'informazione di intensità sonora, considerata sempre la predominante ai fini della localizzazione.

Si tratta di individuare quindi un modello che parametrizzi correttamente l'HRTF, rendendo possibile una descrizione accurata di quanto avviene nel campo vicino al variare della distanza della sorgente dall'ascoltatore.

## 2.2 Sintesi dell'HRTF

Sotto l'ipotesi di una sorgente infinitamente distante dal centro della testa e considerata quest'ultima in prima approssimazione di forma sferica, la risposta relativa ad un punto fisso sulla superficie della sfera è ben descritta dalla seguente funzione di trasferimento, basata sulla formula di diffrazione di Lord Rayleigh:

$$H(\mu, \theta_{inc}) = \frac{1}{\mu^2} \sum_{m=0}^{+\infty} \frac{(-i)^{m-1} (2m+1) P_m(\cos \theta_{inc})}{h'_m(\mu)} \quad (2.1)$$

dove  $\theta_{inc}$  è l'angolo di incidenza dell'onda sonora sulla testa,  $P_m$  e  $h'_m$  sono rispettivamente il polinomio di Legendre e la derivata del primo ordine rispetto al suo argomento della funzione sferica di Hankel di ordine  $m$ -simo,  $\mu = \omega a/c$  la frequenza normalizzata.

Il modello strutturale proposto da Brown e Duda in [7] fornisce un'approssimazione di questa HRTF attraverso un filtro a singolo-polo e singolo-zero, posto in cascata ad un elemento di ritardo che descrive la differenza temporale tra il momento in cui l'onda sonora raggiunge il punto di osservazione e quello in cui raggiungerebbe il centro della testa sferica in condizione di campo libero.

Tale modello di testa sferica non include tuttavia una parametrizzazione sulla distanza, rendendo impossibile il suo utilizzo in condizioni scalate nel campo vicino.

Diverso invece è stato l'approccio utilizzato da Duda e Martens in [12], che, tramite una modifica dell'algoritmo classico per l'HRTF di una sorgente posta a distanza infinita, permette di porre la sorgente ad una distanza arbitraria dalla testa sferica.

Tale soluzione si rivela però troppo pesante dal punto di vista computazionale e quindi poco adatta in condizioni di riproduzione point-to-point real-time.

Per ottenere un'approssimazione che includa la distanza e permetta una più veloce computazione, viene introdotta quindi la STF (Spherical Transfer Function) di cui all'equazione (1.2):

$$H_{sphere}(\rho, \mu, \theta_{inc}) = -\frac{\rho}{\mu} e^{-i\mu\rho} \sum_{m=0}^{+\infty} (2m+1) P_m(\cos \theta_{inc}) \frac{h_m(\mu\rho)}{h'_m(\mu)} \quad (2.2)$$

dove  $\rho = r/a$  è la distanza normalizzata rispetto ad  $a$  raggio della testa.

In figura 2.1 vengono mostrati due esempi di moduli di STF, rispettivamente nel campo vicino e nel campo lontano, calcolati per 19 angoli di incidenza ( $\theta_{inc} = [0^\circ, 180^\circ]$  ad intervalli di 10 gradi).

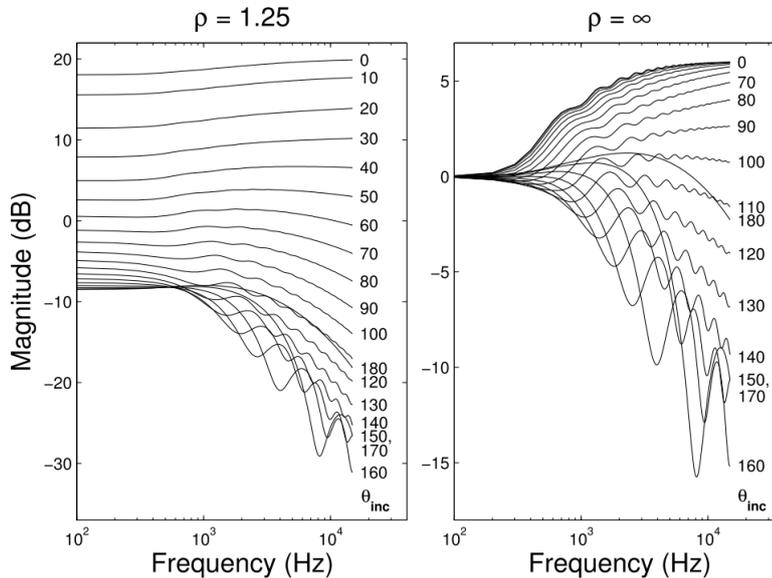


Figura 2.1: STF calcolata per  $\rho = 1.25$  (*near-field*) e  $\rho \rightarrow +\infty$  (*far-field*).

Spagnol e Avanzini in [23] hanno analizzato l'andamento della STF utilizzando l'analisi in componenti principali (PCA), facendo emergere che tale funzione è fortemente dipendente dall'angolo di incidenza e, solo in misura più attenuata, dalla distanza: la STF non si rivela pertanto ancora sufficiente ad isolare l'informazione di distanza, primo obiettivo in questo nostro studio del campo vicino.

In [24], Spagnol *et al.* propongono una normalizzazione della STF che tiene conto della distanza rispetto alla corrispondente risposta sferica nel campo lontano, ottenendo una nuova funzione di trasferimento, denominata Near-Field Transfer Function (NFTF):

$$H_{NF}(\rho, \mu, \theta_{inc}) = \frac{H_{sphere}(\rho, \mu, \theta_{inc})}{H_{sphere}(\infty, \mu, \theta_{inc})} \quad (2.3)$$

Si ottiene così una funzione con andamento pressoché monotono con la frequenza, che risulta meno frastagliata rispetto alla STF.

Nella prossima sezione, si riporta lo studio che valida la NFTF dal punto di vista teorico, dimostrando la sua effettiva dipendenza dal parametro della distanza.

### 2.3 Validazione teorica della NTF

Nonostante una funzione di trasferimento sferica non includa precisamente tutte le interazioni dovute all'anatomia della testa umana (come discusso in precedenza), studi presenti in letteratura hanno attestato la sua idoneità per un'analisi nel campo vicino, specialmente a basse frequenze[31].

Nell'analisi che segue, si fissa il raggio della testa  $a=8.75\text{cm}$ , che rappresenta il valore medio per una testa umana adulta, mantenendo la possibilità eventualmente di scarlo, modificando uniformemente i valori sull'asse delle frequenze.

#### 2.3.1 Comportamento in continua

Come primo obiettivo dell'analisi della NTF ci si propone di studiare il comportamento del guadagno DC ( $\mu=0$ ) al variare della posizione della sorgente lungo una direzione azimutale fissata.

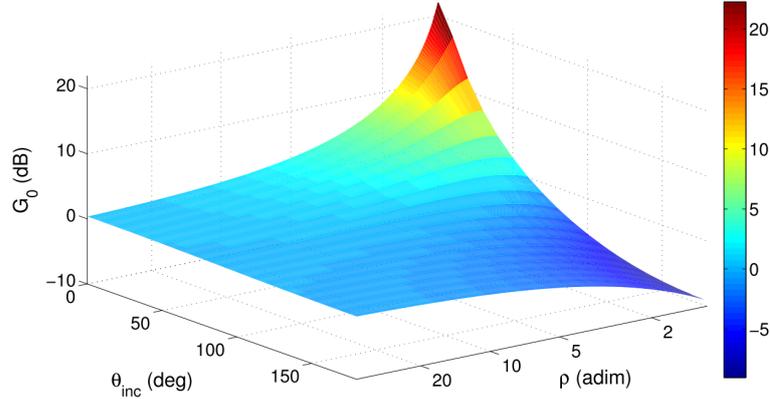


Figura 2.2: Guadagno DC della NTF.

Il guadagno DC in dB  $G_0(\theta_{inc}, \rho)$  si calcola per ciascuno dei 19 angoli di incidenza di cui in figura 2.1 ( $\theta_{inc} = [0^\circ, 180^\circ]$  ad intervalli di 10 gradi), prendendo in considerazione 250 valori di distanza  $\rho$  crescenti in modo esponenziale secondo:

$$\rho = 1.15^{1+\frac{k-1}{10}}, \quad k \in \mathbb{N}, k = 1, \dots, 250 \quad (2.4)$$

In figura 2.2 è riportato il risultato in funzione della distanza e dell'angolo di incidenza: l'andamento di  $G_0(\theta_{inc}, \rho)$  è prevalentemente esponenziale.

Per modellare la dipendenza dalla distanza della NTF in DC si approssima  $G_0(\theta_{inc}, \rho)$  ad una funzione razionale del secondo ordine per i 19 distinti angoli di incidenza, denominata  $\tilde{G}_0(\theta_{inc}, \rho)$  e definita come:

$$\tilde{G}_0(\theta_{inc}, \rho) = \frac{p_{11}(\theta_{inc})\rho + p_{21}(\theta_{inc})}{\rho^2 + q_{11}(\theta_{inc})\rho + q_{21}(\theta_{inc})} \quad (2.5)$$

dove i coefficienti  $p_{11}(\theta_{inc})$ ,  $p_{21}(\theta_{inc})$ ,  $q_{11}(\theta_{inc})$ ,  $q_{21}(\theta_{inc})$  per ciascuno dei 19 angoli di incidenza sono riportati nella tabella 2.1. Per gli angoli intermedi i coefficienti sono stati calcolati utilizzando un'interpolazione lineare.

$\theta_{inc}$	$p_{11}$	$p_{21}$	$q_{11}$	$q_{21}$	RMS [dB]
0°	12.97	-9.691	-1.136	0.219	0.0027
10°	13.19	234.2	18.48	-8.498	0.0223
20°	12.13	-11.17	-1.249	0.346	0.0055
30°	11.19	-9.035	-1.017	0.336	0.0034
40°	9.91	-7.866	-0.83	0.379	0.002
50°	8.328	-7.416	-0.666	0.421	0.0009
60°	6.493	-7.312	-0.503	0.423	0.0002
70°	4.455	-7.278	-0.321	0.382	0.0004
80°	2.274	-7.291	-0.11	0.314	0.0005
90°	0.018	-7.484	-0.13	0.24	0.0005
100°	-2.242	-8.04	0.395	0.177	0.0004
110°	-4.433	-9.231	0.699	0.132	0.0003
120°	-6.488	-11.61	1.084	0.113	0.0002
130°	-8.342	-17.38	1.757	0.142	0.0002
140°	-9.93	-48.42	4.764	0.462	0.0004
150°	-11.29	9.149	-0.644	-0.138	0.0006
160°	-12.22	1.905	0.109	-0.082	0.0003
170°	-12.81	-0.748	0.386	-0.058	0.0002
180°	-13	-1.32	0.45	-0.055	0.0002

Tabella 2.1: Coefficienti per l'equazione di  $\tilde{G}_0$  (2.5) ed errore RMS.

I valori dell'errore RMS (root mean square), che indica quanto  $\tilde{G}_0(\theta_{inc}, \rho)$  si discosti dal guadagno reale  $G_0(\theta_{inc}, \rho)$ , confermano la bontà dell'approssimazione: praticamente in tutti i casi,  $RMS(\tilde{G}_0, G_0) < 0.01$  dB.

### 2.3.2 Comportamento in frequenza

Conclusa l'analisi in DC, si passa quindi all'analisi di come e quanto l'NFTF dipenda dalla frequenza e di come tale dipendenza possa essere efficacemente resa.

A tal fine, si utilizza il guadagno DC reale  $G_0(\theta_{inc}, \rho)$  come fattore di normalizzazione per la NFTF  $H_{NF}(\rho, \mu, \theta_{inc})$ , ottenendo:

$$\hat{H}_{NF}(\rho, \mu, \theta_{inc}) = \frac{H_{NF}(\rho, \mu, \theta_{inc})}{G_0(\theta_{inc}, \rho)} \quad (2.6)$$

L'operazione per ottenere la NFTF normalizzata  $\hat{H}_{NF}(\rho, \mu, \theta_{inc})$  viene eseguita sui 250 valori di distanza  $\rho$  già considerati e in frequenze fino a 15 kHz, campionate ad intervalli di 10 Hz.

La figura 2.3 mostra il comportamento di  $\hat{H}_{NF}$  per una distanza fissata ( $\rho = 1.25$ ), al variare dell'angolo di incidenza  $\theta_{inc}$ .

Si osserva che il modulo della NFTF assume sempre valori negativi, raggiungendo come valore massimo 0 dB: questo significa che il guadagno DC rappresenta il massimo per ciascun valore dell'angolo di incidenza.

Analizzandone l'andamento più nel dettaglio, esso è riconducibile a quello di un filtro shelving per gli angoli di incidenza ipsilaterali (ovvero prossimi a 0°) e ad un filtro passa basso per i controlaterali (prossimi a 180°). Modellare tale comportamento tramite

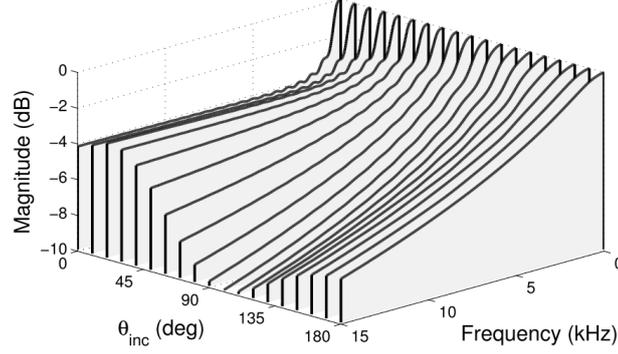


Figura 2.3: Comportamento in frequenza della NFTF normalizzata  $\hat{H}_{NF}$  per  $\rho = 1.25$ .

l'uso di un filtro shelving e di uno passa basso non è consigliato, dal momento che i valori intermedi della frequenza, se non opportunamente trattati, farebbero comparire distorsioni acustiche; inoltre, un filtro passa basso introdurrebbe un'attenuazione troppo elevata per le alte frequenze.

A costo di una perdita di precisione, la soluzione migliore è approssimare la NFTF normalizzata con un filtro shelving del primo ordine ad alta frequenza, secondo l'implementazione proposta in [32]:

$$H_{sh}(z) = 1 + \frac{H_0}{2} \left( 1 - \frac{z^{-1} + a_c}{1 + a_c z^{-1}} \right), \quad (2.7)$$

con

$$a_c = \frac{V_0 \tan(\pi \frac{f_c}{f_s}) - 1}{V_0 \tan(\pi \frac{f_c}{f_s}) + 1}, \quad V_0 = 10^{\frac{G_\infty}{20}}, \quad H_0 = V_0 - 1, \quad (2.8)$$

dove  $f_s$  è la frequenza di campionamento, e i parametri  $f_c$  e  $G_\infty$  (rispettivamente frequenza di taglio e guadagno asintotico in alta frequenza) sono ricavabili da  $\hat{H}_{NF}$ :  $f_c$  è calcolata come la frequenza in corrispondenza della quale  $\hat{H}_{NF}$  possiede un guadagno in dB negativo che approssima i due terzi del guadagno in alta frequenza,  $G_\infty$  corrisponde al guadagno in dB a 15 kHz.

La figura 2.4 mostra la buona qualità dell'approssimazione per mezzo del filtro shelving per i 19 angoli di incidenza a tre diverse distanze ( $\rho = 1.25$ ,  $\rho = 4$ ,  $\rho = 16$ ).

Gli andamenti di  $G_\infty$  e  $f_c$  sono stati a loro volta approssimati con quelli funzioni razionali del secondo ordine, come visto per  $\tilde{G}_0$ :

$$\tilde{G}_\infty(\theta_{inc}, \rho) = \frac{p_{12}(\theta_{inc})\rho + p_{22}(\theta_{inc})}{\rho^2 + q_{12}(\theta_{inc})\rho + q_{22}(\theta_{inc})}, \quad (2.9)$$

$$\tilde{f}_c(\theta_{inc}, \rho) = \frac{p_{13}(\theta_{inc})\rho^2 + p_{23}(\theta_{inc})\rho + p_{33}(\theta_{inc})}{\rho^2 + q_{13}(\theta_{inc})\rho + q_{23}(\theta_{inc})}. \quad (2.10)$$

Le tabelle 2.3.2 e 2.3.2 mostrano i valori dei coefficienti e dell'errore RMS. Anche in questo caso risulta confermata la buona qualità dell'approssimazione effettuata: per

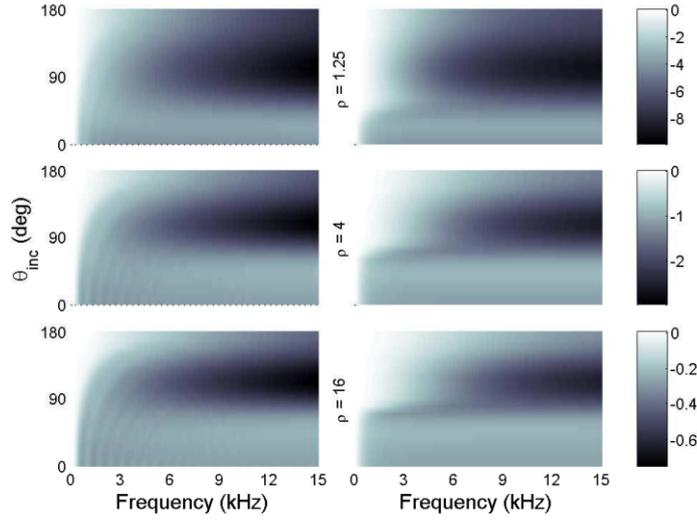


Figura 2.4: Confronto tra la NFTF normalizzata  $\hat{H}_{NF}$  (a sinistra) e la sua approssimazione tramite il shelving filter  $H_{sh}$  (a destra) per  $\rho = 1.25$ ,  $\rho = 4$ ,  $\rho = 16$ .

quanto riguarda  $\tilde{G}_{\infty}$ , il suo errore RMS è sempre inferiore a 0.04 dB, mentre per  $\tilde{f}_c$  l'errore è al di sotto della risoluzione minima di 10 Hz con cui sono state effettuate le analisi per quasi il 70% degli angoli di incidenza considerati.

Come nel caso del guadagno DC, per valori angolari intermedi si utilizza una procedura di interpolazione lineare.

$\theta_{inc}$	$p_{12}$	$p_{22}$	$q_{12}$	$q_{22}$	RMS [dB]
0°	-4.391	2.123	-0.55	-0.061	0.0007
10°	-4.314	-2.782	0.59	-0.173	0.0016
20°	-4.18	4.224	-1.006	-0.021	0.0057
30°	-4.012	3.039	-0.563	-0.316	0.0116
40°	-3.874	-0.566	0.665	-1.129	0.0199
50°	-4.099	-34.74	11.39	-8.301	0.039
60°	-3.868	3.271	-1.571	0.637	0.0151
70°	-5.021	0.023	-0.875	0.325	0.0097
80°	-6.724	-8.965	0.37	-0.083	0.0112
90°	-8.693	-58.38	5.446	-1.188	0.0179
100°	-11.17	11.47	-1.131	0.103	0.0217
110°	-12.08	8.716	-0.631	-0.12	0.0069
120°	-11.13	21.8	-2.009	0.098	0.0018
130°	-11.1	1.91	0.15	-0.401	0.0008
140°	-9.719	-0.043	0.243	-0.411	0.0014
150°	-8.417	-0.659	0.147	-0.344	0.0012
160°	-7.437	0.395	-0.178	-0.184	0.0006
170°	-6.783	2.662	-0.671	0.05	0.0006
180°	-6.584	3.387	-0.836	0.131	0.0008

Tabella 2.2: Coefficienti per l'equazione di  $\tilde{G}_{\infty}$  (2.9) ed errore RMS.

$\theta_{inc}$	$p_{13}$	$p_{23}$	$p_{33}$	$q_{13}$	$q_{23}$	RMS [Hz]
0°	0.457	-0.668	0.174	-1.746	0.699	1.19
10°	0.455	0.142	-0.115	-0.01	-0.348	0.92
20°	-0.87	3404	-1699	7354	-5350	3.36
30°	0.465	-0.913	0.437	-2.181	1.188	7.01
40°	0.494	-0.669	0.658	-1.196	0.256	19.14
50°	0.549	-1.208	2.02	-1.59	0.816	30.67
60°	0.663	-1.756	6.815	-1.296	1.166	21.65
70°	0.691	4.655	0.614	-0.889	0.76	60.32
80°	3.507	55.09	589.3	29.23	59.51	29.59
90°	-27.41	10336	16818	1945	1707	36.16
100°	6.371	1.735	-9.389	-0.058	-1.118	4.54
110°	7.032	40.88	-44.09	5.635	-6.18	2.53
120°	7.092	23.86	-23.61	3.308	-3.392	2.72
130°	7.463	102.8	-92.27	13.88	-12.67	2.33
140°	7.453	-6.145	-1.809	-0.877	-0.19	2.9
150°	8.101	-18.1	10.54	-2.23	1.295	5.28
160°	8.702	-9.05	0.532	-0.96	-0.023	2.15
170°	8.925	-9.03	0.285	-0.905	-0.079	3.71
180°	9.317	-6.888	-2.082	-0.566	-0.398	3.87

Tabella 2.3: Coefficienti per l'equazione di  $\tilde{f}_c$  (2.10) ed errore RMS.

### 2.3.3 Descrizione del modello e considerazioni

Le approssimazioni presentate nelle sezioni precedenti permettono la costruzione del modello finale di filtro nel campo vicino, tramite la loro integrazione con un modello di testa sferica nel campo lontano  $H_{sphere}^\infty$  [7]: esso viene presentato in figura 2.5.

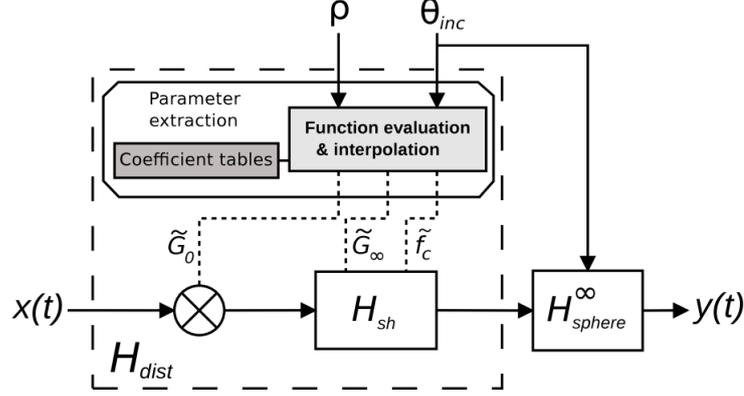


Figura 2.5: Schema a blocchi del modello per gli effetti binaurali nel campo vicino che tiene conto della distanza.

Il blocco computazionale *Parameter extraction* riceve come input la distanza normalizzata  $\rho$  e l'angolo di incidenza  $\theta_{inc}$ . Esso utilizza i coefficienti delle tabelle 2.1, 2.3.2 e 2.3.2 ed interpola le funzioni  $\tilde{G}_0$  (2.5) e  $\tilde{G}_\infty$  (2.9),  $\tilde{f}_c$  (2.10), che vengono rispettivamente utilizzate come fattore moltiplicativo e parametri del filtro shelving  $H_{sh}$ .

Questo primo blocco costituisce il modello  $H_{dist}$ , che viene posto in cascata al filtro  $H_{sphere}^\infty$ , che riceve a sua volta come input  $\theta_{inc}$ .

Per verificare la bontà del filtro  $H_{dist}$  dopo le approssimazioni introdotte, è stata effettuata una misura di distorsione spettrale:

$$SD = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( 20 \log_{10} \frac{|H(f_i)|}{|\tilde{H}(f_i)|} \right)^2} \quad [dB] \quad (2.11)$$

dove  $H$  è la risposta originale ( $H_{NF}$  nel nostro caso),  $\tilde{H}$  la risposta approssimata (ovvero qui  $H_{dist}$ ) e  $N$  il numero di frequenze prese in considerazione (100-15K Hz).

Il modulo di  $H_{dist}$ , calcolato con le solite 250 distanze delle analisi precedenti, prendendo gli angoli tra  $0^\circ$  e  $180^\circ$  ad intervalli di  $5^\circ$ , è stato comparato col modulo della corrispondente NTF originale  $H_{NF}$ . Per la misura dell'errore per le distanze intermedie si è nuovamente fatto uso dell'interpolazione lineare dei parametri.

La figura 2.6 fornisce una rappresentazione della distorsione.

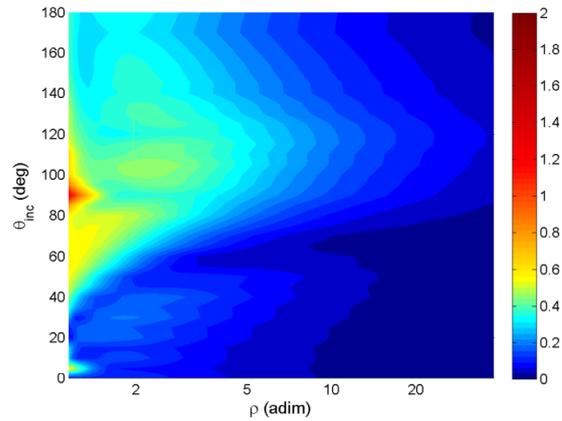


Figura 2.6: Distorsione spettrale introdotta dal modello  $H_{dist}$ .

Dalla figura si evince come l'approssimazione sia ottima, con una distorsione sempre inferiore a 1 dB quasi per l'intero near-field, eccetto per le distanze molto piccole in corrispondenza dei  $90^\circ$ . Questo risultato non stupisce, dal momento che in corrispondenza di questi valori il comportamento delle NFTF normalizzate è una via di mezzo tra quello di un filtro shelving e di un passa basso, come visto in 2.3.2.

Si osserva che per gli angoli intermedi, per definire i parametri dei quali si è fatto uso dell'interpolazione lineare, non ci sono scostamenti importanti, tranne un piccolo salto a  $\theta_{inc} = 5^\circ$  e  $\rho < 1.5$ . Tale risultato conferma l'adeguatezza di un'interpolazione solo del primo ordine e rende superfluo un eventuale campionamento più denso delle funzioni.

Da questa analisi si osserva infine che la distorsione assume valori pressoché nulli per valori di  $\rho$  elevati, col significato che  $H_{dist}$  tende a 1 per  $\rho \rightarrow \infty$ : il modello si rivela così essere valido anche nel campo lontano.

Confermata così la validità del modello dal punto di vista teorico, resta da indagare la sua efficacia attraverso esperimenti psicoacustici. Nei prossimi capitoli verrà descritta la procedura di analisi per testare la corretta parametrizzazione della distanza.



## Capitolo 3

# Design dell'esperimento

In questo capitolo si descrive l'esperimento psicoacustico condotto per: (1) paragonare la performance del modello proposto nel capitolo 2 alla controparte teorica; (2) testare l'efficacia del modello nel rendere la distanza relativa tra sorgenti sonore virtuali, paragonandolo a condizioni di controllo presenti in letteratura.

### 3.1 Apparato sperimentale

#### 3.1.1 Matlab

MATLAB (abbreviazione di MATrix LABoratory) è il software utilizzato per la realizzazione di questo esperimento.

MATLAB è un linguaggio di alto livello e un ambiente interattivo per il calcolo numerico, la programmazione e la visualizzazione di dati. Esso consente di analizzare, sviluppare algoritmi e creare modelli e applicazioni. Inoltre, propone un ambiente grafico, chiamato GUIDE (GUI Design Enviroment), adatto alla realizzazione di semplici e veloci interfacce. MATLAB mette inoltre a disposizione due utili toolbox: uno statistico (Statistics Toolbox) e uno relativo alla manipolazione delle immagini (Image Processing Toolbox).

#### 3.1.2 Cuffie

Le cuffie utilizzate per l'esperimento sono delle Sennheiser HDA 200, mostrate in figura 3.1. Si tratta di cuffie circumaurali, di tipo dinamico e chiuse, con i cuscinetti che circondano completamente la pinna. Sono in grado di isolare parzialmente l'orecchio da rumori esterni e la loro risposta in frequenza va da 20 a 20k Hz (come mostrato nella figura 3.2 in alto), con un'impedenza di 40  $\Omega$ .

In [22], Schiesaro ha studiato come la riproduzione tramite cuffie introduca distorsioni di varia natura. In particolare, cuffie tra cui le Sennheiser HDA 200 presentano un comportamento in frequenza assimilabile a quello di un filtro passa banda con frequenze di taglio intorno ai 150 Hz e 9 kHz, lontano da una risposta ideale a guadagno costante. Un'ulteriore osservazione va fatta riguardo ai riposizionamenti della cuffia sulla testa, che causano una variabilità della HpTF (Headphones Transfer Function), nonché del



Figura 3.1: Sennheiser HDA 200.

corrispondente temporale HpIR (Headphones Impulse Response), particolarmente in alcuni range di frequenze [21]. La figura 3.2 in basso mostra la deviazione standard su 10 riposizionamenti.

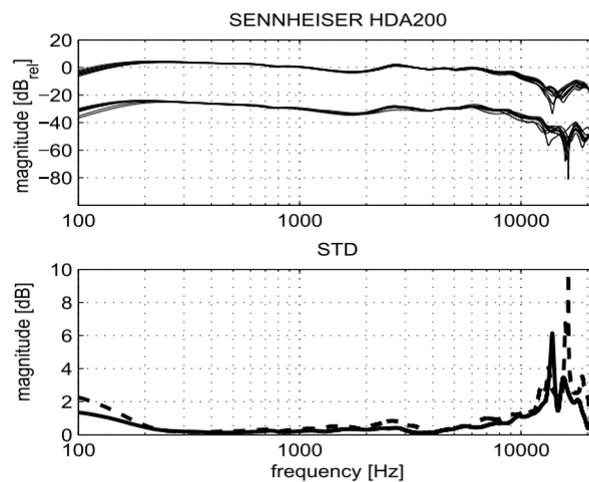


Figura 3.2: Sennheiser HDA 200: risposta in frequenza (in alto) e deviazione standard su 10 riposizionamenti (in basso).

Si osserva come il riposizionamento produca, specialmente per frequenze oltre i 7 kHz, un'equalizzazione non voluta del segnale binaurale: si è reso quindi necessario un intervento di compensazione delle cuffie.

La risposta in frequenza delle cuffie è stata calcolata su un manichino KEMAR senza padiglione auricolare: da ciò consegue che la compensazione effettuata risulta essere verosimile, ma non perfettamente adattata sul singolo soggetto. Infatti, un'esatta riproduzione binaurale è possibile solo per il binomio soggetto-cuffia su cui sono state acquisite le HpIR.

### 3.1.3 Scheda audio esterna

La scheda audio esterna utilizzata è una Roland Edirol AudioCapture UA-101, mostrata in figura 3.3.



Figura 3.3: Roland Edirol 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 96 kHz. 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.

### 3.1.4 Pulsanti USB

Per permettere l'acquisizione delle risposte dei soggetti, si è utilizzata una coppia di pulsanti programmabili (USBButton™), collegati al pc tramite due porte USB. Al momento della configurazione è possibile impostare l'input da tastiera corrispondente ed il colore da attribuire ai pulsanti.



Figura 3.4: USBButton™.

Le specifiche sono disponibili al link <http://www.usbbutton.com/docs/USBButton.pdf>.

### 3.1.5 Cabina Silente

L'esperimento si è svolto all'interno di una cabina insonorizzata Sound Station Pro 45 della Puma s.r.l. (figura 3.5).

Si tratta di una cabina rivestita internamente da materiale poliuretano fonoassorbente autoestinguente a cellule aperte, con le caratteristiche di abbattimento sonoro e di riverbero illustrate in figura 3.6.



Figura 3.5: Sound Station Pro 45 by Puma s.r.l.

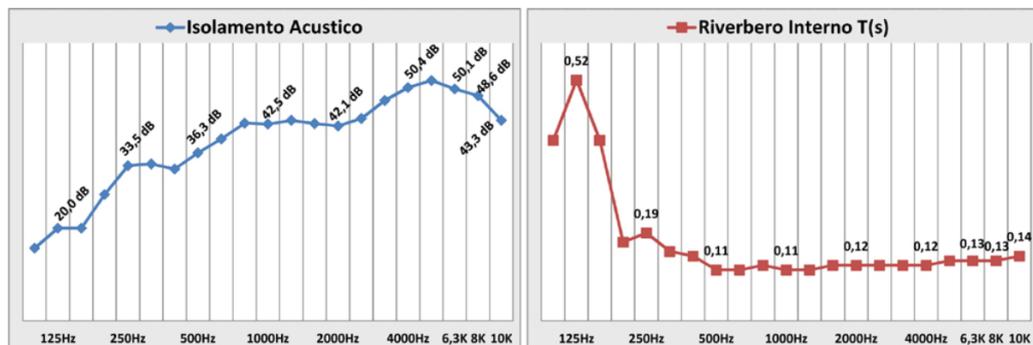


Figura 3.6: Caratteristiche di isolamento acustico e riverbero interno della Sound Station Pro 45 by Puma s.r.l.

All'interno della cabina è stato posizionato un computer, connesso ai pulsanti USB ed alla scheda audio, alla quale sono state a loro volta collegate le cuffie.

## 3.2 Soggetti

L'esperimento si è svolto su un totale di 10 soggetti di entrambi i sessi, con prevalenza maschile. L'età varia tra i 22 ed i 56 anni, con una media di 30,5. Ai soggetti sono state prese le misure antropometriche della testa, al fine di calcolarne il raggio medio  $a_{subject}$ . Per fare ciò, è stato utilizzato il modello lineare presente in [2], che stima il raggio della testa come somma pesata dei tre parametri antropometrici più rilevanti, ovvero larghezza, altezza e profondità:

$$a_{subject} = 0.51 \frac{X_1}{2} + 0.019 \frac{X_2}{2} + 0.18 \frac{X_3}{2} + 3.2, \quad [cm] \quad (3.1)$$

dove  $X_1$  è la larghezza,  $X_2$  l'altezza e  $X_3$  la profondità della testa, tutte misurate in cm.

Nella tabella 3.1 sono presentati i soggetti ed il raggio così calcolato.

ID	Sesso	Età	Raggio testa (cm)
001	M	29	9.30
002	M	41	9.37
003	M	27	8.88
004	M	26	9.56
005	M	24	9.21
006	M	29	9.20
007	F	56	8.61
008	M	28	8.61
009	F	23	8.88
010	M	22	9.23

Tabella 3.1: Soggetti dell'esperimento.

Si segnala che la maggior parte di loro aveva esperienze pregresse in test di psicoacustica.

Prima di iniziare, tutti i soggetti sono stati sottoposti ad una breve sessione di training, svolta in modo analogo alla sessione sperimentale vera e propria (vedasi sezione 3.5).

## 3.3 Stimoli

Lo stimolo è un rumore bianco uniformemente distribuito, della durata di 400 ms, sagomato con rampe lineari di 30 ms sia in capo che in coda. Studi condotti [11] hanno dimostrato come suoni aventi la stessa energia in tutte le bande di frequenza, come il rumore bianco, siano più facilmente localizzabili rispetto a suoni aventi un contenuto spettrale sparso (sinusoidi). Il codice utilizzato per creare tale stimolo è riportato nell'algoritmo A.1 in appendice.

Ciascuno stimolo finale presentato al soggetto, come vedremo nella prossima sezione, è stato ottenuto tramite la convoluzione con una HRTF ed il filtraggio per la compensazione delle cuffie utilizzate.

### 3.4 Condizioni

Gli stimoli sono stati prodotti per testare 4 condizioni, così da poter confrontare i risultati del modello da analizzare con quelli di altre HRTF. L'algoritmo A.2 implementa le quattro condizioni.

1. La condizione  $C_1$  riproduce stimoli filtrati attraverso l'HRTF misurata direttamente su un manichino KEMAR, presente nel database di Qu [19], disponibile per 8 distanze tra 20 e 160 cm, 14 elevazioni e 72 azimut.  
Tale condizione fa uso dell'algoritmo A.3 per l'importazione delle HRIR misurate (written by Zhen Xiao, Copyright by Key Lab of Machine Perception - Ministry of Education, Peking University, 2007-12-05).
2. la condizione  $C_2$  implementa la NFTF teorica di eq.(2.3), partendo dalla HRTF di eq.(1.2). La NFTF così creata viene tradotta in un filtro IIR di ordine 20 tramite la procedura di Yule-Walker. L'intensità degli stimoli ottenuti è stata scalata linearmente con la distanza.  
Il codice è riportato in A.4 per l'HRTF e A.5 per l'NFTF teorica;
3. la condizione  $C_3$  utilizza la NFTF approssimata tramite l'uso del filtro shelving (eq.(2.7)) e delle funzioni  $\tilde{G}_0$  (eq.(2.5)),  $\tilde{G}_\infty$  (eq.(2.9)),  $\tilde{f}_c$  (eq.(2.10)). Come per la condizione precedente, è stata utilizzata la procedura di Yule-Walker per garantire la stabilità dei filtri associati. Anche in questo caso l'intensità degli stimoli ottenuti è stata scalata linearmente con la distanza.  
Il codice è riportato negli algoritmi A.6 per il filtro shelving e A.7 per l'NFTF approssimata;
4. la condizione  $C_4$  effettua solo lo scaling lineare dell'intensità degli stimoli con la distanza.

Negli ultimi tre casi, gli stimoli nel campo vicino sono stati ottenuti applicando le condizioni alle HRTF di Qu misurate nel campo lontano (160 cm).

Per le condizioni  $C_2$  e  $C_3$ , si è scelto di porre verosimilmente le orecchie sul piano orizzontale a  $\pm 100^\circ$  di azimut.

### 3.5 Scelte sperimentali

Le posizioni spaziali scelte per le HRTF, espresse utilizzando il sistema di coordinate verticali polari, sono<sup>1</sup>:

---

<sup>1</sup>azimut compresi tra  $[0^\circ, 180^\circ]$  corrispondono all'emisfero destro, azimut tra  $[180^\circ, 360^\circ]$  all'emisfero sinistro

- per consistenza col database di Qu, le distanze  $r$  di campo vicino testate sono [20, 30, 40, 50, 75, 100] cm, normalizzate sul raggio della testa secondo:

$$r = \frac{[20, 30, 40, 50, 75, 100]}{a_{kemar}} * a_{subject} \quad (3.2)$$

con  $a_{kemar} = 8.75$  cm raggio della testa del manichino KEMAR,  $a_{subject}$  raggio medio della testa del soggetto in cm, calcolato con l'eq. (3.1);

- l'elevazione è fissata sul piano orizzontale ( $\phi = 0$ );
- l'azimut  $\theta_{inc}$  assume i valori [90°, 135°, 180°, 225°, 270°]. Si è scelto di considerare solo il semipiano orizzontale posteriore, visti i frequenti fenomeni di *front-back confusion* (già nominati in 1.6.2) che si verificano nel caso di azimut compresi tra [-90°, 90°], che portano l'ascoltatore ad identificare i suoni come comunque provenienti da dietro.

Gli stimoli sono stati presentati a coppie, separati da pause di 500 ms.

Si è scelto di proporre gli stimoli interni a ciascuna coppia a distanze contigue: precedenti studi [4] suggeriscono tale scelta affermando, infatti, che per discriminare due distanze nel campo vicino la JND (Just Noticeable Difference) sia soltanto del 5% rispetto alla distanza inferiore.

Ogni coppia di stimoli è stata somministrata 10 volte per ciascuna condizione, in particolare 5 volte in un ordine (primo suono più vicino del secondo) e 5 volte nell'altro, così da costituire una statistica sufficiente.

Le coppie di stimoli sono risultate quindi in totale:

$$5coppie\_distanze * 5azimut * 4condizioni * 2ordini * 5ripetizioni = 1000$$

L'esperimento è stato diviso in 5 blocchi da 200 coppie di stimoli, ciascuno della durata di circa 15 minuti. In ogni blocco viene presentata una ripetizione completa (ovvero con gli stimoli in entrambi gli ordini).

L'ordine degli stimoli di ciascun blocco è una permutazione random, ottenuta con l'algoritmo A.8, che sfrutta la funzione *random* messa a disposizione da MATLAB.

A.9 è l'algoritmo utilizzato per la gestione dell'esperimento. Esso richiede come input una stringa numerica, utilizzata come ID per contrassegnare il soggetto coinvolto, il raggio medio della testa  $a$  in cm (calcolato con l'eq.(3.1)) ed un valore di controllo  $n$ , con  $0 \leq n \leq 10$ , che permette di riprendere l'esperimento dall'ultima ripetizione effettuata, in caso di errori hardware/software.

### 3.6 Esecuzione dell'esperimento

Il soggetto viene fatto sedere all'interno della cabina silente. Dopo aver preso le misure della testa per calcolarne il raggio medio, gli vengono fatte indossare le cuffie e di fronte

a lui vengono posti i due pulsanti programmati.

Il soggetto viene istruito a premere il pulsante rosso se ritiene che la sequenza di due suoni che andrà ad ascoltare sia in avvicinamento, quello blu se in allontanamento.

Vengono spente luci e ventola della cabina, così da permettere la massima concentrazione solo sui suoni inviati in cuffia.

Una guida vocale conduce il soggetto attraverso le varie fasi dell'esperimento, ricordando le consegne e scandendo i tempi.

L'esperimento ha inizio nel momento in cui si preme uno dei due pulsanti. Dopo l'ascolto di una coppia di stimoli, premendo il pulsante si invia la risposta su quella che si ritiene essere la posizione relativa delle sorgenti e, contemporaneamente, si permette l'invio di un nuovo stimolo.

Alla fine di ogni blocco, al soggetto viene lasciata una pausa di tre minuti, durante la quale è libero di restare nella cabina oppure di uscire. Il conto alla rovescia per la ripresa dell'esperimento compare in una finestra sul display del computer ed è implementato dall'algoritmo A.10 (Copyright 2007 The MathWorks, Inc.).

Al termine di ogni ripetizione, i dati con le risposte del soggetto vengono salvati all'interno della cartella *results*, in un file di tipo *.mat* e nominato con l'ID del soggetto.

L'esperimento dura un tempo variabile a seconda della velocità di risposta del soggetto, in media intorno ad un'ora e venti.

## Capitolo 4

# Risultati dell'esperimento

In questo capitolo viene presentata l'analisi dei risultati dell'esperimento condotto.

Nell'ordine, si riportano i risultati di ciascun soggetto per coppie di distanze, azimut e condizioni e, a seguire, un'analisi che distingue l'ordine in cui sono stati presentati gli stimoli (avvicinamento/allontanamento), dalla quale sono emersi risultati interessanti.

### 4.1 Risultati dei soggetti

Nelle tabelle seguenti (4.1, 4.2, 4.3) vengono presentati i risultati rispetto a coppie di distanza, azimut e condizioni per ciascun soggetto.

Per facilitarne la lettura, si è adottata la seguente nomenclatura:

- Coppie di distanze:  $D_1=20-30$  cm,  $D_2=30-40$  cm,  $D_3=40-50$  cm,  $D_4=50-75$  cm,  $D_5=75-100$  cm;
- Azimut:  $A_1=90^\circ$ ,  $A_2=135^\circ$ ,  $A_3=180^\circ$ ,  $A_4=225^\circ$ ,  $A_5=270^\circ$ ;
- Condizioni:  $C_i$  con  $i=1,2,3,4$  come definito nella sezione 3.4.

I risultati sono stati ricavati con l'algoritmo A.11.

Nel rappresentare i risultati dei soggetti divisi per ordine di presentazione (ascendente, ovvero stimoli in avvicinamento e discendente, ovvero in allontanamento), si tralascia l'analisi che confronta le diverse condizioni, in quanto non significativamente diversa dalla precedente.

Si osserva come le percentuali d'errore del soggetto 007 siano spesso al di sopra di una soglia accettabile (nel caso di stimoli discendenti confrontati per distanze, ad esempio, la percentuale d'errore raggiunge l'80%). Tali risultati fanno quindi ritenere il soggetto 007 non affidabile per questo esperimento. Per quanto riguarda l'analisi statistica della prossima sezione, si escluderà quindi il suo contributo.

Soggetto	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>
001	13	13	11.5	3	18
002	9	8	6	1.5	13
003	10.5	17	30	21	44
004	32.5	37	39	31	40
005	14	13	14	1	14.5
006	20.5	13.5	20.5	10	35.5
007	44	37	43.5	24	38
008	14.5	10	13.5	3.5	10.5
009	16.5	12.5	19	2.5	12
010	30.5	32.5	38	19	40

Tabella 4.1: Tasso d'errore (%) per **coppie di distanze**.

Soggetto	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>
001	11	11	13.5	14	9
002	8	7.5	7.5	7	7.5
003	21	27.5	32.5	25.5	16
004	31	35.5	34.5	42.5	36
005	8	11.5	15	11.5	10.5
006	16.5	21	26.5	18.5	17.5
007	35.5	41	37.5	36	36.5
008	9	8	12.5	12.5	10
009	12.5	12	13.5	14	10.5
010	31.5	31.5	34	29	34

Tabella 4.2: Tasso d'errore (%) per **azimut**.

Soggetto	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>
001	18.8	10.4	8	9.6
002	16.4	6	4	3.6
003	30	19.2	19.2	29.6
004	39.2	31.6	38.8	34
005	20.4	8.8	9.6	6.4
006	27.2	16	18.4	18.4
007	39.2	35.6	36.8	37.6
008	18.8	6.8	10	6
009	20.8	9.2	11.6	8.4
010	36.8	28.8	27.2	35.2

Tabella 4.3: Tasso d'errore (%) per **condizione**.

Soggetto	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>
001	1	5	5	6	36
002	3	4	4	1	18
003	14	25	56	41	86
004	18	21	27	27	50
005	8	6	5	2	13
006	3	6	23	15	63
007	21	7	7	9	4
008	1	2	4	5	20
009	6	4	5	5	18
010	7	6	18	16	47

Tabella 4.4: Stimoli in ordine **ascendente**: tasso d'errore (%) per **coppie di distanze**.

Soggetto	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>
001	25	21	18	0	0
002	15	12	8	2	8
003	7	9	4	1	2
004	47	53	51	35	30
005	20	20	23	0	16
006	38	21	18	5	8
007	67	67	80	39	72
008	28	18	23	2	1
009	27	21	33	0	6
010	54	59	58	22	33

Tabella 4.5: Stimoli in ordine **discendente**: tasso d'errore (%) per **coppie di distanze**.

Soggetto	A <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>
001	12	14	8	11	8
002	6	6	9	2	7
003	36	55	60	42	29
004	26	39	22	32	24
005	7	8	8	6	5
006	20	31	25	18	16
007	12	15	7	6	8
008	3	6	11	10	2
009	8	11	6	6	7
010	6	25	34	19	10

Tabella 4.6: Stimoli in ordine **ascendente**: tasso d'errore (%) per **azimut**.

## 4.2 Analisi statistica

Prima di procedere all'analisi statistica, si è verificato che i dati raccolti rispettassero le seguenti caratteristiche:

Soggetto	D <sub>1</sub>	D <sub>2</sub>	D <sub>3</sub>	D <sub>4</sub>	D <sub>5</sub>
001	10	8	19	17	10
002	10	9	6	12	8
003	6	0	5	9	3
004	36	32	47	53	48
005	9	15	22	17	16
006	13	11	28	19	19
007	59	67	68	66	65
008	15	10	14	15	18
009	17	13	21	22	14
010	57	38	34	39	58

Tabella 4.7: Stimoli in ordine **discendente**: tasso d'errore (%) per **azimut**.

1. NORMALITÀ, ottenuta mediante una trasformazione arcseno [1] e verificata col test di D'Agostino & Pearson (tale trasformazione implica che l'errore presentato nei grafici di questa sezione non sia più quello percentuale, pur mantenendone sufficientemente la corrispondenza per i valori d'interesse);
2. OMOSCHEDASTICITÀ, verificata col test di Bartlett;
3. SFERICITÀ, verificata col test di Mauchly. Nei casi in cui tale ipotesi non risultasse soddisfatta, è stata applicata una correzione di Greenhouse-Gasser.

#### 4.2.1 Analisi sulle coppie di distanza

L'analisi per quanto riguarda la distanza con gli stimoli in entrambe le direzioni (fig. 4.1) non evidenzia un trend significativo. Si osserva che la coppia di distanze D<sub>4</sub> (50-75 cm) presenta meno errori: una possibile spiegazione potrebbe essere che, essendo questa la prima coppia dove la differenza tra le due distanze aumenta del 50% dopo due coppie con incrementi intorno al 20-30%, la differenza è più netta. L'errore infatti torna a un valore alto per la coppia successiva D<sub>5</sub>, dove la differenza torna al 30%.

I dati sono stati soggetti ad una One-Way Repeated Measures ANOVA (ANalysis Of VAriance), che ha evidenziato un effetto significativo ( $F(4, 32)=13.83$ ,  $p < 0.0001$  \*\*\*\*) della coppia di distanze sull'errore, limitata alla distanza D<sub>4</sub> rispetto alle altre (\*\* con D<sub>1</sub> e D<sub>2</sub>, \*\*\*\* con D<sub>3</sub> e D<sub>5</sub>), come evidenziato dal test post-hoc di Bonferroni.

#### 4.2.2 Analisi sugli azimut

L'errore medio sugli azimut (fig. 4.2) presenta un incremento intorno ai valori centrali, mantenendo pressoché costante la deviazione standard. Tale risultato era prevedibile e già illustrato in letteratura: nel caso di una sorgente posta sul piano mediano, infatti, i valori di ITD e ILD sono prossimi allo zero e quindi l'apporto degli indicatori di campo vicino risulta meno evidente [9].

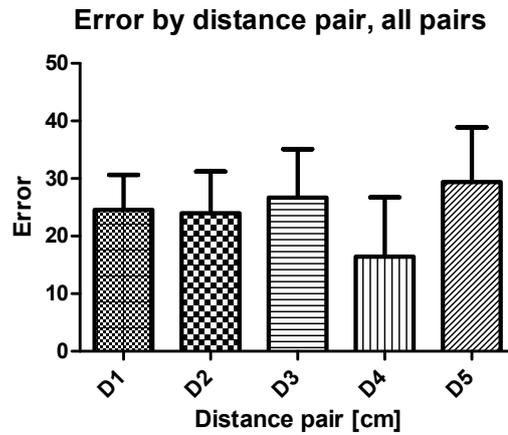


Figura 4.1: Media e deviazione standard dell'errore sulle **coppie di distanze**.

In questo caso, la One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(2, 18)=4.92$ ,  $p=0.02$  \*) degli azimut sull'errore, limitato al valore  $D_3$  ( $180^\circ$ ) rispetto agli azimut laterali  $D_1$  ( $90^\circ$ , \*\*) e  $D_5$  ( $270^\circ$ , \*), come previsto.

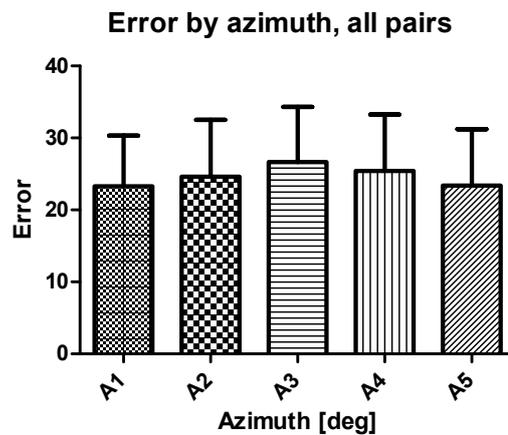


Figura 4.2: Media e deviazione standard dell'errore sugli **azimut**.

### 4.2.3 Analisi sulle condizioni

Per quanto riguarda l'analisi sulle quattro condizioni sperimentali (fig. 4.3), la condizione  $C_1$  presenta i risultati peggiori, mentre le altre tre sono confrontabili. Le cattive prestazioni della condizione  $C_1$  sono da attribuirsi alle misurazioni presenti nel database, riferite ad una testa artificiale e non scalate sul singolo soggetto.

Si conferma la validità del modello teorico di NFTF della condizione  $C_2$ , mentre, a dispetto delle aspettative, il modello approssimato di  $C_3$  dà risultati paragonabili al semplice scaling lineare. Una possibile giustificazione è il fatto che il modello sferico, come accennato in precedenza, pur essendo una buona approssimazione nei modelli di campo lontano, non si rivela invece essere adatto a descrivere le informazioni sulla distanza nel campo vicino.

La One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(3, 22)=20.93$ ,  $p<0.0001$  \*\*\*\*) delle condizioni sull'errore, limitato alla condizione  $C_1$  rispetto alle altre (\*\*\*\*).

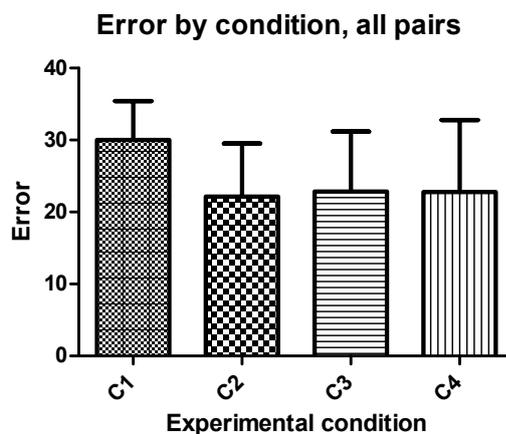


Figura 4.3: Media e deviazione standard dell'errore sulle **condizioni**.

#### 4.2.4 Analisi sulle coppie di distanza: stimoli in ordine ascendente/discendente

Analizzando i risultati tenendo conto dell'ordine degli stimoli, emergono risultati interessanti ed inaspettati.

I tassi d'errore sulle distanze dipendono nettamente dall'ordine: nel caso di stimoli ascendenti, il trend è decisamente crescente all'aumentare delle distanze (fig. 4.4), mentre è opposto nel caso di stimoli discendenti (fig. 4.5).

Come giustificazione al trend decrescente per gli stimoli discendenti si può adottare la spiegazione "evoluzionistica" secondo la quale, percependo un suono vicino, ci si aspetta un seguente avvicinamento della sorgente e viceversa con un suono lontano.

Per gli stimoli in ordine ascendente, la One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(2, 13)=25.31$ ,  $p<0.0001$  \*\*\*\*) della coppia di distanze sull'errore, per la coppia di distanze  $D_5$  rispetto alle altre (\*\*\*\*).

Per gli stimoli in ordine discendente, la One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(2, 19)=26.01$ ,  $p<0.0001$  \*\*\*\*) della coppia di di-

stanze sull'errore, sia per la coppia di distanze  $D_4$  rispetto a  $D_1$ ,  $D_2$ ,  $D_3$  (\*\*\*\*), sia per la coppia  $D_5$  rispetto a  $D_1$  (\*\*\*\*) e  $D_2$  e  $D_3$  (\*\*).

**Error by distance pair, decreasing distance**

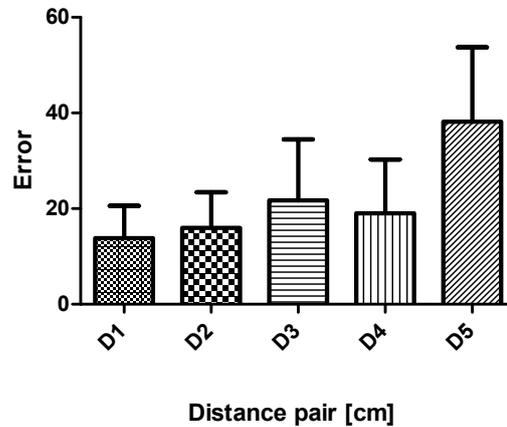


Figura 4.4: Stimoli in ordine **ascendente**: media e deviazione standard dell'errore sulle **coppie di distanze**.

**Error by distance pair, increasing distance**

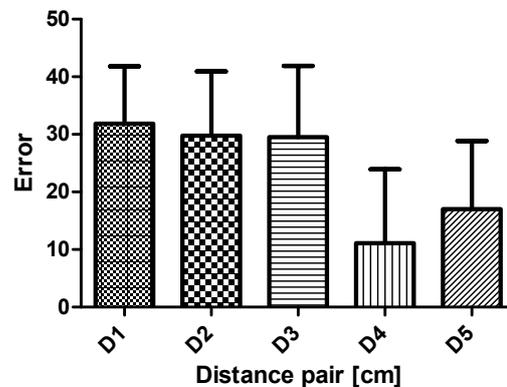


Figura 4.5: Stimoli in ordine **discendente**: media e deviazione standard dell'errore sulle **coppie di distanze**.

#### 4.2.5 Analisi sugli azimut: stimoli in ordine ascendente/discendente

Analizzando i trend degli errori sull'azimut, sempre nel caso di stimoli proposti in un singolo ordine, si osserva un effetto piccolo e non particolarmente significativo: in media,

nel caso di stimoli ascendenti, si hanno più errori per gli azimuth posti sul lato destro; viceversa, per gli stimoli discendenti, aumentano gli errori sul lato sinistro.

Probabilmente si tratta di errori da parte dei soggetti che, involontariamente, hanno premuto il pulsante corrispondente al lato da cui proveniva lo stimolo. Infatti, il pulsante per indicare l'avvicinamento era posto sulla sinistra e gli stimoli ascendenti sono stati maggiormente indicati come provenienti da quel lato. La stessa osservazione vale per il pulsante relativo all'allontanamento sul lato destro e gli stimoli discendenti.

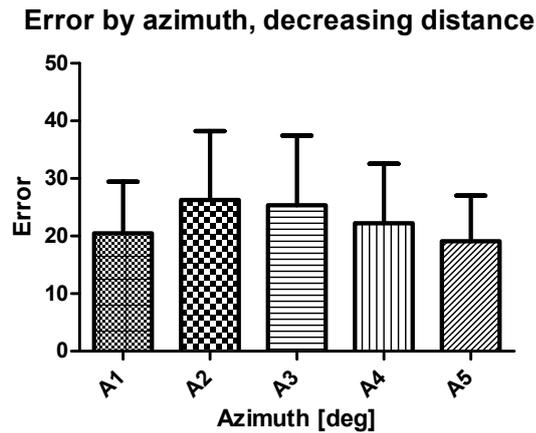


Figura 4.6: Stimoli in ordine **ascendente**: media e deviazione standard dell'errore sugli **azimut**.

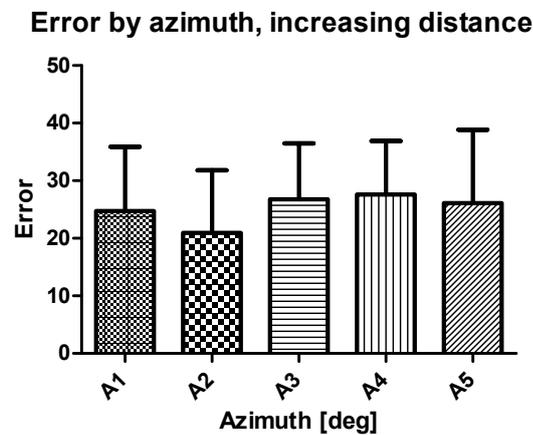


Figura 4.7: Stimoli in ordine **discendente**: media e deviazione standard dell'errore sugli **azimut**.

Per gli stimoli in ordine ascendente, la One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(2, 20)=5.13$ ,  $p=0.02$  \*) degli azimuth sull'errore,

per il valore  $A_5$  rispetto ad  $A_2$  (\*\*) e ad  $A_3$  (\*).

Per gli stimoli in ordine discendente, la One-Way Repeated Measures ANOVA ha evidenziato un effetto significativo ( $F(3, 24)=3.90$ ,  $p=0.02$  \*) degli azimut sull'errore, per il valore  $A_2$  rispetto a  $A_3$  e  $A_4$  (\*).

#### 4.2.6 Interazione tra variabili indipendenti

Per individuare eventuali correlazioni tra le variabili indipendenti, sono state effettuate ulteriori analisi, del tipo Two-Way Repeated Measures ANOVA.

Per quanto riguarda la correlazione **coppie di distanze e condizioni**, è risultato che l'interazione è significativa ( $F(12, 120)=13.15$ ,  $p<0.0001$ ).

L'interazione tra **azimut e condizione** non risulta invece significativa ( $F(12, 120)=0.72$ ,  $p=0.73$ ).

Si sono studiate, inoltre, le interazioni tra l'ordine di presentazione degli stimoli e le altre variabili.

Le interazioni tra **condizione e ordine** e **azimut e ordine** non sono risultate significative, confermando le ipotesi precedenti che avevano rispettivamente escluso l'analisi One Way delle condizioni al variare dell'ordine di presentazione ed attribuito il lieve decentramento dell'errore sull'azimut alla posizione dei pulsanti.

L'interazione tra **distanza ed ordine di presentazione** è risultata invece altamente significativa ( $F(4, 40)=10.20$ ,  $p<0.0001$ ), come da aspettativa.

Si è considerata, infine, l'interazione tra **le condizioni e le ripetizioni**, che non è in nessun caso risultata significativa.

Ci si è domandati, in conclusione, se le diverse misure del raggio medio della testa dei soggetti potesse aver influito sensibilmente sulle HRTF delle quattro condizioni. Confrontando le risposte degli stimoli somministrati ai soggetti con rispettivamente raggio minore (8.61 cm) e maggiore (9.56), non si notano tuttavia differenze significative.



## Capitolo 5

# Conclusioni e sviluppi futuri

Questa tesi ha avuto come scopo la validazione dal punto di vista pratico del modello di sorgente sonora virtuale nel campo vicino proposto nel capitolo 2.

I risultati mostrano come le performance di tale modello siano paragonabili a quelle ottenute usando il modello teorico presente in letteratura. Le approssimazioni introdotte si rivelano quindi corrette e sufficienti, seppure solo del primo ordine, rendendo così possibile la formalizzazione di un modello low-cost.

Per quanto riguarda l'efficacia nel rendere la distanza relativa, i dati non mostrano miglioramenti rispetto al modello teorico o al semplice scaling lineare. È emerso inoltre un risultato interessante su come l'ordine di presentazione di due stimoli influenzi la percezione delle diverse coppie di distanze.

Alla fine di questo lavoro, si considerano aperte nuove strade di ricerca in due direzioni:

1. dal momento che l'intensità dello stimolo sonoro si è rivelato essere l'indizio percettivo dominante per discriminare tra le distanze, ci si potrebbe proporre di ripetere l'esperimento eliminando l'informazione di intensità, al fine di indagare sugli ulteriori effetti binaurali che concorrono all'identificazione della distanza;
2. i risultati sui trend opposti riguardo alla somministrazione di stimoli nei due ordini incoraggiano a condurre un esperimento psicofisico che metta in luce più chiaramente il legame tra distanza ed ordine di presentazione. Al momento, si suppone che tale fenomeno possa essere legato non direttamente alla posizione della sorgente nello spazio, quanto all'intensità dello stimolo.

Ci si propone, infine, di riuscire ad inquadrare il lavoro svolto in un contesto più ampio: la completa validazione di questo modello di campo vicino permetterà la sua integrazione in un framework di tipo strutturale per la parametrizzazione ed il rendering personalizzato di HRTF [13, 25], accoppiandolo ad altri modelli studiati dal gruppo di ricerca *SMC* negli ultimi anni [14, 26].



# Appendice A

## Algoritmi MATLAB

In questa appendice è riportato il codice relativo alle funzioni MATLAB utilizzate.

### A.1 stimulus

```
function stimulus(len, ramp_len, eq, pinna)

% Algorithm for stimulus creation
% INPUTS:
% len           - Burst duration (s)
% ramp_len     - Ramp duration (s)
% eq           - Headphones equalization (0=no, 1=yes)
% pinna        - Equalization with or without pinna (1-0)
% OUTPUT:
% created_stimulus.wav - Stimulus

if (pinna == 1) % load eq filter
    load('eqfilter_HDA200_pinna.mat');
else
    load('eqfilter_HDA200_nopinna.mat');
end

fs=48000;
len_s=round(fs*len);
ramp_s=round(fs*ramp_len);

ramp = 0:1/ramp_s:1-1/ramp_s;
reverse_ramp = 1-1/ramp_s:-1/ramp_s:0;

sigl = 2*rand(1, len_s)-1;
sigl(1:ramp_s) = sigl(1:ramp_s).*ramp;
sigl(len_s-ramp_s+1:len_s) = sigl(len_s-ramp_s+1:len_s).*reverse_ramp;

sigl = [zeros(1, round(fs/50)) sigl zeros(1, round(fs/50))];
```

```

if (eq == 1) % equalize headphones
    [r,c] = size(sig1);
    sigout = zeros(c,2);
    sigout(:,1) = fftfilt(data.IR(:,1),sig1);
    sigout(:,2) = fftfilt(data.IR(:,2),sig1);
else % do not equalize
    sigout = [sig1' sig1'];
end

wavwrite(sigout/10,fs,'created_stimulus');
wavplay(sigout,fs);

end

```

## A.2 create\_hrtfs

```

function create_hrtfs(acm)

% Algorithm for the 4 conditions (Sec. 3.4)
% INPUT:
% acm          - Head radius (cm)
% OUTPUT:
% distX_aziY.wav - Stimulus at dist X and azi Y for each condition

hold off;
close all;

target_fs = 48000; % sampling frequency
signal = wavread('created_stimulus.wav'); % stimulus
dist = [20 30 40 50 75 100 130 160]; % distances in Qu db

% create 1st set of stimuli (Qu db HRTFs @ different distances)
for azi = 90:45:270
    for i = 1:8
        hrir_l_load = readhrir('./', dist(i), 0, azi, 'l'); % load hrirs
        hrir_r_load = readhrir('./', dist(i), 0, azi, 'r');
        hrir_l = resample(hrir_l_load,target_fs,65536); % resample hrirs to
            target_fs
        hrir_r = resample(hrir_r_load,target_fs,65536);
        sig_l = fftfilt(hrir_l,signal(:,1)); % filter stimulus with hrir
        sig_r = fftfilt(hrir_r,signal(:,2));
        wavwrite([sig_l sig_r], target_fs, 32, ['./set_1/dist' int2str(i) '
            _azi' int2str(azi) '.wav']);
    end
end

%create 2nd set of stimuli (HRTFs scaled with theoretical NRTF)
for azi = 90:45:270
    for i = 1:6
        [y, fs] = wavread(['./set_1/dist' int2str(8) '_azi' int2str(azi) '
            .wav']);
    end
end

```

```

[B,A] = theo_nftf(azi,i,acm);
sig_l2 = (160/dist(i))*filter(B(:,1),A(:,1),y(:,1));
sig_r2 = (160/dist(i))*filter(B(:,2),A(:,2),y(:,2));
wavwrite([sig_l2 sig_r2], fs, 32, ['./set_2/dist' int2str(i) '_azi'
int2str(azi) '.wav']);
end
end

% create 3rd set of stimuli (normalized HRTFs approximated through a first-
order high-frequency shelving filter)
for azi = 90:45:270
for i = 1:6
[z, fs] = wavread(['./set_1/dist' int2str(8) '_azi' int2str(azi) '
.wav']);
[F,E,g_0] = approx_nftf(azi,i,acm);
sig_l3 = filter(F(:,1),E(:,1),z(:,1));
sig_r3 = filter(F(:,2),E(:,2),z(:,2));
sig_l3 = sig_l3*(10^(g_0(1)/20))*(160/dist(i));
sig_r3 = sig_r3*(10^(g_0(2)/20))*(160/dist(i));
wavwrite([sig_l3 sig_r3], fs, 32, ['./set_3/dist' int2str(i) '_azi'
int2str(azi) '.wav']);
end
end

% create 4th set of stimuli (linear scaling)
for azi = 90:45:270
for i = 1:6
[y, fs] = wavread(['./set_1/dist' int2str(8) '_azi' int2str(azi) '
.wav']);
sig_l4 = (160/(dist(i)))*y(:,1);
sig_r4 = (160/(dist(i)))*y(:,2);
wavwrite([sig_l4 sig_r4], fs, 32, ['./set_4/dist' int2str(i) '_azi'
int2str(azi) '.wav']);
end
end
end
end

```

## A.3 hrir

```

function hrir = readhrir(filepath, dist, elev, azi, lr)
% written by Zhen Xiao
% Copyright by Key Lab of Machine Perception (Ministry of Education), Peking
University 2007-12-05
% read Qu-HRTF database
% filepath is the path of the database
% dist is the distance including 20 30 40 50 75 100 130 160
% elev is the elevation, values from -40 to 90 in step of 10
% azi is the azimuth, values from 0 to 355 in step of 5 (elev <= 50), from 0
to 350 in step of 10 (elev == 60), from 0 to 345 in step of 15 (elev ==
70), from 0 to 330 in step of 30 (elev == 80), and 0 (elev == 90)

```

```

% lr represents the data from right ear or left ear. lr == l means left, and
%   lr == r means right and lr == lr means both.
% the length of each HRIR is 1024 points.

if nargin ~= 5
    error('Wrong number of input arguments')
end;
filename = [filepath, 'dist', int2str(dist), '\elev', int2str(elev), '\azi',
    int2str(azi), '_elev', int2str(elev), '_dist', int2str(dist), '.dat'];
p = fopen(filename, 'r');
if lr == 'l'
    hrir = fread(p, 1024, 'double');
else if lr == 'r'
    fseek(p, 1024*8, 'bof');
%     fseek(p, 1024*8, 'bof');
    hrir = fread(p, 1024, 'double');
else
    hrir = fread(p, 'double');
    hrir = reshape(hrir, 1024, 2);
end;
end;
fclose(p);

```

## A.4 hrtf

```

function H = hrtf(a, r, theta, f, c, threshold)

% Evaluation algorithm for Eq.(2.2) (from [12])
% INPUT:
% a           - Head radius (m)
% r           - Sphere center-source distance (m)
% theta       - Incidence angle (degrees)
% f           - Frequency or frequency array (Hz)
% c           - Sound speed (typically 343 m/s)
% threshold   - Threshold for the accuracy of computation (typically 10e-5)
% OUTPUT:
% H           - Evaluation of Eq. (2.2) in f

[row, len] = size(f);
H = zeros(1, len);
i = sqrt(-1);
theta = (theta/180)*pi;
for k = 1:len
    x = cos(theta);
    mu = (2 * pi * f(k) * a) / c;
    rho = r / a;
    zr = 1 / (i * mu * rho);
    za = 1 / (i * mu);
    Qr2 = zr;
    Qr1 = zr * (1 - zr);
    Qa2 = za;

```

```

Qa1 = za * (1 - za);
P2 = 1;
P1 = x;
sum = 0;
term = zr / (za * (za - 1));
sum = sum + term;
term = (3 * x * zr * (zr - 1)) / (za * (2 * za^2 - 2 * za + 1));
sum = sum + term;
oldratio = 1;
newratio = abs(term) / abs(sum);
m = 2;
while ((newratio > threshold) | (oldratio > threshold))
    Qr = - (2 * m - 1) * zr * Qr1 + Qr2;
    Qa = - (2 * m - 1) * za * Qa1 + Qa2;
    P = ((2 * m - 1) * x * P1 - (m - 1) * P2) / m;
    term = ((2 * m + 1) * P * Qr) / ((m + 1) * za * Qa - Qa1);
    sum = sum + term;
    m = m + 1;
    Qr2 = Qr1;
    Qr1 = Qr;
    Qa2 = Qa1;
    Qa1 = Qa;
    P2 = P1;
    P1 = P;
    oldratio = newratio;
    newratio = abs(term) / abs(sum);
end
H(k) = (rho * exp(-i * mu) * sum) / (i * mu);
end

```

## A.5 theo\_nftf

```

function [B,A] = theo_nftf(azi,i,acm)

% Algorithm for the 2nd condition
% INPUT:
% azi          - Incidence angle (degrees)
% i            - # of near-field distances
% acm          - Head radius (cm)
% OUTPUT:
% [B, A]      - Filter coefficients for theoretical NFTF

%%% audio parameters
fs = 48000; % sampling rate (Hz)
sp = 100; % frequency spacing (Hz)
freqs = (sp:sp:fs/2); % frequency vector (for evaluation)
threshold = .001; % threshold for convergence

%%% physical parameters
a = acm/100; % sphere radius
c = 343; % speed of sound

```

```

r = [0.2 0.3 0.4 0.5 0.75 1]; % near-field distances

%%% initialization
yw_order = 20; % filter order for Yule-Walker procedure
B = zeros(yw_order+1,2);
A = zeros(yw_order+1,2);

% create left NTF filter
Hl = hrtf(a, r(i), abs(abs(80-azi)-180), freqs, c, threshold);
H0 = hrtf(a, 1000, abs(abs(80-azi)-180), freqs, c, threshold);
Hl = abs(Hl)./abs(H0);
[B(:,1),A(:,1)] = yulewalk(yw_order,[0 2*freqs/fs],[Hl(1) Hl]);

% create right NTF filter
Hr = hrtf(a, r(i), abs(abs(280-azi)-180), freqs, c, threshold);
H0 = hrtf(a, 1000, abs(abs(280-azi)-180), freqs, c, threshold);
Hr = abs(Hr)./abs(H0);
[B(:,2),A(:,2)] = yulewalk(yw_order,[0 2*freqs/fs],[Hr(1) Hr]);

end

```

## A.6 shelving

```

function [b, a] = shelving(fc, G, fs)

% Determine the coefficients of shelving filter in Eq.(2.7) e (2.8)
% INPUTS:
% fc      - Cutoff frequency
% G       - Asymptotic gain at high frequency
% fs      - Sampling rate (Hz)
% OUTPUT:
% a, b    - Coefficients of shelving filter

V0 = 10^(G/20);
H0 = V0 - 1;
c = (V0*tan(pi*fc/fs) - 1)/(V0*tan(pi*fc/fs) + 1);

b0 = 1+H0/2-c*H0/2;
b1 = c*H0/2+c-H0/2;
a1 = c;

%return values
a = [ 1, a1];
b = [ b0, b1];

end

```

## A.7 approx\_nftf

```

function [F,E,g_0] = approx_nftf(azi,i,acm)

% Algorithm for the 3rd condition
% INPUT:
% azi          - Incidence angle (degrees)
% i            - # of near-field distances
% acm          - Head radius (cm)
% OUTPUT:
% [F, E]       - Filter coefficients for approximated NFTF

f = 100:100:(48000/2);

% coefficients tab. 2.1
g0_m = [12.97    -9.691    -1.136    0.2187;
        13.19    234.2     18.48    -8.498;
        12.13   -11.17    -1.249    0.3464;
        11.19   -9.035    -1.017    0.3357;
        9.91    -7.866    -0.8295   0.3786;
        8.328   -7.416    -0.6659   0.4206;
        6.493   -7.312    -0.5028   0.4226;
        4.455   -7.278    -0.3205   0.3819;
        2.274   -7.291    -0.1098   0.3141;
        0.018   -7.484    -0.1298   0.2402;
        -2.242  -8.04     0.3948    0.1771;
        -4.433  -9.231    0.699     0.1315;
        -6.488  -11.61    1.084     0.1125;
        -8.342  -17.38    1.757     0.1423;
        -9.93   -48.42    4.764     0.4616;
        -11.29  9.149    -0.6444   -0.1376;
        -12.22  1.905    0.1087    -0.08168;
        -12.81  -0.7484   0.3864    -0.05841;
        -13     -1.32     0.4504    -0.05468];

% coefficients tab. 2.2
ginf_m=
[-4.39088003383214,2.12314028967817,-0.550342011542206,-0.0609632179570666;
-4.31411485612902,-2.78237618701472,0.590420385683966,-0.172746985684833;
-4.17982017240275,4.22435073006052,-1.00570152593507,-0.0208248019461848;
-4.01243724959128,3.03884632583880,-0.562827215497413,-0.315530209554668;
-3.87391498129897,-0.565735408932855,0.665289421975951,-1.12899967239948;
-4.09898173717451,-34.7433999513258,11.3912041461328,-8.30082997910485;
-3.86796262102190,3.27082228979764,-1.57062373148895,0.637154297736675;
-5.02116461002581,0.0229089737344803,-0.875130417188532,0.324620950590807;
-6.72409501648114,-8.96510126454168,0.369882704093707,-0.0828109372589683;
-8.69329910871191,-58.3796219261361,5.44580276485252,-1.18751981453947;
-11.1704912914291,11.4661521394652,-1.13052911969852,0.103070861420148;
-12.0776735297249,8.71628051050344,-0.630829712107489,-0.120072103973373;
-11.1271806105540,21.7984190318694,-2.00892449298097,0.0976450923633144;
-11.1000584080813,1.91012238813865,0.150021173069516,-0.400728513723333;

```

```
-9.71880518920443,-0.0433437010436329,0.243468707131706,-0.411250064654372;
-8.41681657550721,-0.659406261905253,0.146851520570887,-0.344276628796043;
-7.43660115238932,0.394518952280143,-0.177624744628400,-0.184003297507725;
-6.78277805190077,2.66233867498760,-0.670941006086112,0.0496730869687261;
-6.58368619600661,3.38722705743118,-0.835925966690698,0.131450137750410];
```

```
% coefficients tab. 2.3
```

```
fc_m =
```

```
[0.456738130610664,-0.667885376513498,0.174406021661890,-1.74567049816251,
 0.699218675019255;
0.454541844259773,0.142100988668044,-0.114720536235189,-0.0104690116229417,
 -0.347938699848861;
-0.870223549286121,3404.33438831235,-1699.02892572387,7354.12964681382,
 -5350.79892271667;
0.465442406611391,-0.912574612595962,0.436686828365789,-2.18089699800851,
 1.18770006687586;
0.493539971789342,-0.669496562105813,0.658274961227184,-1.19621715721824,
 0.255814415541582;
0.548924474163761,-1.20757593224935,2.02008334817018,-1.58986075214055,
 0.815838454378319;
0.662601864931785,-1.75638044635979,6.81461697098597,-1.29559191628743,
 1.16551507219983;
0.690569034774204,4.65473838116712,0.614259281095880,-0.889347730665654,
 0.759819015198427;
3.50669670067451,55.0856937352457,589.266194095640,29.2279037740886,
 59.5136301332242;
-27.4119376541283,10336.0491696565,16818.0206632154,1945.01083639478,
 1706.85442950588;
6.37070533615590,1.73520272229387,-9.38878974680223,-0.0581158584335969,
 -1.11804659509488;
7.03218876856814,40.8779294864441,-44.0922128791881,5.63542133212250,
 -6.17990554759501;
7.09158371449472,23.8623791668355,-23.6077449317165,3.30838835448727,
 -3.39224360314035;
7.46349400352057,102.798771547131,-92.2670858226236,13.8800175747823,
 -12.6675586271947;
7.45251342671091,-6.14513486489671,-1.80898601244964,-0.877374036262434,
 -0.189682819247414;
8.10082028895926,-18.1016208822406,10.5353693724314,-2.22980638933625,
 1.29476162012514;
8.70222637863090,-9.05021354116830,0.531552694684510,-0.960077164011900,
 -0.0226235472464294;
8.92494024924663,-9.02995716950749,0.285348127991700,-0.905258967584000,
 -0.0787223215378096;
9.31741576337106,-6.88848142002089,-2.08217108815710,-0.566051658783982,
 -0.398029037445496];
```

```
%% physical parameters
```

```
a = acm/100; % sphere radius
```

```
r = [0.2 0.3 0.4 0.5 0.75 1]; % near-field distances
```

```
B = zeros(2,2);
```

```
A = zeros(2,2);
```

```

F = zeros(21,2);
E = zeros(21,2);
g_0 = zeros(1,2);

if rem(azi, 10) == 0
% create left shelving filter
% Eq.(2.5) left
g_0(1) = (g0_m(abs(abs(80-azi)-180)/10+1,1)*(r(i)/a) + g0_m(abs(abs(80-azi)-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs(80-azi)-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(80-azi)-180)/10+1,4));
% Eq.(2.9) g_inf
g_inf = (ginf_m(abs(abs(80-azi)-180)/10+1,1)*(r(i)/a) + ginf_m(abs(abs(80-azi)-180)/10+1,2)) / (((r(i)/a)^(2)) + ginf_m(abs(abs(80-azi)-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(80-azi)-180)/10+1,4));
% Eq.(2.10) f_c
f_c = ((fc_m(abs(abs(80-azi)-180)/10+1,1)*((r(i)/a)^(2)) + fc_m(abs(abs(80-azi)-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs(80-azi)-180)/10+1,3)) / ((r(i)/a)^(2)) + fc_m(abs(abs(80-azi)-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(80-azi)-180)/10+1,5))*1000;
% determine the coefficients of shelving filter
[B(:,1), A(:,1)] = shelving(f_c*(0.0875/a), g_inf, 48000);

hl = freqz(B(:,1),A(:,1), (f/48000)*2*pi);
[F(:,1),E(:,1)] = yulewalk(20,[0 2*f/48000],[hl(1) hl]);

% create right shelving filter
% Eq.(2.5) right
g_0(2) = (g0_m(abs(abs(280-azi)-180)/10+1,1)*(r(i)/a) + g0_m(abs(abs(280-azi)-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs(280-azi)-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(280-azi)-180)/10+1,4));
% Eq.(2.9) g_inf
g_inf = (ginf_m(abs(abs(280-azi)-180)/10+1,1)*(r(i)/a) + ginf_m(abs(abs(280-azi)-180)/10+1,2)) / (((r(i)/a)^(2)) + ginf_m(abs(abs(280-azi)-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(280-azi)-180)/10+1,4));
% Eq.(2.10) f_c
f_c = ((fc_m(abs(abs(280-azi)-180)/10+1,1)*((r(i)/a)^(2)) + fc_m(abs(abs(280-azi)-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs(280-azi)-180)/10+1,3)) / ((r(i)/a)^(2)) + fc_m(abs(abs(280-azi)-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(280-azi)-180)/10+1,5))*1000;
% determine the coefficients of shelving filter
[B(:,2), A(:,2)] = shelving(f_c*(0.0875/a), g_inf, 48000);

hr = freqz(B(:,2),A(:,2), (f/48000)*2*pi);
[F(:,2),E(:,2)] = yulewalk(20,[0 2*f/48000],[hr(1) hr]);

else
% linear interpolation for angles 45,135,225,315

% create left shelving filter
% Eq.(2.5) left

```

```

g_0a(1) = (g0_m(abs(abs(80-(azi-5))-180)/10+1,1)*(r(i)/a) + g0_m(abs(
abs(80-(azi-5))-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs
(80-(azi-5))-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(80-(azi-5))
-180)/10+1,4));
g_0b(1) = (g0_m(abs(abs(80-(azi+5))-180)/10+1,1)*(r(i)/a) + g0_m(abs(
abs(80-(azi+5))-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs
(80-(azi+5))-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(80-(azi+5))
-180)/10+1,4));
g_0(1) = (g_0a(1) + g_0b(1))/2;
% Eq.(2.9) g_inf
g_infa = (ginf_m(abs(abs(80-(azi-5))-180)/10+1,1)*(r(i)/a) + ginf_m(
abs(abs(80-(azi-5))-180)/10+1,2)) / (((r(i)/a)^(2)) + ginf_m(abs(
abs(80-(azi-5))-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(80-(azi-5))
)-180)/10+1,4));
g_infb = (ginf_m(abs(abs(80-(azi+5))-180)/10+1,1)*(r(i)/a) + ginf_m(
abs(abs(80-(azi+5))-180)/10+1,2)) / (((r(i)/a)^(2)) + ginf_m(abs(
abs(80-(azi+5))-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(80-(azi+5))
)-180)/10+1,4));
g_inf = (g_infa + g_infb)/2;
% Eq.(2.10) f_c
f_ca = ((fc_m(abs(abs(80-(azi-5))-180)/10+1,1)*((r(i)/a)^(2)) + fc_m
(abs(abs(80-(azi-5))-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs(80-(azi
-5))-180)/10+1,3)) / (((r(i)/a)^(2)) + fc_m(abs(abs(80-(azi-5))
-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(80-(azi-5))-180)/10+1,5)))
*1000;
f_cb = ((fc_m(abs(abs(80-(azi+5))-180)/10+1,1)*((r(i)/a)^(2)) + fc_m
(abs(abs(80-(azi+5))-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs(80-(azi
+5))-180)/10+1,3)) / (((r(i)/a)^(2)) + fc_m(abs(abs(80-(azi+5))
-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(80-(azi+5))-180)/10+1,5)))
*1000;
f_c = (f_ca + f_cb)/2;
% determine the coefficients of shelving filter
[B(:,1), A(:,1)] = shelving(f_c*(0.0875/a), g_inf, 48000);

hl = freqz(B(:,1),A(:,1), (f/48000)*2*pi);
[F(:,1),E(:,1)] = yulewalk(20, [0 2*f/48000], [hl(1) hl]);

% right shelving filter
% Eq.(2.5) right
g_0a(2) = (g0_m(abs(abs(280-(azi-5))-180)/10+1,1)*(r(i)/a) + g0_m(abs
(abs(280-(azi-5))-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs
(280-(azi-5))-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(280-(azi-5))
-180)/10+1,4));
g_0b(2) = (g0_m(abs(abs(280-(azi+5))-180)/10+1,1)*(r(i)/a) + g0_m(abs
(abs(280-(azi+5))-180)/10+1,2)) / (((r(i)/a)^(2)) + g0_m(abs(abs
(280-(azi+5))-180)/10+1,3)*(r(i)/a) + g0_m(abs(abs(280-(azi+5))
-180)/10+1,4));
g_0(2) = (g_0a(2) + g_0b(2))/2;
% Eq.(2.9) g_inf
g_infa = (ginf_m(abs(abs(280-(azi-5))-180)/10+1,1)*(r(i)/a) + ginf_m(
abs(abs(280-(azi-5))-180)/10+1,2)) / (((r(i)/a)^(2)) + ginf_m(abs
(abs(280-(azi-5))-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(280-(azi
-5))-180)/10+1,4));

```

```

g_infb = (ginf_m(abs(abs(280-(azi+5))-180)/10+1,1)*(r(i)/a) + ginf_m(
    abs(abs(280-(azi+5))-180)/10+1,2)) / ((r(i)/a)^(2)) + ginf_m(abs
    (abs(280-(azi+5))-180)/10+1,3)*(r(i)/a) + ginf_m(abs(abs(280-(azi
    +5))-180)/10+1,4));
g_inf = (g_infa + g_infb)/2;
% Eq. (2.10) f_c
f_ca = ((fc_m(abs(abs(280-(azi-5))-180)/10+1,1)*((r(i)/a)^(2)) +
    fc_m(abs(abs(280-(azi-5))-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs
    (280-(azi-5))-180)/10+1,3)) / ((r(i)/a)^(2)) + fc_m(abs(abs(280-(
    azi-5))-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(280-(azi-5))-180)
    /10+1,5))*1000;
f_cb = ((fc_m(abs(abs(280-(azi+5))-180)/10+1,1)*((r(i)/a)^(2)) +
    fc_m(abs(abs(280-(azi+5))-180)/10+1,2)*(r(i)/a) + fc_m(abs(abs
    (280-(azi+5))-180)/10+1,3)) / ((r(i)/a)^(2)) + fc_m(abs(abs(280-(
    azi+5))-180)/10+1,4)*(r(i)/a) + fc_m(abs(abs(280-(azi+5))-180)
    /10+1,5))*1000;
f_c = (f_ca + f_cb)/2;
% determine the coefficients of shelving filter
[B(:,2), A(:,2)] = shelving(f_c*(0.0875/a), g_inf, 48000);

hr = freqz(B(:,2),A(:,2), (f/48000)*2*pi);
[F(:,2),E(:,2)] = yulewalk(20,[0 2*f/48000],[hr(1) hr]);

end
end

```

## A.8 random\_sort

```

function new_vect = random_sort(vect)

% Randomizing algorithm

[r,c] = size(vect);
if (r==1) vect_len = c;
else vect_len = r;
end

rem = vect_len;
new_vect = zeros(1,vect_len);

for i = 1:vect_len
    indx = round(0.5+rem*rand(1));
    new_vect(i) = vect(indx);
    vect(indx) = vect(rem);
    rem = rem-1;
end

end

```

## A.9 run\_exp

```

function run_exp(subject_id,acm,status)

% Algorithm to run the experiment:
% INPUT:
% subject_id    - 3-character string (starting from 001)
% acm           - head radius (cm)
% status        - 0 -> new subject, calculate hrtfs
%               0 < n <= 10 -> new starting point for the experiment
%               in case of troubles during the execution
% OUTPUT: id_workspace containing the following structures
% res_table     - table of results(correct/wrong answers)
% rep           - order of the stimulus couples during the 10 repetitions
% flags         - stimulus order within each couple
% I             - internal cycle counter (for restoring in case of troubles)

hold off;
close all;

%initialization
if (status == 0)
    create_hrtfs(acm);

    % couple number = 5 distance couples * 5 azi * 4 conditions * 10
    repetitions = 1000
    rep = zeros(100,10);
    rep(:,1) = random_sort(1:100); % 1st repetition order
    for i = 2:10
        rep(:,i) = random_sort(rep(:,i-1)); % repetition order from 2 to 10
    end

    % assign balanced flags for stimulus order (5 in one direction, 5 in the
    opposite)
    flags = NaN(100,10);
    for r = 1:5
        flags(:,2*r-1) = round(rand(100,1));
        flags(:,2*r) = 1-flags(:,2*r-1);
    end

    % table order: first distance couples, second azimuth, last conditions
    res_table = NaN(100,13);
    res_table(:,1) = mod(0:99,5)+1; % first column: identify distance couples
    (1-5)
    res_table(:,2) = mod(round(((0:99)/5)-0.49)*45-225,225)+90; % second
    column: azimuth
    res_table(:,3) = round(((0:99)/25)-0.49)+1; % third column: id condition

    % initial instructions
    voice = wavread('./TTS/instr1.wav');
    wavplay(voice./5, 48000);

```

```

    voice = wavread('./TTS/instr2.wav');
    wavplay(voice./5, 48000);
    voice = wavread('./TTS/instr3.wav');
    wavplay(voice./5, 48000);

else
    load(['./results/' subject_id '.mat']);
end

% confirm the start of the experiment
disp('When ready press a button');
voice = wavread('./TTS/ready.wav');
wavplay(voice./5, 48000);
waitforbuttonpress;

% identify the repetition to start with
if (status == 0)
    start = 1;
    status = 1;
else
    start = mod(I+1,100);
end

% external cycle (repetition)
for r = status:10
    if (mod(r,2) && start == 1)
        voice = wavread(['./TTS/session' int2str((r+1)/2) '.wav']);
        wavplay(voice./5, 48000);
    end
    % internal cycle (100 stimulus couples)
    for i = start:100
        dist_pair = res_table(rep(i,r),1);
        azi = res_table(rep(i,r),2);
        set = res_table(rep(i,r),3);
        flag = flags(rep(i,r),r);
        if flag == 0
            [y, fs] = wavread(['./set_' int2str(set) '/dist' int2str(
                dist_pair) '_azi' int2str(azi) '.wav']);
            pause(0.5);
            wavplay(y,fs);
            [y, fs] = wavread(['./set_' int2str(set) '/dist' int2str(
                dist_pair+1) '_azi' int2str(azi) '.wav']);
            pause(0.3);
            wavplay(y,fs);
            waitforbuttonpress;
            ch = get(gcf,'CurrentCharacter');
            if strncmp(ch,'a',1)==1
                k = 0;
            else
                k = 1;
            end
            res_table(rep(i,r),3+r) = k;
        else

```

```

[y, fs] = wavread(['./set-' int2str(set) '/dist' int2str(
    dist_pair+1) '_azi' int2str(azi) '.wav']);
pause(0.5);
wavplay(y,fs);
[y, fs] = wavread(['./set-' int2str(set) '/dist' int2str(
    dist_pair) '_azi' int2str(azi) '.wav']);
pause(0.3);
wavplay(y,fs);
waitforbuttonpress;
ch = get(gcf, 'CurrentCharacter');
if strncmp(ch, 'a', 1) == 1
    k = 1;
else
    k = 0;
end
res_table(rep(i,r), 3+r) = k;
end
% update variables and save workspace at each cycle iteration
I = i;
save(['./results/' subject_id '.mat'], 'rep', 'flags', 'I', '
    res_table');
end
% pause between each experiment block
if (~mod(r,2) && r < 9)
    close all;
    voice = wavread(['./TTS/session' int2str(r/2) 'over.wav']);
    wavplay(voice./5, 48000);
    disp('Take a short break');
    voice = wavread(['./TTS/break.wav']);
    wavplay(voice./5, 48000);
    countdown(3,0, 'The end', 0);
    pause(181);
    close all;
    disp('When ready press a button');
    voice = wavread(['./TTS/ready.wav']);
    wavplay(voice./5, 48000);
    waitforbuttonpress;
end
start=1;
end

% end experiment
voice = wavread(['./TTS/the end.wav']);
wavplay(voice./5, 48000);
close all;

end

```

## A.10 countdown

```

function countdown(mins,secs,endMssg,showIllusion)
% Copyright 2007 The MathWorks, Inc.
% Creates a figure to countdown remaining time (minutes/seconds only).
%
% By default, the figure includes an optical illusion created with lines
% and patches. This can be turned off.
%
% SYNTAX:
%   COUNTDOWN(MINS)
%   COUNTDOWN(MINS,SECS)
%   COUNTDOWN(MINS,SECS,ENDMSSG)
%   COUNTDOWN(MINS,SECS,ENDMSSG,SHOWILLUSION)
%   (SHOWILLUSION is a binary input to SHOW (1=default) or SUPPRESS (0)
%   optical illusion.)
%
% EXAMPLE:
%   %1) 15 minutes
%   countdown(15)
%
%   %2) 15 minutes, 30 seconds, then flash message 'Pencils Down!!!'
%   countdown(15,30,'Pencils Down!!!')
%   % Note: same as countdown(15.5,0,'Pencils Down!!!')
%
%   %3) 5 minutes, 0 seconds; default ending message; suppress optical
%   illusion
%   countdown(5,0,[],0)

% Created by Brett Shoelson
% brett.shoelson@mathworks.com
% 12/31/07

% Revisions:
% 01/01/08 Added new input argument (in position 3) to allow easy setting of
% ending message. Also changed default ending message, and added
% examples.)

if nargin < 2
    secs = 0;
end
secs = secs + rem(mins,1)*60;
mins = floor(mins);

if nargin < 3 || isempty(endMssg)
    endMssg = 'Time's Up!';
end

if nargin < 4
    showIllusion = 1; %DEFAULT
end

```

```

countdownfig = figure('numbertitle','off','name','COUNTDOWN',...
    'color','w','menubar','none','toolbar','none',...
    'pos',[65 65 400 400],'closerequestfcn',@cleanup);

if showIllusion
    createIllusion;
    edtpos = [0.1 0.75 0.8 0.2];
else
    edtpos = [0.1 0.1 0.8 0.8];
end
edtbox = uicontrol('style','edit','string','STARTING','units','normalized',
    ...
    'position',edtpos,'fontsize',62,'foregroundcolor','r','backgroundcolor',
    'k');
timerobj = timer('timerfcn',@updateDisplay,'period',1,'executionmode','
    fixedrate');
secsElapsed = 0;
start(timerobj);

function updateDisplay(varargin)
    secsElapsed = secsElapsed + 1;
    if secsElapsed > secs + mins*60
        set(edtbox,'string',endMssg);
        tmp = get(0,'screensize');
        set(countdownfig,'pos',[65 65 400 400]);
        set(edtbox,'foregroundcolor',1-get(edtbox,'foregroundcolor')); %
            , 'backgroundcolor',1-get(edtbox,'backgroundcolor'))
    else
        set(edtbox,'string',...
            datestr([2003 10 24 12 mins secs-secsElapsed],'MM:SS'));
    end
end

function cleanup(varargin)
    stop(timerobj);
    delete(timerobj);
    closereq;
end

function createIllusion(varargin)
    %%%%%%%%%%%
    %PARAMETERS
    pct = 0.7;%Vertical size of axis (Pct of figure)
    nlines = 9;%Number of lines
    fc = [0 0 0];%Facecolor
    lc = [1 1 1]*0.5;%Linecolor
    sz = 0.04;%Patchwidth
    lw = 2; %Linewidth
    %%%%%%%%%%%

    subplot('position',[0 0 1 pct]);
    y = 1:nlines;

```

```

axis off;
hold on;
xs = [0 1 1 0 0, reshape(repmat(reshape([3:2:floor(1/sz);2:2:floor(1/
    sz)-1],[],1),1,2)', [],1)']*sz;
ys = [0 0 1 1 0, repmat([0 1 1 0],1, (numel(xs)-5)/4)];
h = zeros(1,nlines);
for jj=1:nlines
    h(jj) = patch(xs,ys+jj,fc);
end
line(xlim,[y;y], 'color',lc, 'linewidth',lw);
shifts = repmat([1 2 3 2],1,3);
for ii = 1:max(y)
    set(h(ii), 'xdata',get(h(ii), 'xdata')+ sz/shifts(ii));
end
set(gca, 'xlim',[0 1]);
end
end

```

## A.11 error\_count

```

function rates = error_count(subject)

% INPUT:
% subject      - Subject ID
% OUTPUT:
% Error rate by distance pair
% Error rate by azimuth
% Error rate by condition
% Error rate by stimulus order
% Error rate by distance pair, only upward sequences
% Error rate by azimuth, only upward sequences
% Error rate by condition, only upward sequences
% Error rate by distance pair, only downward sequences
% Error rate by azimuth, only downward sequences
% Error rate by condition, only downward sequences

file_ws = [ './results/' subject '.mat'];
load(file_ws);
rates = zeros(4,5);

res_table_up = zeros(100,5);
res_table_down = zeros(100,5);
for c=1:10
    for i=1:100
        if flags(i,c) == 0
            res_table_down(i,ceil(c/2)) = res_table(i,c+3);
        else
            res_table_up(i,ceil(c/2)) = res_table(i,c+3);
        end
    end
end

```

```

    end
end

d = zeros(1,5);
for i=1:20
    for j = 1:5
        d(j) = d(j) + sum(res_table((i-1)*5+j,4:13));
    end
end
disp('Error rate by distance pair:');
disp((200-d)/2);
rates(1,:) = (200-d)/2;

a = zeros(1,5);
for i=1:4
    for j = 1:5
        a(j) = a(j) + sum(sum(res_table((i-1)*25+(j-1)*5+1:(i-1)*25+(j-1)
            *5+5,4:13)));
    end
end
disp('Error rate by azimuth:');
disp((200-a)/2);
rates(2,:) = (200-a)/2;

c = [sum(sum(res_table(1:25,4:13))), sum(sum(res_table(26:50,4:13))), sum(sum
    (res_table(51:75,4:13))), sum(sum(res_table(76:100,4:13)))];
disp('Error rate by condition:');
disp((250-c)/2.5);
rates(3,1:4) = (250-c)/2.5;

err0 = 1-sum(sum((1-flags).*res_table(:,4:13))/sum(sum(1-flags)));
err1 = 1-sum(sum(flags.*res_table(:,4:13))/sum(sum(flags)));
disp('Error rate by stimulus order:');
disp([err0*100 err1*100]);

%when the sequence is downward and when upward (in percent)
rates(4,1:2) = [err0*100 err1*100];

%upward sequences

d = zeros(1,5);
for i=1:20
    for j = 1:5
        d(j) = d(j) + sum(res_table_up((i-1)*5+j,:));
    end
end
disp('Error rate by distance pair, only upward sequences:');
disp(100-d);

a = zeros(1,5);
for i=1:4
    for j = 1:5

```

```

        a(j) = a(j) + sum(sum(res_table_up((i-1)*25+(j-1)*5+1:(i-1)*25+(j-1)
            *5+5,:)));
    end
end
disp('Error rate by azimuth, only upward sequences:');
disp(100-a);

c = [sum(sum(res_table_up(1:25,:))), sum(sum(res_table_up(26:50,:))), sum(sum
    (res_table_up(51:75,:))), sum(sum(res_table_up(76:100,:)))]];
disp('Error rate by condition, only upward sequences:');
disp((125-c)/1.25);

%downward sequences

d = zeros(1,5);
for i=1:20
    for j = 1:5
        d(j) = d(j) + sum(res_table_down((i-1)*5+j,:));
    end
end
disp('Error rate by distance pair, only downward sequences:');
disp(100-d);

a = zeros(1,5);
for i=1:4
    for j = 1:5
        a(j) = a(j) + sum(sum(res_table_down((i-1)*25+(j-1)*5+1:(i-1)*25+(j
            -1)*5+5,:)));
    end
end
disp('Error rate by azimuth, only downward sequences:');
disp(100-a);

c = [sum(sum(res_table_down(1:25,:))), sum(sum(res_table_down(26:50,:))), sum
    (sum(res_table_down(51:75,:))), sum(sum(res_table_down(76:100,:)))]];
disp('Error rate by condition, only downward sequences:');
disp((125-c)/1.25)

end

```



# Bibliografia

- [1] W. H. Ahrens, D. J. Cox, and G. Budhwar. Use of the arcsine and square root transformations for subjectively determined percentage data. *Weed*, (38):452–458, 1990.
- [2] V. R. Algazi, C. Avendano, and R. O. Duda. Estimation of a spherical-head model from anthropometry. *J. Audio Eng. Soc.*, 49(6):472–479, June 2001.
- [3] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang. Approximating the head-related transfer function using simple geometric models of the head and torso. *J. Acoust. Soc. Am.*, 112(5):2053–2064, November 2002.
- [4] D. H. Ashmead, D. LeRoy, and R. D. Odom. Perception of the relative distances of nearby sound sources. *Percept. Psychophys.*, 47(4):326–331, April 1990.
- [5] F. Avanzini and G. D. Poli. *Algorithms for sound and music computing*. 2010.
- [6] D. R. Begault, E. M. Wenzel, and M. R. Anderson. Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. *J. Audio Eng. Soc.*, 49(10):904–916, October 2001.
- [7] C. P. Brown and R. O. Duda. A structural model for binaural sound synthesis. *IEEE Trans. Speech Audio Process.*, 6(5):476–488, September 1998.
- [8] D. S. Brungart. Auditory localization of nearby sources. III. Stimulus effects. *J. Acoust. Soc. Am.*, 106(6):3589–3602, December 1999.
- [9] D. S. Brungart, N. I. Durlach, and W. M. Rabinowitz. Auditory localization of nearby sources. II. Localization of a broadband source. *J. Acoust. Soc. Am.*, 106(4):1956–1968, October 1999.
- [10] D. S. Brungart and B. D. Simpson. Auditory localization of nearby sources in a virtual audio display. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, pages 107–110, New Paltz, New York, USA, October 2001.
- [11] D. Doukhan and A. Sédès. CW\_binaural~: A binaural synthesis external for Pure Data. In *Proc. 3rd Puredata Int. Conv. (PdCon09)*, São Paulo, Brazil, July 2009.

- [12] R. O. Duda and W. L. Martens. Range dependence of the response of a spherical head model. *J. Acoust. Soc. Am.*, 104(5):3048–3058, November 1998.
- [13] M. Geronazzo, S. Spagnol, and F. Avanzini. A head-related transfer function model for real-time customized 3-D sound rendering. In *Proc. INTERPRET Work., SITIS 2011 Conf.*, pages 174–179, Dijon, France, November-December 2011.
- [14] M. Geronazzo, S. Spagnol, and F. Avanzini. Mixed structural modeling of head-related transfer functions for customized binaural audio delivery. In *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, Santorini, Greece, July 2013.
- [15] D. J. Kistler and F. L. Wightman. A model of head-related transfer functions based on principal components analysis and minimum-phase reconstruction. *J. Acoust. Soc. Am.*, 91(3):1637–1647, March 1992.
- [16] D. H. Mershon and J. N. Bowers. Absolute and relative cues for the auditory perception of egocentric distance. *Perception*, 8(3):311–322, 1979.
- [17] L. Mion, F. Avanzini, B. Mantel, B. Bardy, and T. A. Stoffregen. Real-time auditory-visual distance rendering for a virtual reaching task. In *Proc. ACM VRST 2007*, pages 179–182, 2007.
- [18] J. R. Pierce. *The Science of Musical Sound*. W.H. Freeman and Company, NY, 1983.
- [19] T. Qu, Z. Xiao, M. Gong, Y. Huang, X. Li, and X. Wu. Distance-dependent head-related transfer functions measured with high spatial resolution using a spark gap. *IEEE Trans. Audio, Speech, Lang. Process.*, 17(6):1124–1132, August 2009.
- [20] S. Scaiella. Valutazione parametrica di un modello strutturale di orecchio esterno per il rendering spaziale del suono. Master’s thesis, University of Padova, 2014.
- [21] Z. Schärer and A. Lindau. Evaluation of equalization methods for binaural signals. In *Proc. 126th Conv. Audio Eng. Soc.*, Munich, Germany, May 2009.
- [22] G. Schiesaro. Registrazione ed analisi di headphone impulse response per la spazializzazione del suono. Bachelor’s Thesis, University of Padova, 2013.
- [23] S. Spagnol and F. Avanzini. Real-time binaural audio rendering in the near field. In *Proc. 6th Int. Conf. Sound and Music Computing (SMC09)*, pages 201–206, Porto, Portugal, July 2009.
- [24] S. Spagnol, M. Geronazzo, and F. Avanzini. Hearing distance: A low-cost model for near-field binaural effects. In *Proc. EUSIPCO 2012 Conf.*, pages 2005–2009, Bucharest, Romania, September 2012.
- [25] S. Spagnol, M. Geronazzo, and F. Avanzini. On the relation between pinna reflection patterns and head-related transfer function features. *IEEE Trans. Audio, Speech, Lang. Process.*, 21(3):508–519, March 2013.

- [26] S. Spagnol, M. Geronazzo, D. Rocchesso, and F. Avanzini. Extraction of pinna features for customized binaural audio delivery on mobile devices. In *Proc. 11th Int. Conf. Advances Mobile Comput. & Multim. (MoMM 2013)*, pages 514–517, Vienna, Austria, December 2013.
- [27] G. Stewart. The acoustic shadow of a rigid sphere with certain applications in architectural acoustics and auditions. *Phys. Rev.*, (33):467–479, 1911.
- [28] G. Stewart. Phase relations in the acoustic shadow of a rigid sphere: Phase difference at the ears. *Phys. Rev.*, (34):252–258, 1911.
- [29] E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman. Localization using nonindividualized head-related transfer functions. *J. Acoust. Soc. Am.*, 94(1):111–123, July 1993.
- [30] F.L. Wightman and D.J. Kistler. The dominant role of low-frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.*, 91(3):1648–1661, 1992.
- [31] P. Zahorik, D. S. Brungart, and A. W. Bronkhorst. Auditory distance perception in humans: a summary of past and present research. *Acta Acustica united with Acustica*, 91(3):409–420, May/June 2005.
- [32] U. Zölzer, editor. *DAFX - Digital Audio Effects*. J. Wiley & Sons, New York, NY, USA, 2002.
- [33] D. N. Zotkin, R. Duraiswami, and L. S. Davis. Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia*, 6(4):553–564, 2004.