

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE  
LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

**PERSONALIZZAZIONE DI PROFILI  
ACUSTICI PER LA REALTÀ AUMENTATA  
IN CAMPO MOBILE**

*Laureando:*  
Sorato Giacomo

*Relatore:*  
Ch.mo Prof.  
Federico Avanzini

*Matricola:*  
1086981

*Correlatore:*  
Dott. Michele Geronazzo

Anno Accademico 2015/2016  
5/12/2016



# Prefazione

Quando ci accingiamo ad ascoltare con le cuffie un brano musicale registrato in studio possiamo accorgerci di come la relazione tra suono e spazio sia marcatamente artificiale. Se ascoltassimo la registrazione di un concerto e la relazionassimo all'ascolto live, ci accorgeremmo ancora meglio di quanto questa relazione tra suono e spazio non venga acquisita dai più comuni metodi di registrazione. Vi sono tuttavia delle tecniche che mirano a realizzare una scena sonora coinvolgente, capace di far provare una sensazione di totale immersione nell'ambiente acustico desiderato. A tal fine, la tecnologia dell'audio binaurale È stata una delle possibilità maggiormente sviluppata negli ultimi 20 anni. Una registrazione binaurale viene effettuata mediante l'uso di una simulazione di testa antropomorfa munita di due microfoni inseriti nei padiglioni auricolari. Il suono così registrato cattura le informazioni dell'ambiente e del ricevitore stesso. Avendo a disposizione entrambe queste informazioni è possibile modellare un suono virtuale in modo da renderlo come inserito in una scena acustica e spazializzato all'interno di essa.

Negli ultimi anni le tecnologie per la realtà virtuale e aumentata sono state notevolmente migliorate, tanto da farci immaginare un futuro in cui queste tecnologie pervaderanno le nostre vite diventando un prodotto di massa e di uso quotidiano.

Affinché un contesto di realtà aumentata o virtuale risulti credibile vi è la necessità che suoni sintetici vengano inseriti operando su di essi una modellazione che li renda perfettamente omogenei con la scena percepita e conceda loro una spazializzazione tridimensionale efficace. Questo lavoro di tesi ha il principale obiettivo di validare in un ambiente controllato un'applicazione Android al fine di acquisire i parametri antropometrici soggettivi di un utente per permettere la personalizzazione della spazializzazione tridimensionale. Tali parametri allo sono acquisibili professionalmente solo attraverso attrezzature difficilmente accessibili in termini di costi e di spazi, ecco perché È di grande importanza sviluppo di una applicazione lato consumer che permetta in modo semplice ed economico l'acquisizione di tali parametri. L'interazione acustica uomo e ambiente viene così acquisita dall'applicazione in diversi tipi di ambiente per individuare limiti e pregi.

Viene infine definita una procedura di auto-calibrazione per eliminare degli effetti di percezione dovuti alla colorazione introdotta dall'indossare le cuffie, con l'obiettivo

di renderle “trasparenti”. Infatti indossando le cuffie si modifica la struttura del canale uditivo che da aperto passa a chiuso (tappato dalle cuffie stesse) producendo effetti che colorano il suono. Una procedura di resta trasparente mira ad eliminare i contributi apportati dall’headset, e dagli effetti che genera indossandolo, al suono ascoltato. Avendo a disposizione i parametri ottenuti dal profilo acustico dell’ascoltatore, dell’ambiente e della cuffia, si ha che, in contesti di realtà aumentata, si può integrare un suono sintetico nella scena acustica.

# Sommario

Questa tesi è strutturata come segue.

**Capitolo 1** - In questo capitolo viene presentata una breve introduzione al campo di ricerca dello “spatial hearing”. Viene introdotta una descrizione delle componenti anatomiche che concorrono alla resa spaziale del suono e successivamente viene presentata un’introduzione alla percezione legata alla localizzazione spaziale soffermandosi sulla localizzazione nel piano orizzontale e verticale.

**Capitolo 2** - In questo capitolo si approfondisce il concetto di rendering spaziale del suono definendo la funzione head-related transfer function. Successivamente si descrivono le principali tecniche di acquisizione.

**Capitolo 3** - Il capitolo terzo introduce i principali concetti di realtà aumentata, realtà aumentata acustica e realtà aumentata acustica in campo mobile, descrivendone la composizione in termini di strumentazione necessaria per poterla sviluppare e di effetti di distorsione o colorazione acustica introdotti.

**Capitolo 4** - In questo capitolo viene descritto il funzionamento dell’applicazione mobile SelfEar il cui scopo è rendere accessibile la personalizzazione di un ascolto immersivo nell’ambito della realtà acustica aumentata attraverso l’acquisizione guidata di profili acustici.

**Capitolo 5** - In questo capitolo viene presentata una validazione acustica sperimentale dell’applicazione in ambiente controllato su manichino. Vengono definiti strumenti utilizzati, accorgimenti impiegati e procedure seguite per l’analisi delle misurazioni.

**Capitolo 6** - In questo capitolo viene presentata una validazione dell’applicazione con un soggetto umano in quattro tipi di ambiente, ideando una procedura di compensazione dell’ambiente che sia facilmente gestibile da un utente poco esperto.

**Capitolo 7** - In questo capitolo viene descritta una procedura di auto-calibrazione ideata

per condurre l'utente a definire un'equalizzazione da applicare all'headset per la resa trasparente delle cuffie.

**Capitolo 8** - In questo ultimo capitolo vengono riassunti i risultati ottenuti dai vari studi condotti in questo lavoro di tesi e proposti i più promettenti sviluppi futuri.

**Appendice A** - Contiene i principali grafici prodotti nelle sessioni sperimentali.

**Appendice B** - Contiene le porzioni di codice maggiormente rilevanti per il procesamiento delle informazioni.

# Publicazioni

Il contributo originale di questo lavoro di tesi ha dato origine alle seguenti pubblicazioni su atti di conferenze internazionali:

- Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. Acoustic selfies for extraction of external ear features in mobile audio augmented reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 23–26. ACM, 2016
- Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. The selfear project: a mobile application for low-cost pinna-related transfer function acquisition. *SMC conference*, 2016

e alle seguenti pubblicazioni su atti di conferenze nazionali:

- Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. Selfie acustiche con il progetto selfear: un'applicazione mobile per l'acquisizione a basso costo di pinna-related transfer function. *XXI CIM conference*, 2016



## Ringraziamenti

Desidero qui ricordare tutti coloro che direttamente o indirettamente hanno apportato un contributo a questo mio lavoro di tesi.

Innanzitutto ringrazio la mia famiglia: papà, mamma, e sorella; che nonostante tutte le difficoltà che ha incontrato è sempre rimasta unita, donandomi la possibilità di arrivare sino a qui.

Ringrazio il professor Avanzini, relatore, attraverso il quale ho avuto la possibilità di coronare un sogno che avevo sin dai primi anni delle superiori. Ringrazio il dottor Geronazzo, co-relatore per avermi affiancato, supportato e sopportato durante i mesi che ho passato in laboratorio, e per avermi accompagnato alla fine di questo mio lavoro di tesi.

Ringrazio i colleghi Fantin Jacopo, Baldovino Guido e Beraldo Filippo con i quali ho avuto il piacere di condividere gli spazi del laboratorio e di collaborare per la stesura di alcune pubblicazioni.

Ringrazio tutti gli amici che alla fine hanno ceduto e si sono resi disponibili per sottoporsi alle sessioni sperimentali di questa tesi: Mauro Mazzaro, Maverick Mugnol, Giulia Grillo, Daniele Muzzolin, Chiara Stevanato, Davide Scappini, Marco Del Longo, Michele Geronazzo, Alvisè Fecchio, Nicola Salmaso, Riccardo Tamai, Riccardo Valentini, Silvia Bertoldo, Simone Zennaro. Ringrazio anche tutti coloro che avrebbero avuto piacere di partecipare, ma che non sono riusciti.

Ringrazio tutti gli amici che hanno sempre creduto in me, che quotidianamente mi accompagnano e sostengono nella vita.

Ringrazio anche tutti i ragazzi che mi sopportano come animatore ed educatore per aver acceso in me un fuoco che con la sua energia ha contribuito a darmi la spinta necessaria ad arrivare alla fine di questo mio percorso.

Ringrazio il Signore per tutti i doni che mi ha fatto e per avermi guidato, attraverso le sue vie, a diventare la persona che sono oggi.

Giacomo



# Indice

<b>Prefazione</b>	<b>i</b>
<b>Sommario</b>	<b>iii</b>
<b>Pubblicazioni</b>	<b>v</b>
<b>Ringraziamenti</b>	<b>vii</b>
<b>1 Percezione e rendering binaurale del suono</b>	<b>1</b>
1.1 Spazializzazione del suono . . . . .	2
1.2 Localizzazione . . . . .	3
1.2.1 Localizzazione orizzontale . . . . .	8
1.2.2 Localizzazione verticale . . . . .	8
1.2.3 Percezione della distanza . . . . .	9
<b>2 Rendering spaziale</b>	<b>11</b>
2.1 Head-related transfer function . . . . .	11
2.2 Metodi di acquisizione di HRTF . . . . .	12
2.2.1 Modelli strutturali misti . . . . .	12
2.2.2 Modelli di analisi ed estrazione . . . . .	12
2.2.3 Modelli di selezione . . . . .	14
2.2.4 Modelli di sintesi . . . . .	15
<b>3 Realtà aumentata acustica</b>	<b>17</b>
3.1 Headset . . . . .	18
3.1.1 Mixer . . . . .	18
3.1.2 Display acustici virtuali . . . . .	20
3.2 Realtà aumentata acustica in campo mobile . . . . .	21

3.2.1	La sfida dell'esternalizzazione . . . . .	22
3.2.2	Acustica del canale uditivo chiuso . . . . .	22
3.2.3	Effetto di occlusione . . . . .	23
3.2.4	Effetto di isolamento . . . . .	24
3.2.5	Dispersione . . . . .	24
<b>4</b>	<b>Il progetto SelfEar</b>	<b>27</b>
4.1	Un'applicazione mobile . . . . .	28
4.2	Funzionamento dell'App . . . . .	30
<b>5</b>	<b>Validazione del progetto SelfEar</b>	<b>35</b>
5.1	Hardware . . . . .	35
5.2	Setup . . . . .	39
5.3	Misurazioni . . . . .	40
5.3.1	Diffuse field . . . . .	40
5.3.2	Sessioni di misurazione . . . . .	41
5.4	Analisi . . . . .	41
5.5	Risultati . . . . .	42
5.6	Conclusioni . . . . .	45
<b>6</b>	<b>Interazione uomo-macchina in Selfear</b>	<b>47</b>
6.1	Hardware . . . . .	47
6.2	Misurazioni . . . . .	48
6.2.1	Setup . . . . .	48
6.2.2	Sessioni sperimentali . . . . .	49
6.2.3	Analisi . . . . .	51
6.3	Risultati . . . . .	53
6.4	Discussione . . . . .	55
<b>7</b>	<b>Trasparentizzazione headset</b>	<b>57</b>
7.1	Hardware . . . . .	57
7.2	Elaborazione del segnale . . . . .	58
7.2.1	Segnali . . . . .	58
7.2.2	Pure Data . . . . .	58
7.2.3	Filtri e implementazione . . . . .	59
7.3	Protocollo di comunicazione: OSC . . . . .	59
7.4	Procedura guidata . . . . .	60
7.4.1	GUI di Matlab . . . . .	62
7.4.2	Patch di Pure Data . . . . .	66

7.5	Sessioni sperimentali . . . . .	67
7.5.1	Prima sessione sperimentale . . . . .	67
7.5.2	Seconda sessione sperimentale . . . . .	68
7.6	Soggetti . . . . .	69
7.7	Risultati . . . . .	69
7.8	Conclusioni . . . . .	71
<b>8</b>	<b>Conclusioni</b>	<b>77</b>
	<b>Appendici</b>	<b>81</b>
<b>A</b>	<b>Risultati</b>	<b>81</b>
A.1	Risultati Cap 6 . . . . .	81
A.2	Risultati Cap 7 . . . . .	91
	<b>Appendici</b>	<b>104</b>
<b>B</b>	<b>Codice</b>	<b>105</b>
B.1	Matlab . . . . .	105
B.2	preProcessing.m . . . . .	105
B.3	postProcessing.m . . . . .	107
B.4	Filtri . . . . .	107
B.4.1	HPfilter.m . . . . .	107
B.4.2	Pfilter.m . . . . .	108
B.4.3	Nfilter.m . . . . .	109
B.5	hrirsElaboration.m . . . . .	110
B.6	Ambiente.m . . . . .	111
B.7	DfVariation.m . . . . .	115
B.8	Reciprocity.m . . . . .	118
B.9	SubjectContribute.m . . . . .	121
B.10	KemarContribute.m . . . . .	124



# Elenco delle figure

1.1	Sistema di riferimento . . . . .	2
1.2	Visualizzazione grafica di ITD ed ILD . . . . .	4
1.3	Stima ITD semplificata . . . . .	5
1.4	Modulo sfera . . . . .	6
1.5	Orecchio esterno . . . . .	7
1.6	Torso . . . . .	7
2.1	Grafici dell'intensità per due esempi di HRTF in dipendenza dell'elevazione $\Phi$ (a) e dell'azimuth $\theta$ (b). . . . .	12
2.2	Acquisizione di HRTF in una camera anecoica tramite un dispositivo professionale . . . . .	13
2.3	Misure antropometriche della pinna . . . . .	15
3.1	Diagramma di un sistema di realtà aumentata acustica . . . . .	19
3.2	Funzione di trasferimento senza equalizzazione. . . . .	20
3.3	Funzione di trasferimento con equalizzazione. . . . .	20
3.4	Curva di equalizzazione generica, generata a partire dalle informazioni prese da tutti e quattro i soggetti di test. [4, 5] . . . . .	21
4.1	Percorso dell'impulso dagli altoparlanti del dispositivo ai microfoni degli auricolari attraverso le trasformazioni acustiche coinvolte nel fenomeno di riproduzione. . . . .	28
4.2	Visione schematica del progetto SelfEat in un sistema mAAR. . . . .	29
4.3	Diagramma a blocchi per la procedura di acquisizione delle risposte allo sweep. . . . .	33
5.1	Cabina silente: Sound Station Pro 45 . . . . .	36
5.2	Caratteristiche della Sound Station Pro 45 . . . . .	36
5.3	Motu 896mk3 . . . . .	37
5.4	Cuffiette Roland CS-10EM . . . . .	38
5.5	Adattatore A3 Soundman . . . . .	38

5.6	KEMAR . . . . .	38
5.7	Risposta in frequenza del Loudspeaker Genelect 8030A . . . . .	39
5.8	Setup per le misurazioni. Posizioni di sorgente e ricevitore. . . . .	39
5.9	Setup di misurazione di Selfear, con selfie-stick incorporato. . . . .	39
5.10	Struttura per calcolo diffuse field con Roland . . . . .	41
5.11	Confronto tra i moduli (in dB SPL) delle BRTF (linee spesse) e le relative PRTF (linee sottili) ottenute usando: come ricevitore - il microfono destro delle cuffie Roland (H); come sorgente - il loudspeaker dello smartphone (S, linee tratteggiate) e il loudspeaker Genelec (L, linee continue). . . . .	42
5.12	Moduli delle PRTF per i dieci riposizionamenti della cuffia Roland sul manichino KEMAR. La linea spessa rappresenta l'ampiezza media. La deviazione standard è spostata di -60 dB per agevolarne la lettura. . . .	43
5.13	Confronto tra i moduli delle seguenti PRTF: a) PRTF media della Fig.5; b) sorgente: smartphone - ricevitore: microfoni delle cuffie ; c) sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio destro della KEMAR (con canale uditivo); d) sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio sinistro della KEMAR (senza canale uditivo). . . . .	44
5.14	PRTF nel piano mediano. (a) acquisizione con SelfEar senza compensazione; (b) acquisizione con SelfEar compensata con diffuse-field; (c) CIPIC KEMAR, Subject 165 - con compensazione free-field. Vengono visivamente identificati i principali picci (P1-2) e notch (N1-3), ove presenti. . . . .	45
6.1	Scheda audio Fast Track Pro . . . . .	48
6.2	Setup: (a) supporto per lo smartphone; (b) supporto per le Roland; . . .	49
6.3	Procedura di acquisizione di diffuse field . . . . .	50
6.4	Patch per la registrazione dei segnali di SelfEar . . . . .	51
6.5	Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, in una posizione (straight) e nella posizione opposta (opposit) per entrambi canali (left e right). Smartphone: Zenfone 2 . . . . .	52
6.6	Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, con la presenza della KEMAR e senza. Smartphone: Zenfone . . . . .	52

6.7	Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto e senza. Smartphone: Zenfone . . . . .	53
6.8	Misurazioni effettuate in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto. Smartphone: Wiko. Smartphone attaccato al corpo . . . . .	53
6.9	Misurazioni effettuate in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto. Smartphone: Wiko. Distanza dello smartphone dal corpo: 10 cm . . . . .	53
6.10	confronto tra le acquisizioni di tre smartphone. L'acquisizione dello smartphone HTC è quella ottenuta nel cap. 5 (quindi ci si aspetta essere migliore delle altre) . . . . .	54
6.11	Risultati relativi ad ambiente controllato . . . . .	55
7.1	Collegamenti tra i vari componenti della procedura . . . . .	61
7.2	View 1 . . . . .	62
7.3	View 2 . . . . .	63
7.4	View 3 . . . . .	65
7.5	View 4 . . . . .	65
7.6	View 5 . . . . .	66
7.7	Risultati del Soggetto 9 . . . . .	71
7.8	Patch 1 . . . . .	74
7.9	Patch 2 . . . . .	75
A.1	Risultati relativi ad ambiente controllato . . . . .	83
A.2	Misurazione diffuse field acquisite in ambiente controllato con Zenfone . . . . .	84
A.3	Misurazione diffuse field acquisite in ambiente controllato con Wiko . . . . .	84
A.4	Risultati relativi ad ambiente domestico . . . . .	85
A.5	Misurazione diffuse field acquisite in ambiente domestico con Zenfone . . . . .	86
A.6	Misurazione diffuse field acquisite in ambiente domestico con Wiko . . . . .	86
A.7	Risultati relativi ad ambiente riverberante . . . . .	87
A.8	Misurazione diffuse field acquisite in ambiente riverberante con Zenfone . . . . .	88
A.9	Misurazione diffuse field acquisite in ambiente riverberante con Wiko . . . . .	88
A.10	Risultati relativi ad ambiente esterno . . . . .	89
A.11	Misurazione diffuse field acquisite in ambiente esterno con Zenfone . . . . .	90
A.12	Misurazione diffuse field acquisite in ambiente riverberante con Wiko . . . . .	90
A.13	Risultati relativi a Soggetto 6 . . . . .	93
A.14	Risultati relativi a Soggetto 7 . . . . .	95

A.15 Risultati relativi a Soggetto 8 . . . . .	97
A.16 Risultati relativi a Soggetto 9 . . . . .	99
A.17 Risultati relativi a Soggetto 10 . . . . .	101
A.18 Risultati relativi a Soggetto 11 . . . . .	103

# Elenco delle tabelle

2.1	Principali basi di dati pubbliche di HRTF con loro caratteristiche [6] . . .	14
5.1	Distorsione spettrale tra PRTF di Figure 5.13. Tutti i valori in dB. . . .	43
6.1	Distorsione spettrale delle risposte diffuse field acquisite mediante smart- phone. . . . .	54
6.2	Frequenze centrali dei notch alle varie angolazioni . . . . .	56
7.1	Caratteristiche del loudspeaker Energy . . . . .	57
7.2	Soggetti . . . . .	69
7.3	Risultati - parametri dei filtri . . . . .	70
7.4	Confronto tra parametri trovati dai soggetti e approssimazioni dei para- metri originali. Il trattino indica che non è stato possibile quantificare un dato valore. . . . .	72



# Capitolo 1

## Percezione e rendering binaurale del suono

Il problema della realizzazione di una buona riproduzione sonora con effetto di immersione nello spazio acustico è stato affrontato nello studio di Schroeder, Gottlob e Siebrasse (1974) con costruzione di una testa di manichino dotato di orecchio (pinna e canale uditivo) e di microfoni posizionati alla fine del canale uditivo [7]. Inviando la registrazione fatta mediante i microfoni alle cuffie di un ascoltatore umano (uscita del microfono sinistro collegato alla cuffia sinistra e quello destro alla cuffia destra) si riuscivano ad approssimare i segnali nei canali uditivi dell'ascoltatore come se egli fosse effettivamente presente nel luogo dove è stata effettuata la registrazione. Questa particolare tecnica in cui due segnali, registrati mediante una struttura che emula fisicamente i connotati umani (oppure realizzata proprio su un soggetto reale), vengono inviati ad un ascoltatore umano, prende il nome di audio binaurale. Tale tecnologia ottiene dunque la massima efficienza attraverso la riproduzione in cuffia, che mantiene intatte le caratteristiche del segnale, senza riflessioni e riverberi dell'ambiente. Questo approccio però è valido soggettivamente, la registrazione effettuata per un soggetto non è detto sia compatibile con un secondo ascoltatore. Il motivo principale è che le caratteristiche anatomiche di ciascuno di noi, dove l'orecchio e la testa ne rappresentano la parte più importante, variano in forma e caratteristiche da individuo a individuo. L'effetto sarebbe quindi migliore se si riuscisse ad effettuare la registrazione direttamente con le proprie caratteristiche anatomiche (e quindi con il proprio corpo) piuttosto che con quelle di un manichino.

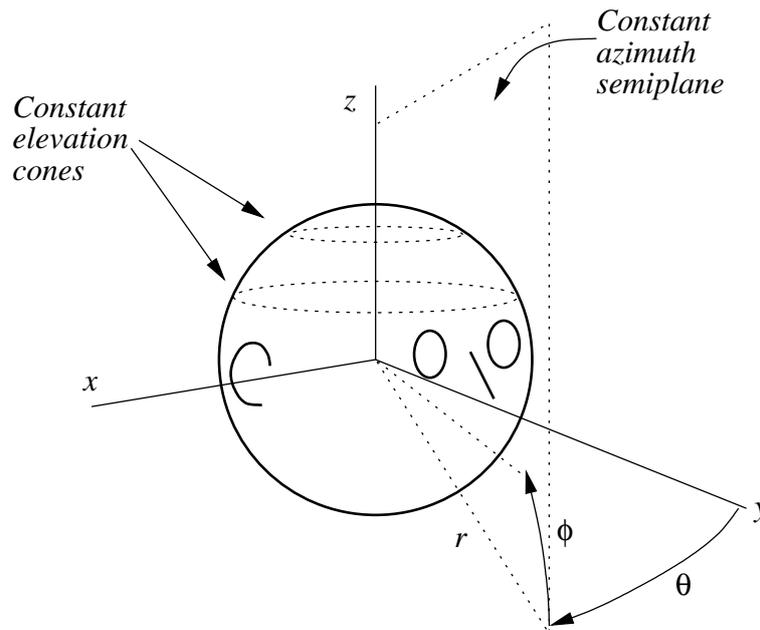


Figura 1.1: Sistema di riferimento

## 1.1 Spazializzazione del suono

Forniamo innanzitutto le basi acustiche e matematiche su cui si basano le tecnologie oggi adottate per la resa di sorgenti sonore virtuali. La via più promettente per questo scopo è basata sulle tecniche di registrazione e riproduzione binaurale, in quanto permettono di fornire le informazioni necessarie all'auralizzazione tridimensionale. Al fine di una trattazione sommaria e non eccessivamente dettagliata che si estranerebbe dalle intenzioni di questo elaborato, adotteremo la semplificazione di considerare i segnali che rappresentano la variazione di pressione acustica ai timpani<sup>1</sup> come completi di tutte le informazioni che vengono ricevute da un soggetto umano e usate per elaborare la propria percezione uditiva, comprese quelle spaziali [8]. Per semplificare le considerazioni derivanti dalla meccanica delle onde, assumiamo anche che la forma della testa sia sferica e quella del busto semiellissoidale, e che la sorgente sonora sia sufficientemente lontana da poter considerare le onde da essa generate come onde piane. La Figura 1.1

<sup>1</sup>La pressione acustica al ricevitore (Sound Pressure Level) è definita con la formula

$$SPL = 20 \log\left(\frac{P}{P_0}\right)$$

dove  $P_0$  è una pressione di riferimento, nel nostro caso  $P_0 = 210_5 \text{ Pa}$ , ovvero la pressione corrispondente alla soglia di udibilità per l'orecchio umano, e risulta uguale, in termini di intensità sonora, a

$$SPL = 10 \log\left(\frac{I}{I_0}\right)$$

con  $I_0 = 10^{12} \text{ W/m}^2$  intensità sonora di riferimento, in questo caso anch'essa corrispondente alla soglia di udibilità. la funzione  $\log$  indica il logaritmo in base 10 [8].

mostra qual è il sistema di coordinate spaziali adottato, si tratta di un sistema verticale polare dove una sorgente viene individuata dalle coordinate sferiche: azimuth  $\theta$ , elevazione  $\phi$  e distanza  $r$ . In questo sistema l'azimut  $\theta$  corrisponde all'angolo tra il piano yz, piano mediano, e il semipiano iniziante dall'asse verticale z e contenente la sorgente sonora ( $\theta \in [180, +180]$ ). Su questo semipiano, l'elevazione  $\phi$  corrisponde all'angolo tra la semiretta di intersezione con il piano orizzontale xy e la semiretta iniziante all'origine degli assi e contenente la sorgente sonora ( $\phi \in [90, 90]$ ). In questo sistema di coordinate, le superfici ad azimut costante sono dei semipiani aventi l'asse verticale z come retta delimitatrice, mentre le superfici ad elevazione costante sono coni concentrici attorno allo stesso asse. Questa esposizione concettuale porterà a capire quali sono gli strumenti utili a rendere un suono in cuffia tridimensionale e come impiegarli per la creazione di display acustici virtuali.

## 1.2 Localizzazione

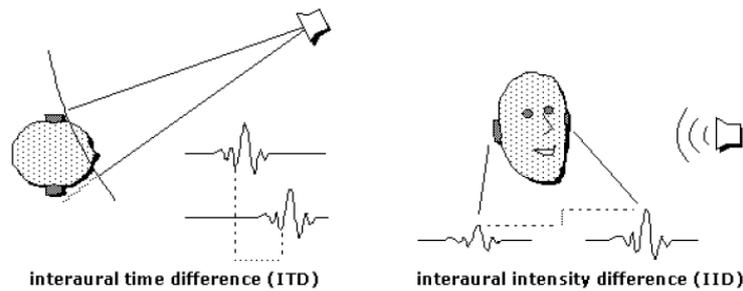
L'abilità di localizzare la sorgente di un determinato suono, elemento di grande rilievo nella vita di tutti i giorni, è determinata da due elementi principali. Il primo dovuto a fenomeni fisici associabili all'ambiente in cui il suono si propaga, descritti dalla *room impulse response* (RIR). Un esempio è dato dal riverbero che svolge un ruolo fondamentale nell'individuazione di alcune caratteristiche di una ipotetica stanza nella quale inserita la sorgente: dimensioni, tipo di materiale delle pareti e cos via; ci fornisce anche dei suggerimenti sulla distanza e direzione dalle sorgente, e un senso di immersione nel suono.

Il secondo è dovuto al contributo di parti anatomiche quali la testa, la pinna e il torso, che permettono la localizzazione anche in ambienti anecoici (ovvero in assenza di qualsivoglia riflessione ambientale dell'onda sonora). Si evidenziano quindi aspetti antropometrici dell'ascoltatore, descritti nella *head-related impulse response* (HRIR), dovuti a componenti anatomiche il cui contributo influisce fortemente nella registrazione binaurale per la precisione di spazialità che riescono a veicolare all'ascoltatore.

Le due risposte all'impulso che abbiamo descritto compongono la *binaural room impulse response* (BRIR) che da la resa dell'ambiente acustico completa.

Vediamo qual è l'influenza delle componenti anatomiche grazie alle quali è possibile localizzare le sorgenti sonore.

**Testa** Di certo gli elementi che concorrono maggiormente alla percezione del suono sono le nostre orecchie. Esse però non sono oggetti isolati nello spazio, sono posizionate ad una distanza di circa 17.5cm alla stessa altezza ai due lati della testa, che ha



**Figura 1.2:** Visualizzazione grafica di ITD ed IID

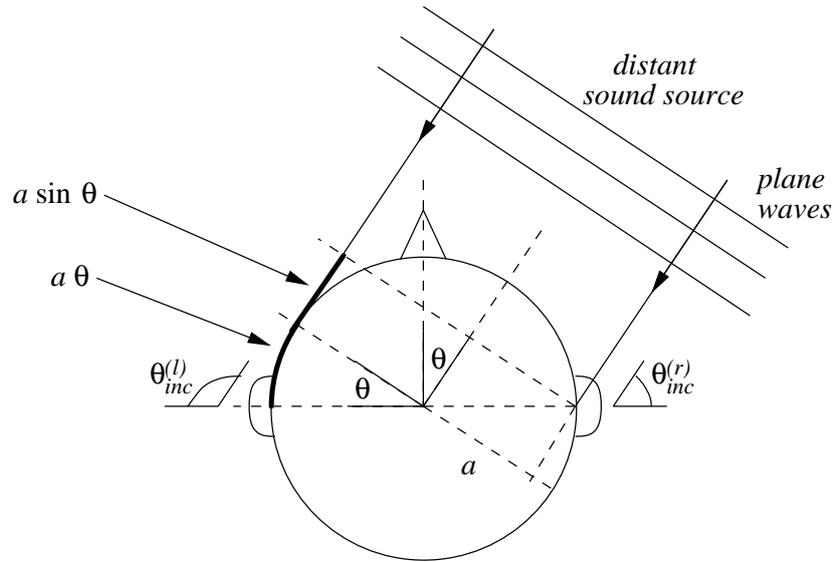
un'alta impedenza acustica. Questa costituisce un ostacolo per la libera propagazione del suono e ha due principali effetti sulla percezione sonora, mostrati schematicamente in Figura 1.2:

- Interaural Time Difference (ITD), causato dal fatto che le onde sonore devono percorrere una maggiore distanza per raggiungere l'orecchio più lontano. Uno studio approssimato ma abbastanza preciso di questo fenomeno si può effettuare facendo delle ipotesi semplificative: si suppongono la sorgente lontana, la testa sferica, e le orecchie poste agli antipodi. La prima assunzione implica che le onde sonore che colpiscono la testa saranno onde piane; l'ulteriore distanza  $\Delta x$  che un raggio sonoro dovrà percorrere per raggiungere l'orecchio più lontano, viene stimata da considerazioni geometriche elementari (vedi Figura 1.3):

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

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

- Interaural Level Difference (ILD), causato dal fatto che all'orecchio più lontano giunge un suono di minore intensità causato dall'introduzione da parte della testa di un'attenuazione. Mentre è accettabile sopporre l'ITD come un parametro indipendente dalla frequenza (agendo nelle basse frequenze), l'ILD è invece fortemente legato alla frequenza (agendo nelle alte frequenze): a basse frequenze (per lunghezze d'onda dell'ordine del diametro della testa) alle due orecchie giunge una pressione sonora di differenza trascurabile l'una dall'altra, mentre, ad alte frequenze, la differenza diventa significativa. Come già fatto in precedenza, l'ILD può essere analizzato nel caso ideale di testa perfettamente sferica (di raggio  $a$ )



**Figura 1.3:** Stima di ITD nel caso di sorgente lontana (onde piane) e testa sferica

e con la sorgente sonora situata ad una distanza  $r > a$  dal centro della sfera. È solito utilizzare costanti normalizzate:  $\mu = \frac{\omega a}{c}$  (frequenza normalizzata) e  $\rho = \frac{r}{a}$  (distanza normalizzata). Se si considera un punto sulla sfera allora la diffrazione di un'onda acustica causata dalla sfera nel punto scelto si può esprimere con la funzione di trasferimento:

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

dove  $P_m$  e  $h_m$  sono, rispettivamente, i polinomi di Legendre di ordine m-esimo e la funzione sferica di Hankel;  $\theta_{inc}$  rappresenta l'angolo di incidenza. È noto che la funzione di Hankel  $h_m(x)$  ammette un'approssimazione asintotica quando l'argomento  $x$  tende all'infinito. Dal grafico in Figura 1.4 si nota che a bassa frequenza ( $\mu < 1$ ) la funzione di trasferimento non è dipendente dalla direzione e il suo modulo  $|H_{sphere}(\rho, \theta_{inc}, \mu)|$  è pressoché costante ed uguale a 1 qualunque sia l'angolo. Se  $\mu > 1$  la dipendenza dall'angolo di incidenza diventa molto sensibile. È da notare inoltre che la risposta minima non si presenta per  $\theta_{inc} = \pm\pi$ : questo è dovuto al fatto che tutte le onde sonore che si propagano attorno alla sfera arrivano in quel punto in fase; a frequenze elevate questo fenomeno diventa molto meno presente e la parte posteriore della sfera risulta essere in una zona di ombra sonora. La differenza tra i moduli di questa funzione di trasferimento per l'orecchio destro e l'orecchio sinistro definisce l'ILD.

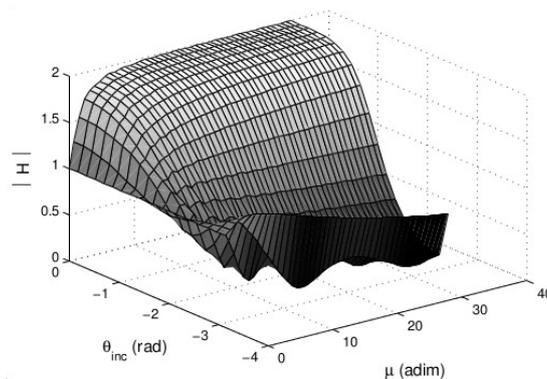
**Orecchio esterno** Quando ci si riferisce all'orecchio esterno si sta sostanzialmente parlando sia del padiglione auricolare (pinna in inglese), che rappresenta la parte effettivamente visibile, sia del condotto uditivo che è chiuso da una estremità dal timpano. Diversi studi hanno accertato che la pinna modifica significativamente il suono in arrivo, specialmente alle alte frequenze: alle basse-medie frequenze la pinna non influisce in modo incisivo sulla pressione acustica al timpano, soprattutto per il fatto che le lunghezze d'onda sono troppo elevate per poter essere catturate da un organo di dimensioni relativamente limitate, mentre alle alte frequenze (circa da 3kHz) si verifica interferenza distruttiva che produce dei notch in frequenza.

Tutto ciò concorre ovviamente a influire sulla nostra capacità di localizzare i suoni.

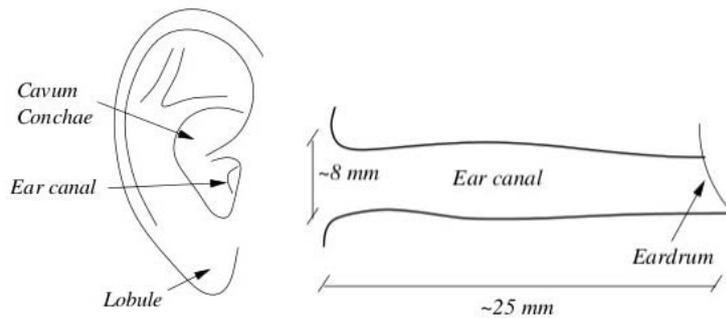
Quando un suono arriva al nostro orecchio si osserva che, dopo aver attraversato il condotto uditivo, raggiunge il timpano provocandone la vibrazione. La Figura 1.5 mostra la struttura dell'orecchio esterno, da cui si osserva che la pinna presenta dei bassorilievi visibili ed è direttamente connessa al condotto uditivo. Quest'ultimo può essere descritto in prima approssimazione come un tubo di larghezza costante caratterizzato da pareti ad alta impedenza acustica: mentre la pinna svolge la funzione di antenna acustica, esso si comporta come un risonatore monodimensionale [9].

Le cavità della pinna, che possono essere geometricamente molto differenti da persona a persona, hanno la capacità di amplificare alcune frequenze, o di introdurre interferenze che ne attenuano altre. Osservando la differenza di lunghezza dei percorsi che compiono il suono diretto e le sue riflessioni si ha che, quando questa differenza è nulla o pari ad un multiplo della lunghezza d'onda, i segnali riflessi arrivano in fase con quello diretto, mentre in caso contrario si ottiene interferenza distruttiva che porta alla presenza di notch nello spettrogramma della risposta.

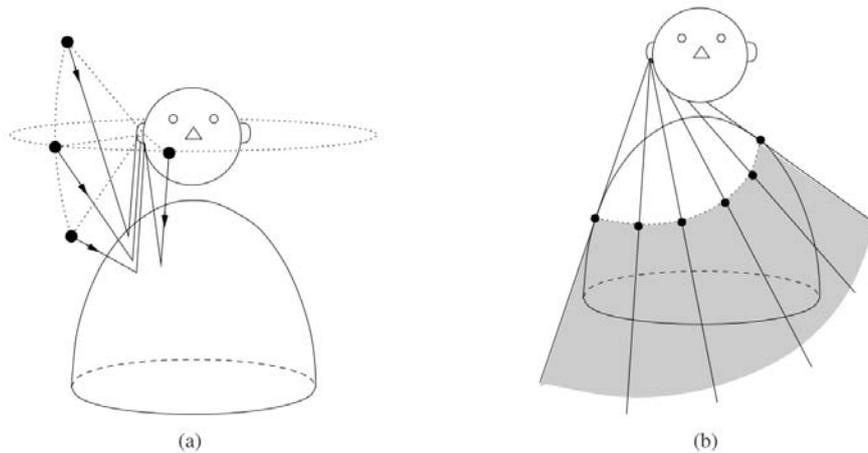
Un altro elemento da tenere in considerazione è l'intensità percepita alle due orecchie, che dipende da come è posizionata la testa (con l'orecchio diretto verso la sorgente si



**Figura 1.4:** Modulo di una sfera  $H_{sphere}(\infty, \theta_{inc}, \mu)$  nel caso di una sorgente a distanza infinita



**Figura 1.5:** Struttura dell'orecchio esterno: pinna e condotto uditivo



**Figura 1.6:** Fenomeni che coinvolgono il torso. (a) Riflessioni del torso. (b) Oscuramento da parte del torso per suoni provenienti dal basso. Le rette tangenti al torso che passano per il timpano definiscono un cono d'ombra.

percepisce un'intensità maggiore che non con l'orecchio posto in altre posizioni).

Tutte queste osservazioni si traducono in un effetto di filtraggio da parte della pinna che può essere effettivamente pensata come un filtro la cui risposta in frequenza dipende in genere dalla distanza e dalla direzione del suono che giunge all'orecchio. Tale effetto può essere interpretato studiando le alterazioni dei raggi sonori introdotte dalla pinna, oppure mediante uno studio nel dominio della frequenza.

**Torso** Il torso interagisce con le onde acustiche essenzialmente in due modi: riflettendo il segnale diretto e "nascondendo" la testa ai suoni provenienti dal basso. Tramite il fenomeno di riflessione, al segnale diretto si somma un treno di segnali riflessi, ciascuno ritardato di una quantità direttamente proporzionale all'elevazione (ad un ritardo maggiore corrisponderà una minore frequenza dell'onda riflessa). Il ritardo non cambia di molto se la sorgente si muove lungo una circonferenza, indicata nella Figura 1.6(a), sul piano orizzontale. Lo spettro delle riflessioni ad opera del torso è approssimabile un filtro comb che introduce notch periodici nello spettro complessivo al timpano. Da

una certa linea di incidenza in poi (Figura 1.6(b)) i fenomeni di riflessione scompaiono, mentre emergono quelli di "oscuramento", che consistono in diffrazione e dispersione attenuando le alte frequenze. Per quanto riguarda la frequenza, dunque, gli effetti del torso sono da considerarsi complementari a quelli della pinna, andando ad agire più sulle basse frequenze.

### 1.2.1 Localizzazione orizzontale

Il ruolo di ILD e ITD è stato dimostrato essere il maggior responsabile per la localizzazione nel piano orizzontale da Lord Rayleigh nel 1907 [10]. Consideriamo una disposizione dove al centro dell'ambiente di ascolto vi è la nostra testa, mentre un altoparlante, posto davanti a noi, spostato verso la nostra destra emette delle onde sonore. Il suono emesso dal trasduttore raggiunge entrambe le orecchie, con diverse intensità, e percorrendo cammini differenti. Questo comporta che il segnale arriverà dapprima all'orecchio destro, che è più vicino alla sorgente, quindi a quello sinistro, cioè giungerà alle due orecchie in tempi diversi. Ciò causa una differenza interaurale di tempo che dipende dalla posizione della sorgente rispetto all'ascoltatore. Se la sorgente si trovasse sul piano mediano, cioè con angolazione orizzontale nulla, la differenza interaurale di tempo sarebbe uguale a zero, mentre sarebbe massima quando la sorgente si trova sulla linea che congiunge le due orecchie. Da questa differenza di tempo con la quale lo stimolo sonoro arriva alle due orecchie deriva anche una differenza di fase delle onde sonore, in presenza di suoni periodici continui. L'orecchio è in grado di riconoscere questa differenza di fase solo, però, per quelle frequenze che hanno una lunghezza d'onda maggiore della distanza che separa le orecchie, e quindi fino a circa 1500 Hz [11]. Per suoni di frequenza superiore la testa inizia a comportarsi come un ostacolo per la propagazione delle onde sonore. Questo fa sì che l'intensità all'orecchio più vicino aumenti rispetto a quella che si avrebbe se non vi fosse la testa, mentre quella all'orecchio più lontano, essendo schermata da essa, diminuisca.

### 1.2.2 Localizzazione verticale

Il nostro sistema di ascolto ci porta ad avere una capacità di localizzare maggiore per suoni situati nel piano orizzontale rispetto a suoni situati su quello verticale. La risoluzione nella percezione di un angolo in elevazione è minore rispetto a quella di un angolo azimutale [12]. La spiegazione deriva dal fatto che, approssimando la testa ad una sfera, si ha che per stesse elevazioni nel piano mediano ITD e ILD rimangono invariati. Nella realtà ciò non accade veramente, poichè la conformità della nostra testa introduce delle variazioni. L'elemento che porta ad eliminare disambiguità nel piano verticale è

la pinna. Mentre per la percezione orizzontale del suono, quindi l'azimuth della sorgente, è preponderante il contributo della testa, per la percezione dell'elevazione risulta dunque fondamentale la presenza della pinna e delle sue insenature. Sperimentalmente è stato infatti dimostrato come con le pinne occluse la localizzazione lungo il piano mediano peggiora, mentre se viene inibito una delle orecchie solo le performance di riconoscimento spaziale rimangono accettabili (questo fatto suggerisce che le informazioni spaziali fornite dal padiglione sono essenzialmente monaurali). È possibile un riconoscimento dell'elevazione di una sorgente solamente se questa possiede sufficiente energia alle alte frequenze, a partire dalle quali la pinna contribuisce maggiormente a trasformare il segnale. Esistono varie teorie circa il ruolo della pinna, tutte però concordano nel fatto che l'elevazione comporta lo spostamento di notch e picchi nello spettro del segnale. Si è dimostrato anche che un suono, per avere una localizzazione verticale accurata, deve avere una sostanziale energia nelle alte frequenze, questo perché la pinna ha dimensioni limitate nello spazio e le onde con lunghezza d'onda maggiore delle dimensioni della pinna non sono affette da essa. Sotto i 3 kHz generalmente si ritiene che la pinna non modifichi un suono [10].

### 1.2.3 Percezione della distanza

L'elemento più influente nella percezione della distanza di un suono è l'intensità con cui viene percepito [13], essa infatti dipende dalla distanza tra emettitore ed ascoltatore per via di attenuazioni dovute alla propagazione nell'aria. Tale attenuazione agisce maggiormente alle alte frequenze che non alle basse, tale considerazione porta ad osservare che un suono a cui è stato applicato un filtro passa alto è percepito più vicino di uno a cui è stato applicato un passa basso [13]. Nielsen [14] attraverso i suoi studi ha mostrato che un ascoltatore tendenzialmente sovrastima la distanza di sorgenti vicine e sottostima quello di sorgenti lontane. Se si immagina una situazione in cui vi sia una sola sorgente e nessun riverbero, si avrebbe che la percezione della distanza sarebbe basata sulla dipendenza in frequenza di ITD e ILD [12] e sull'intensità con cui il suono arriva. Nel caso in cui invece vi siano fenomeni riverberanti gli elementi che influiscono alla percezione della distanza aumentano. Nielsen [14], in tal caso, ha evidenziato che l'elemento di maggior influenza è il rapporto tra il suono diretto ed il riverbero della stanza. Ulteriori elementi che concorrono nella percezione della distanza sono la visione della sorgente, dove quindi entra in gioco anche il senso della vista che è più allenato nella stima delle distanze, e la familiarità con i suoni in quanto la nostra esperienza ci permette di effettuare stime più o meno veritiere dipendenti dal fatto che un dato suono sia a noi familiare o meno. In conclusione si ha che l'intensità e lo spettro di un suono sono considerati come elementi per una percezione della distanza relativa poiché solo mediante il confronto con

un suono di riferimento si in grado di ottenere informazioni. Il riverbero invece è considerato con un elemento per una percezione assoluta poiché permette immediatamente una stima della distanza.

# Capitolo 2

## Rendering spaziale

### 2.1 Head-related transfer function

Tutti gli effetti descritti nel capitolo precedente, che concorrono alla nostra percezione di un evento acustico, possono essere descritti da opportune funzioni di trasferimento che possono essere trattate come lineari e che, se sommate, possono ben simulare la sovrapposizione degli effetti stessi in un unico sistema lineare. La funzione di trasferimento di questo sistema viene chiamata *Head-Related Transfer Function (HRTF)*. Formalmente, le HRTF sono costituite di una componente sinistra e una destra, corrispondenti alla pressione acustica prodotta da una sorgente sonora arbitraria al timpano sinistro e destro rispettivamente, e vengono indicate, sinteticamente, con  $H^{(l),(r)}(r, \theta, \phi, \omega)$ , dove  $r, \theta, \phi$  sono le coordinate spaziali usate finora e  $\omega$  è la frequenza del suono al timpano.

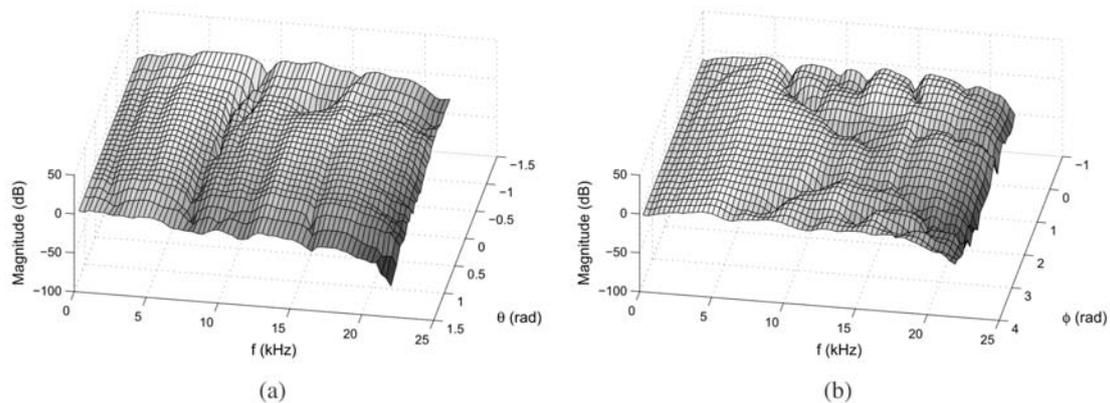
Quando  $r$  tende ad infinito, che per molte applicazioni si traduce in  $r > 1m$ , si parla di condizione di campo lontano (far field) per la sorgente. In questo caso, le HRTF non dipenderanno più dalla distanza. Ipotizzando poi perfetta simmetria della testa, possiamo considerare le due componenti identiche a meno dell'angolo azimutale, che sarà uno l'opposto dell'altro. La notazione precedente diventa allora  $H^{(l)}(\theta, \phi, \omega) = H(\theta, \phi, \omega)$ ,  $H^{(r)}(\theta, \phi, \omega) = H(-\theta, \phi, \omega)$ .

L'HRTF è definita come la funzione di trasferimento acustica tra sorgente sonora e timpano, caratterizzata da una dipendenza in frequenza e spazio, e adeguatamente compensata in campo libero [15]. Matematicamente, è il rapporto tra l'SPL  $\Phi^{(l),(r)}(\theta, \phi, \omega)$  al timpano corrispondente, e l'SPL  $\Phi_f(\omega)$  nel punto occupato dal centro della testa in condizione di free-field<sup>1</sup>.

$$H^{(l)}(\theta, \phi, \omega) = \frac{\Phi^l(\theta, \phi, \omega)}{\Phi_f(\omega)} \quad H^{(r)}(\theta, \phi, \omega) = \frac{\Phi^r(\theta, \phi, \omega)}{\Phi_f(\omega)} \quad (2.1)$$

---

<sup>1</sup>Un ambiente in cui non avvengono riflessioni delle onde sonore, come ad esempio in camera anecoica, dove le pareti sono pressoché totalmente assorbenti



**Figura 2.1:** Grafici dell'intensità per due esempi di HRTF in dipendenza dell'elevazione  $\Phi$  (a) e dell'azimuth  $\theta$  (b).

Dalla Figura 2.1 si può osservare come le HRTF siano in generale funzioni assai complesse, riunendo assieme gli effetti acustici relativi a testa, pinna e torso in un'unica funzione. Esse fanno parte di funzioni di trasferimento ancora più complesse, le Binaural Room Transfer Function (BRTF), che contengono sia le informazioni acustiche sull'individuo (HRTF) sia quelle dell'ambiente circostante in cui avviene l'evento sonoro (Room Transfer Function, RTF). Dalle HRTF si possono ricavare a loro volta funzioni di trasferimento indicatrici esclusivamente dell'apporto della pinna, le Pinna-Related Transfer Function (PRTF). Tutte queste funzioni nel dominio della frequenza hanno il loro corrispettivo nel dominio del tempo: Head-Related Impulse Response (HRIR), Binaural Room Impulse Response (BRIR), Room Impulse Response (RIR), Pinna-Related Impulse Response (PRIR).

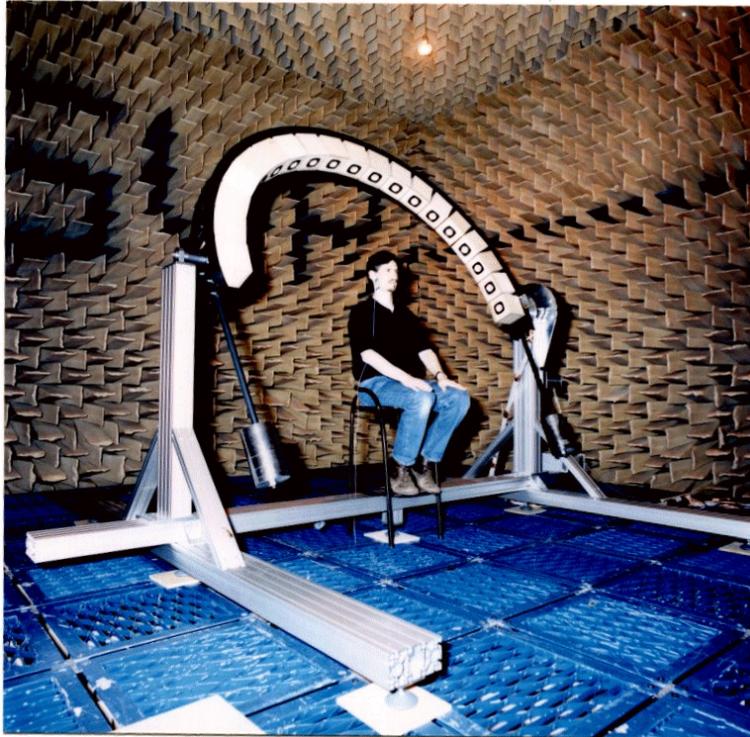
## 2.2 Metodi di acquisizione di HRTF

### 2.2.1 Modelli strutturali misti

I modelli strutturali misti (MSM) utilizzano una combinazione tra sintesi e misurazioni. Con questi procedimenti, che uniscono filtri sintetizzati appositamente e contributi misurati acusticamente, è possibile realizzare display uditivi adattabili all'antropometria dell'ascoltatore per la resa verticale del suono [16].

### 2.2.2 Modelli di analisi ed estrazione

Il pattern di analisi consiste in una serie più o meno ricca di misurazioni acustiche, in genere compiute in ambiente anecoico (o quantomeno controllato, come una stanza iso-



**Figura 2.2:** *Acquisizione di HRTF in una camera anecoica tramite un dispositivo professionale*

lata acusticamente che porti ad un ulteriore abbattimento dell'intensità del rumore proveniente dall'esterno, o provvisto di pareti fonoassorbenti per avvicinarsi all'ambiente anecoico) per annullare gli effetti del riverbero e delle riflessioni ambientali che si attue- rebbero in campo aperto e per non venire influenzati dal rumore di disturbo proveniente dall'esterno della camera. Le misurazioni avvengono grazie ad un set di altoparlanti disposti variamente attorno ad un sistema di microfoni (situazione di free-field) o al soggetto umano di cui interessa misurare la HRTF, opportunamente equipaggiato con microfoni applicati all'ingresso dell'orecchio esterno Figura 2.2. In alternativa ad un individuo reale vengono impiegati manichini, come già accennato nel capitolo 1, caratterizzati da misure antropometriche standard e medie, uno dei più adottati dei quali è la KEMAR (Knowles Electronics Manikin for Auditory Research), adottato anche per le misurazioni trattate in questa tesi [17]. Le registrazioni ottenute con questa procedura sono di qualità molto elevata, sono individuali e richiedono minimo sforzo computazionale per l'estrazione delle HRIR dell'ascoltatore in quanto la componente delle RIR viene annullata dall'anechoicità della camera. Tuttavia, come si può immaginare, questo metodo è molto dispendioso in termini economici dato che si ha a che fare con attrezzatura di livello professionale, e talvolta in termini di tempo (allo stato dell'arte vi sono strumentazioni che permettono l'acquisizione in soli dieci minuti [18], altre strumentazioni impiegano invece un tempo sostanzialmente più lungo), specialmente

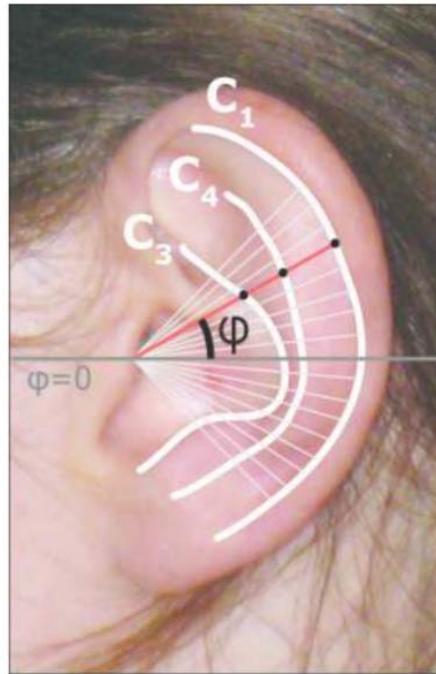
Database	CIPIC [19]	LISTEN	FIU	MIT	AALTO	ARI	PKU&IOA
Sampling Frequency	44100 Hz	44100 Hz	96000 Hz	44100 Hz	48000 Hz	48000 Hz	65536 Hz
HRIR length	200pts	512pts	256pts	512pts	8192pts	256pts	1024pts
Coordinate system	IP	VP	VP	VP	VP	VP	VP
Spatial grid	25A,50E	24A,10E	12A,6E	72A,14E	72A,19E	90A,22E	72A,14E
No. of directions	1250	187	72	710	240	1550	793
No. of distance	1	1	1	1	2	1	8
Distance	1 m	1.95 m		1.4 m	0.68,1.35 m	1 m	0.21.6 m
Stimulus type	GC	LS	GC	ML	LS	LS	SG
Mic position	BEC	BEC	BEC	BEC	BEC	BEC,BTE	BEC
Data format	.mat	.wav, .mat	.txt	.wav	.mat	.mat	.dat
No. of subject	45	51	15	2	21	92	1
Raw data	no	yes	no	yes	no	yes	no
Onset/ITD data	yes	no	yes	no	no	no	no
Anthropometric data	yes	yes	yes	no	no	yes	no

**Tabella 2.1:** *Principali basi di dati pubbliche di HRTF con loro caratteristiche [6]*

se consideriamo che per delle misurazioni ottimali sono necessarie diverse decine di campioni.

### 2.2.3 Modelli di selezione

Il modello di selezione fa uso di HRTF pre-registrate in ambiente anecoico e organizzate in basi di dati opportune. Molto spesso tali basi di dati comprendono misurazioni effettuate su manichini che vengono poi utilizzate per un utilizzo delle HRTF “generico” e che in media possa essere efficace. Per avere le HRTF, l’utente deve semplicemente estrarre quelle di suo interesse tra quelle a disposizione dalla base di dati. Esistono database pubblici di HRTF che è possibile sfruttare per la selezione, alcuni dei più famosi sono riportati in Tabella 2.1 che oltre ad elencarli ne descrive gli aspetti salienti. È possibile migliorare la resa da parte di questo modello con algoritmi di selezione binaurale. Un sistema efficace consiste nel partire da misure antropometriche dell’orecchio dell’utente ottenute tramite una foto del padiglione esterno per selezionare, all’interno di un database che includa questo tipo di informazioni, le HRTF, o sue componenti, che corrispondono ai valori antropometrici più vicini a quelli misurati dalla fotografia. I risultati ottenuti con due set di HRTF così selezionate mostrano come, grazie alla maggiore individualizzazione introdotta, le performance di localizzazione delle sorgenti sonore migliorino significativamente con questo modello rispetto a quelle ottenute tramite selezione di HRTF di manichini [20]. Il metodo di selezione rappresenta dunque un buon compromesso tra la qualità delle HRTF ottenute, garantita dalla provenienza da esperienze in camera anecoica, e i costi richiesti, molto ridotti rispetto a quelli per i modelli di misurazione (le misurazioni sono già state fatte e gli utenti possono usufruire gratuitamente dei dati nelle collezioni pubbliche). Tuttavia, vale la pena notare che, pur potendo acquisire un certo grado di individualità tramite algoritmi opportuni, le risposte ottenute per selezione non sono in nessun caso esattamente quelle dell’utente. Resta



**Figura 2.3:** Misure antropometriche della pinna

ancora dimostrane che le HRTF individuali siano migliori (in termini di localizzazione verticale) di una qualsiasi altra HRTF generica presente nel mondo.

## 2.2.4 Modelli di sintesi

Questa è la tipologia di modelli meno sviluppata dal punto di vista implementativo, conseguenza data dal fatto che spesso le risposte che vengono restituite tramite questi modelli non possiedono l'accuratezza necessaria ad un rendering audio in tre dimensioni. Condividono il carattere puramente sintetico delle HRTF estrapolate, non richiedendo quindi misurazioni ad-hoc e forniscono risposte comunque su misura per l'utente. Le ricerche più promettenti per questo tipo di procedure di acquisizione riguardano:

- Misure antropometriche: algoritmi che ricevono in input dimensioni antropometriche (Figura 2.3) come lunghezza del bordo esterno della pinna e diametro del canale uditivo e restituiscono in output una HTRF dipendente solo da quei parametri;
- Approssimazioni ai primi ordini di HRTF: algoritmi di sintesi di HRTF basati unicamente sulla progettazione di filtri appositi per la modellizzazione matematica completa delle trasformazioni che il suono subisce tra testa, pinna e torso. I più diffusi modelli sono il modello pole-zero, che crea HTRF semplificate attraverso l'impiego di criteri percettivi, e l'espansione in serie, che sfruttano il *Principal*

*Component Analysis* (PCA) per sintetizzare HRTF direzionali partendo da una serie di misurazioni già portate a termine in precedenza;

Il fatto che questi modelli siano ancora privi di una realizzazione fa intuire che la sintesi di HRTF sia il percorso più complicato da seguire tra quelli esistenti, per via degli innumerevoli fattori determinanti di cui molto ancora non conosciamo e che finora ci hanno reso inevitabile ricorrere alla misurazione acustica. Non un caso che, se funzionanti e impiegabili per l'effettiva resa spaziale del suono, essi rappresenterebbero probabilmente le soluzioni pi economiche, veloci ed affidabili a disposizione.

# Capitolo 3

## Realtà aumentata acustica

Il termine realtà aumentata (in inglese augmented reality AR) è stato definito nel 1997 da Azuma [21] e si identifica per tre importanti caratteristiche:

1. Combinazione tra realtà ed elementi virtuali.
2. Interazione in tempo reale.
3. Estensione in tre dimensioni.

In altri termini la realtà aumentata sovrappone elementi virtuali alla nostra percezione del mondo reale, avendo cura di renderli invariante rispetto ai nostri movimenti. Volendo ottenere un'interpretazione più restrittiva si può dire che gli elementi virtuali inseriti in un contesto di realtà aumentata accrescono il mondo reale originario in modo significativo. Negli anni sono stati studiati vari modi per ottenere realtà aumentata, la prima sulla quale si è concentrata la ricerca è la realtà aumentata visiva, che permette, utilizzando ad esempio un display capace di mostrare il mondo reale, di inserire elementi virtuali sovrapponendoli di fatto al mondo reale che si sta osservando.

Un display di un cellulare può quindi, grazie alla videocamera integrata, visualizzare quando si sta riprendendo, con la possibilità di inserire elementi di tipo visivo che aumentano le informazioni da noi percepite guardando il display. Un esempio può essere la recensione di un dato ristorante visualizzato a display quando viene inquadrato dalla videocamera. In tal caso la realtà aumentata diventa di tipo mobile.

Un altro elemento che concorre alla realizzazione di realtà aumentata è l'audio. Esso può essere combinato con elementi visivi per ottenere applicazioni multimodali, oppure può essere utilizzato singolarmente per applicazioni unimodali, in tal caso si parla di audio augmented reality. Sarter (2006) [22] ha spiegato quali sono le principali differenze tra canali uditivi e visivi nella presentazione di informazioni ad un utente finale. In primo luogo, la nostra percezione uditiva ha un carattere omnidirezionale: siamo in

grado di percepire i suoni da qualsiasi direzione in qualsiasi tempo. In secondo luogo, i suoni hanno in genere una natura transitoria: un cartello scritto può essere letto più volte finché è visibile, ma una frase parlata può essere ascoltata una sola volta, a meno che non venga ripetuta. Infine, mentre un messaggio visivo è facilmente ignorabile, cambiando ad esempio la direzione del nostro sguardo o semplicemente chiudendo gli occhi, difficilmente si ha la possibilità di ignorare un messaggio sonoro. Un'altra fondamentale differenza viene citata da Kolarik et al. (2015) [23]: il suono può normalmente viaggiare attraverso ostacoli. Nella realtà aumentata, il carattere omnidirezionale della percezione uditiva è spesso un vantaggio rispetto agli approcci visivi, le applicazioni possono infatti guidare l'attenzione dell'utente e trasmettere informazioni in qualsiasi direzione ed in qualsiasi momento. Questo vantaggio diventa maggiore se consideriamo applicazioni per smartphone in cui l'output visivo è limitato a causa di display di piccole dimensioni.

### 3.1 Headset

L'estensione della percezione reale con l'inserimento di elementi virtuali, nel campo della realtà aumentata acustica, viene effettuato mediante l'uso di un equipaggiamento speciale a cui si sovrappone un'azione di signal processing. Viene definito quindi un headset composto da due microfoni miniaturizzati integrati alle estremità esterne di un paio di cuffie in-ear, uno per ogni cuffia. Il segnale acquisito dai microfoni viene riprodotto mediante le cuffiette in tempo reale senza alterazioni significative, in tal modo l'utente ha idealmente la possibilità di ascoltare una copia della realtà acustica che lo circonda. Tale copia prende il nome di ambiente pseudo-acustico [24].

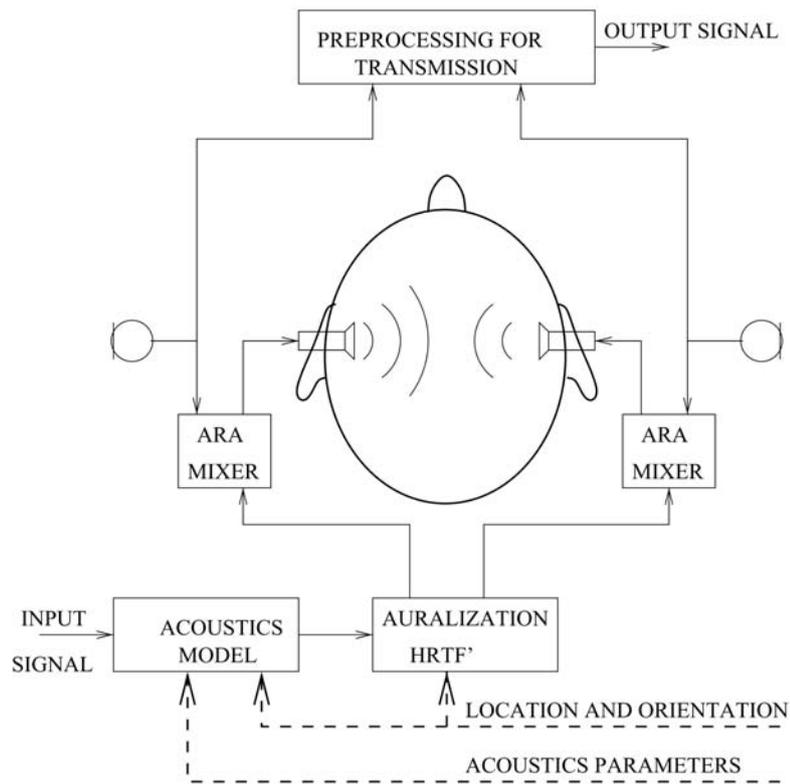
#### 3.1.1 Mixer

Ovviamente tutto ciò avviene idealmente, nella realtà l'headset non è completamente trasparente e causa delle colorazioni alla rappresentazione pseudo-acustica dell'ambiente reale.

Nell'ambiente pseudo-acustico possono quindi essere inseriti elementi virtuali che, se opportunamente spazializzati, possono essere scambiati per reali. Tale inserimento viene effettuato mediante un dispositivo denominato AAR mixer.

Tale mixer ha il compito di instradare e mixare i segnali ed agisce come interfaccia tra l'utente ed il mondo esterno. Il mixer ha come compiti principali i seguenti:

- Instradare i segnali microfonici agli auricolari



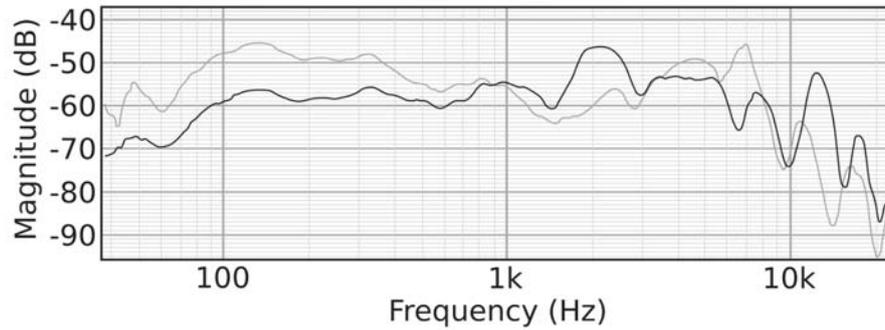
**Figura 3.1:** Diagramma di un sistema di realtà aumentata acustica

- Instradare i segnali microfonici ad un sistema esterno, come un computer o uno smartphone
- Ricevere i segnali audio in ingresso da un sistema esterno, aggiungendoli alla rappresentazione pseudo-acustica

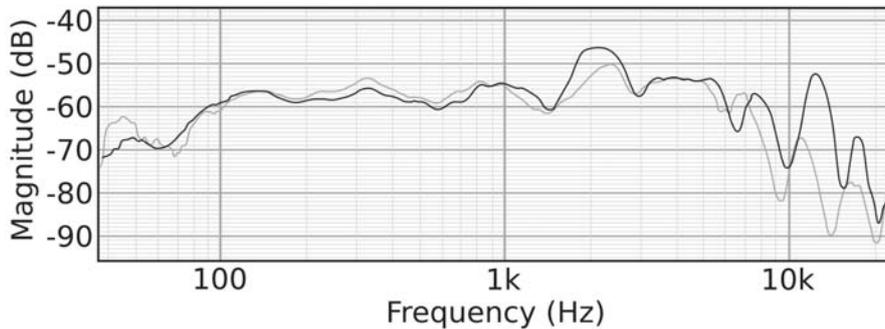
è auspicabile che il mixer AAR sia piccolo e leggero in modo che l'utente non ne venga appesantito. Il consumo di corrente deve poi essere il più basso possibile per garantire una lunga durata di funzionamento. La Figura 3.1 mostra un diagramma base di un'applicazione AAR. Al fine di migliorarne l'usabilità è preferibile installare una connessione wireless per la comunicazione con dispositivi esterni.

Se si vuole che gli oggetti sonori virtuali inseriti nell'ambiente pseudo-acustico vengano percepiti come reali, si ha che devono risultare invarianti rispetto alla posizione dell'ascoltatore, vi è perciò la necessità di tenere traccia dei movimenti e della posizione dell'utente. Vi sono vari modi per poterlo fare ed il processo è chiamato "head-tracking and positioning" [25].

In [4, 5] vengono analizzati i componenti che costituiscono il mixer AAR. Viene calcolata la funzione di trasferimento da un loudspeaker (posto di fronte ad un soggetto di test) e da un microfono miniaturizzato inserito all'interno del canale uditivo ad una



**Figura 3.2:** Funzione di trasferimento senza equalizzazione.



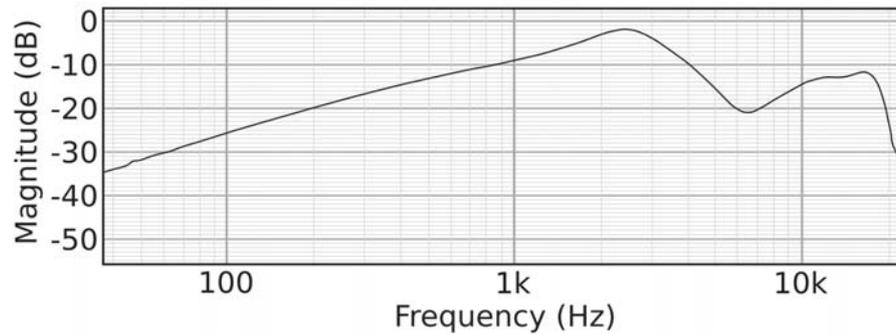
**Figura 3.3:** Funzione di trasferimento con equalizzazione.

profondità di 5mm in due casi: nel primo si lascia il canale uditivo aperto, nel secondo si chiude il canale con l'headset e lo si utilizza per la riproduzione. Sempre in [4, 5] sono stati eseguiti quattro test su quattro soggetti differenti, il risultato per uno di questi è mostrato in Figura 3.2 dove viene mostrata la funzione di trasferimento misurata da un altoparlante davanti a un soggetto di prova ad un microfono all'interno condotto uditivo del soggetto. La linea nera corrisponde alla risposta con il canale uditivo occluso, la linea grigia corrisponde invece alla risposta con l'headset indossato, senza equalizzazione. L'obiettivo è, mediante l'applicazione di un filtro compensativo, di rendere quanto più simile possibile la linea nera a quella grigia.

In Figura 3.3 (La linea nera corrisponde alla risposta con il canale uditivo occluso, la linea grigia corrisponde invece alla risposta con l'headset indossato, con equalizzazione.) gli autori ci riescono applicando tre filtri: un filtro passa alto del primo ordine, un filtro peak del secondo ordine ed un filtro notch del secondo ordine, ottenendo il filtro di Figura 3.4

### 3.1.2 Display acustici virtuali

Abbiamo già parlato degli effetti che agiscono su un suono che arriva al timpano, abbiamo anche già constatato che questi effetti, ad esempio la early reflection, contengono informazioni circa l'ambiente in cui il suono viene emesso. Lo stesso tipo di effetti pos-



**Figura 3.4:** Curva di equalizzazione generica, generata a partire dalle informazioni prese da tutti e quattro i soggetti di test. [4, 5]

sono essere applicati ad un suono che viene trasmesso tramite cuffie, in tal modo gli si conferiscono delle caratteristiche che non lo fanno pi percepire come emesso dalle cuffie, ma come emesso da una sorgente nello spazio attorno all'ascoltatore. Questo tipo di sistema chiamato Display acustico virtuale (VAD). Le applicazioni di AAR utilizzano tecniche VAD per inserire sorgenti virtuali all'interno dell'ambiente pseudo-acustico.

## 3.2 Realtà aumentata acustica in campo mobile

In un sistema di tipo realtà acustica aumentata in campo mobile (mAAR), l'ascoltatore può usufruire di un mix di sorgenti sonore reali e virtuali. Le prime vengono riprese dai microfoni dell'headset previo filtraggio naturale del soggetto con conseguente applicazione di una compensazione in frequenza-fase che tenga conto dell'errore introdotto dalla differenza di posizione dei microfoni rispetto al punto d'ingresso del canale uditivo [26]. Le sorgenti virtuali necessitano invece di un processo di auralizzazione dinamico per ottenere una sovrapposizione ideale con la realtà. L'auralizzazione comprende un rendering tramite la risposta binaurale della stanza, risultante nel filtraggio tramite la risposta della stanza e la risposta personale del soggetto che devone essere personalizzate in accordo con l'ambiente circostante [27] e con l'ascoltatore [16]. Gli algoritmi di signal processing (DSP) implementano filtri correttivi che compensano i microfoni, gli speaker e le loro interazioni, prendendo in considerazione gli effetti psicoacustici e di colorazioni causati dall'indossare la cuffia rispetto alle normali condizioni di ascolto senza di essa [28]. In particolare, una procedura che permette ad un qualsiasi soggetto di personalizzare l'ascolto andando ad agire sui filtri correttivi per la compensazione dell'headset, nel capitolo 7 di questa tesi viene proposta un'applicazione per effettuare questa operazione.

### 3.2.1 La sfida dell'esternalizzazione

Produrre scenari acustici virtuali e aumentati in cuffia con attenzione alle caratteristiche di ambiente e di esternalizzazione rimane una delle sfide più importanti in questo dominio di ricerca, a causa delle forti interconnessioni presenti tra i vari componenti del sistema mAAR. Con il termine "esternalizzazione" indichiamo la percezione di un suono localizzato fuori dalla testa, piuttosto che al suo interno. Le sfide e i criteri per una corretta esternalizzazione possono essere riassunti in quattro categorie:

- sistema ergonomico: le cuffie ideali dovrebbero essere acusticamente trasparenti: l'ascoltatore non si dovrebbe rendere conto della loro presenza né dell'emissione di suoni dagli speaker. A tale scopo sono essenziali cuffie poco invasive nel peso e nelle dimensioni;
- tracciamento: i movimenti della testa nell'ascolto binaurale producono segnali interaurali dinamici [29]; tracciare la posizione dell'utente nell'ambiente permette il riconoscimento dell'interazione acustica e una rappresentazione spaziale coerente tra scena reale e simulazione virtuale;
- conoscenza dell'ambiente: la percezione spaziale della scena acustica in cui si è immersi richiede la conoscenza di riessioni e riverberazioni tipiche dell'ambiente stesso [30]; queste informazioni risultano essenziali per una impressione realistica dello spazio in cuffia [31];
- individualizzazione: testa e orecchio del soggetto filtrano in maniera individuale i suoni; va presa in considerazione una correzione individuale per l'accoppiamento acustico tra cuffie e orecchio esterno [32].

### 3.2.2 Acustica del canale uditivo chiuso

Il suono che si propaga dall'ingresso dell'orecchio al timpano si comporta come un'onda piana e come tale, quando giunge al timpano, in parte si riflette e torna indietro sino alla parte aperta dell'orecchio [33]. Il canale uditivo si comporta analogamente ad un tubo risonatore semi-aperto di lunghezza di 25 millimetri con un diametro medio di 7 millimetri [34]. La lunghezza della prima frequenza di risonanza è lunga un quarto del tubo semi-aperto e approssimativamente si trova circa a 3.0 kHz. Tale frequenza e le frequenze di risonanza vicine sono responsabili di picchi nello spettro del segnale. In un tubo risonatore semi-aperto, le onde stazionarie si presentano a frequenze secondo la seguente formula, dove  $L$  è la lunghezza del canale,  $n$  è un numero dispari positivo, e  $c$  è la velocità del suono.

$$f(n) = \frac{nc}{4L} \quad (3.1)$$

Quando viene inserita la cuffietta nel canale uditivo, l'acustica del canale stesso cambia e, analogamente alla situazione precedente, il canale si comporta come un tubo risonatore chiuso. Questo cambiamento di comportamento porta degli slittamenti delle frequenze di risonanza lungo lo spettro del segnale [33]. La lunghezza d'onda della risonanza principale diventa metà della lunghezza del canale, lunghezza significativamente maggiore di quella che si aveva nel canale semi-aperto. La frequenza di tale risonanza quindi raddoppia portandosi circa a 6 kHz. Analogamente a prima, in un tubo risonatore chiuso, le frequenze di risonanza appaiono in accordo con la seguente formula, dove  $L$  è la lunghezza del canale,  $n$  è un numero positivo, e  $c$  è la velocità del suono.

$$f(n) = \frac{nc}{2L} \quad (3.2)$$

Lokki e Huhtakallio [33] hanno analizzato le frequenze di risonanza di un simulatore di canale uditivo, che presenta tutte le caratteristiche del canale uditivo umano, occluso e non occluso. Nelle misure a canale aperto si identificano nella risposta impulsiva due evidenti picchi, il primo dei quali dovuto al raggiungimento diretto del suono al timpano, il secondo invece dovuto ad una prima riflessione del suono nel condotto uditivo. Comunque, nel caso di canale uditivo non occluso vi sono diversi picchi significativi dovuti alle riflessioni. Le frequenze di tali picchi si verificano approssimativamente a  $1/4\lambda$ ,  $3/4\lambda$  e  $5/4\lambda$ , dove  $\lambda$  è la lunghezza d'onda. In corrispondenza si ha che a canale uditivo occluso le frequenze di risonanza si trovano a  $1/2\lambda$  e  $1/\lambda$ . Tali misurazioni confermano le teorie precedentemente analizzate.

Tutte queste caratteristiche devono essere tenute in considerazione nella progettazione di cuffie per la realtà aumentata acustica, altrimenti il suono si colora e viene percepito come non naturale. Le possibilità per compensare tale fenomeno acustico, riducendo quindi la colorazione, possono essere:

- l'applicazione di un filtro generico pre-definito
- l'utilizzo di algoritmi di compensazione
- utilizzare una procedura guidata di auto-calibrazione

### 3.2.3 Effetto di occlusione

Quando parliamo, la nostra voce giunge al timpano in due modi, attraverso la propagazione del suono attraverso l'aria e attraverso le nostre ossa. Nel caso in cui il nostro canale uditivo sia aperto, il maggior contributo nella percezione del suono viene apportato dalla trasmissione attraverso l'aria. Tuttavia, quando il canale uditivo è bloccato,

l'impedenza acustica al timpano diventa maggiore a causa delle ossa ossa e della conduzione del tessuto. In questo caso le pareti dei canali auricolari sono direttamente collegati alle cavità bloccati. Questo fenomeno è chiamato effetto di occlusione [35], ed dovuto al fatto che il canale uditivo si comporta come una camera di pressione. La vibrazione delle pareti causa una aumento della pressione acustica (SPL) se comparata al caso con il canale aperto. Tali vibrazioni sono quindi propagate al timpano. In aggiunta si ha che la conduzione ossea diventa dominante per frequenze sotto 1 kHz con un'amplificazione anche di 30 dB [36]. Il risultato finale ci porta a percepire la nostra voce come innaturale.

### 3.2.4 Effetto di isolamento

Quando indossiamo un paio di cuffiette, esse ci isolano dai suoni dell'ambiente a noi circostante per permettere di concentrarci sui suoni che emettono. Tale effetto di isolamento è influenzato da fattori acustici e psicoacustici. Bernier e Voix [35] ne hanno identificati quattro. In primo luogo si ha che le cuffiette isolano maggiromente alle alte frequenze che alle basse, secondariamente si ha che l'effetto di occlusione causa uno slittamento delle principali frequenze di risonanza nello spettro del segnale, in terzo luogo si considera il fatto che le cuffiette potrebbero attenuare troppo i segnali ambientali (portando a pericoli), e in ultima si ha che la percezione non lineare dell'intensità può causare un'attenuazione uniforme che non viene percepita uniformemente nel dominio della frequenza. Tali fenomeni sono presenti in un headset AAR poiché le cuffie hanno un potere isolante rispetto all'ambiente esterno. Dunque se un ambiente pseudo-acustico prodotto non è naturale, tutto l'ambiente AAR può sembrare scollegato dal reale ambiente circostante.

### 3.2.5 Dispersione

Quando indossiamo delle cuffiette si ha che parte del suono che essere riproducano esca dal canale uditivo. In una situazione di AAR, dove un microfono posizionato sulla parte opposta della cuffietta registra l'ambiente circostante, avviene che parte del segnale possa rientrare generando un loop. Al timpano arriva quindi la somma tra il segnale riprodotto dalla cuffietta e il segnale che ha subito un effetto di dispersione [34]. Quando il ritardo dell'ambiente pseudo-acustico riprodotto è eccessivamente grande, la sua combinazione con un il suono affetto da dispersione provoca un effetto di filtraggio a pettine. Questo effetto provoca una colorazione che rende fastidioso il suono percepito. Pertanto, l'elaborazione real time dell'ambiente pseudo-acustico è di fondamentale importanza. Harma et al. [37] menzionano due approcci per ridurre questo effetto. Un

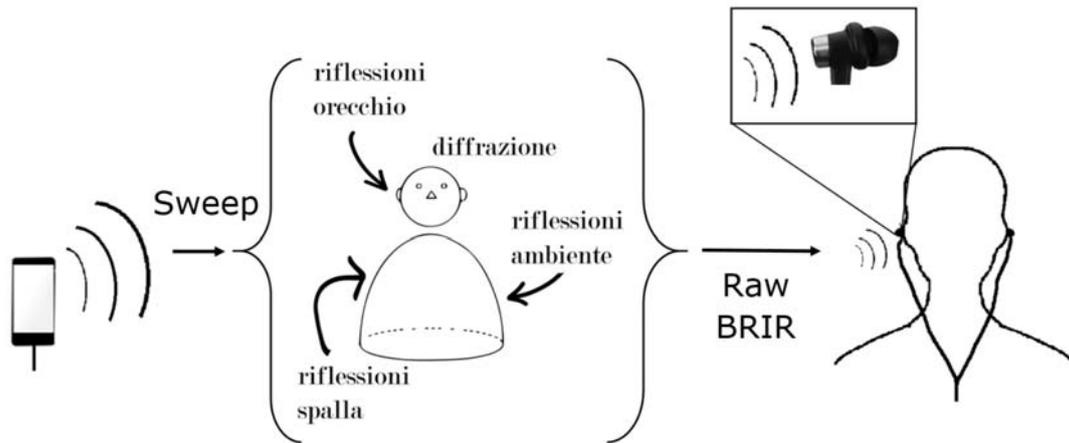
primo approccio si basa sullo studio di design delle cuffie che permettano la minor dispersione possibile, il secondo approccio invece prevede di rafforzare il livello sonoro pseudo-acustico per mascherare l'effetto di dispersione.



# Capitolo 4

## Il progetto SelfEar

Dalla trattazione del Capitolo 2 si ha che ottenere delle HRTF di qualità sufficientemente alta da essere applicabili a contesti di display acustici virtuali, ossia in grado di dare effettivamente la percezione della posizione delle sorgenti sonore durante l'ascolto in cuffia, non è un compito banale, specialmente in ambiente domestico. L'analisi e l'estrazione di HRTF in ambiente controllato rappresenta la soluzione migliore in termini di qualità ma anche la più costosa sia in tempo che in denaro; la selezione di HRTF da basi di dati esistenti garantisce la qualità del metodo di estrazione in ambiente controllato, ma garantisce una minore individualità delle risposte; algoritmi di sintesi che creino HRTF partendo dalle sole misure antropometriche non sono ancora abbastanza sviluppati da poter essere utilizzati; soluzioni strutturali miste possono migliorare e compensare aspetti in cui la singola soluzione è carente; in particolare, la metodologia di modellazione esposta nel paragrafo 2.2.4 è, al momento, una delle soluzioni più equilibrate ed efficaci, ma manca ancora di applicazione concreta che permetta di usare effettivamente questo metodo. Dal momento che la direzione che ha intrapreso già da tempo la realtà virtuale e aumentata è di rapida crescita, tanto da farci immaginare la presenza massiccia di ambienti virtuali anche nella vita di tutti i giorni in un futuro più prossimo che remoto, è richiesto uno sviluppo di pari passo anche da parte dell'audio immersivo, componente fondamentale per una resa il più realistica possibile dell'ambiente. Nello specifico, per l'aspetto quotidiano che si prevede queste tecnologie acquisiranno, diventano di particolare rilevanza la rapidità e l'economicità del processo di ottenimento delle HRTF, soprattutto pensando ad applicazioni di massa. L'idea, quindi, è quella di escogitare una soluzione per ottenere HRTF individuali che possano essere sia facilmente accessibili ad un pubblico vasto, che una valida alternativa alle tecniche già esistenti, e che siano effettivamente utilizzabili per una resa acustica tridimensionale ottimizzata per l'utente.



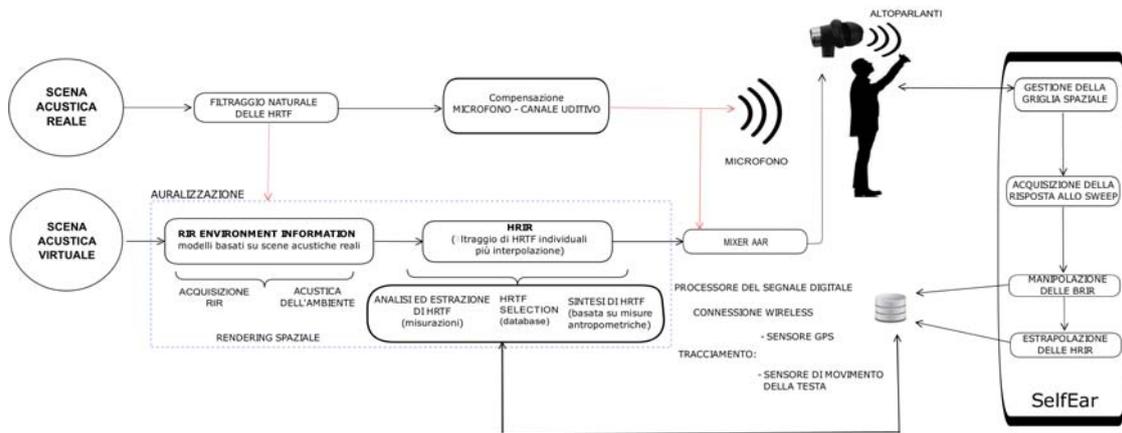
**Figura 4.1:** Percorso dell'impulso dagli altoparlanti del dispositivo ai microfoni degli auricolari attraverso le trasformazioni acustiche coinvolte nel fenomeno di riproduzione.

## 4.1 Un'applicazione mobile

La soluzione proposta, dal nome in codice SelfEar, è un'applicazione per smartphone e tablet che in grado di scattare una "fotografia acustica" all'orecchio dell'utente. Il nome "SelfEar" richiama i selfie fotografici, e vuole proprio evidenziare come l'obiettivo sia fare "un selfie al profilo acustico delle orecchie dell'ascoltatore". Mentre con una foto è possibile ricostruire una funzione di trasferimento misurando le grandezze antropometriche, tramite un impulso acustico possiamo direttamente registrare la risposta che il nostro personale sistema busto-testa-orecchie dà a tale impulso, e ricavare le HRTF a partire da questa risposta. C'è dunque bisogno di utilizzare sia gli altoparlanti del dispositivo, per riprodurre il segnale impulsivo, sia un paio di cuffie in-ear binaurali provviste di microfono dietro l'altoparlante<sup>1</sup>, in modo da registrare il segnale che sarebbe entrato nel canale uditivo. Essendo disposti appena all'esterno del condotto uditivo, i microfoni cattureranno l'impulso solo dopo esser stato processato tramite effetti di riflessione, diffrazione e risonanza a causa della presenza di busto, testa e la pinna, come mostrato in Figura 4.1 [38]. Tale segnale trasformato può essere trattato come una prima forma grezza di BRIR (raw BRIR). È da notare che questa risposta impulsiva è la personale risposta di soggetto e ambiente assieme, il che rende le HRIR in esso contenute completamente individuali. Come segnale impulsivo è stato scelto uno sweep logaritmico che copre uniformemente in 1s tutte le frequenze tra 20Hz e 20kHz in modo che la risposta impulsiva contenga informazioni per tutto lo spettro delle frequenze di interesse.

In Figura 4.2 è mostrato il progetto in questione all'interno di un sistema di mobile Audio Augmented Reality (mAAR), una delle possibili applicazioni di SelfEar. Grazie

<sup>1</sup>Nel nostro caso un paio di cuffie in-ear Roland CS-10EM, dal costo di circa 70 euro.



**Figura 4.2:** *Visione schematica del progetto SelfEar in un sistema mAAR.*

all'usabilità che una applicazione per dispositivi mobile possiede, essa rappresenta il mezzo ideale per realizzare questa alternativa. Tra i vantaggi principali che ogni app presenta intrinsecamente, i più utili al caso dell'acquisizione HRTF sono i seguenti:

**Accessibilità:** Grazie alla larga diffusione di dispositivi mobile, SelfEar sarebbe facilmente accessibile da grandi masse di utenti;

**Versatilità:** Un'applicazione per smartphone o tablet offre una elevata facilità di utilizzo, che accorcia notevolmente i tempi necessari per il processo di acquisizione delle HRTF, e la mobilità che contraddistingue questo tipo di dispositivi;

**Economicità:** Se commercializzata, un'app come SelfEar costerebbe pochi euro, ai quali andrebbero ad aggiungersi quelli per un paio di cuffie binaurali: una cifra comunque irrisoria rispetto a quelle che si spenderebbero per ottenere delle HRTF in altri modi. Naturalmente, per quanto bene sia stato progettato il sistema di speaker del dispositivo, la qualità del segnale in uscita da esso non sarà mai paragonabile a quella che si può ottenere con degli altoparlanti professionali e appositamente progettati. Inoltre, bisogna tenere conto del fatto che il processo tramite SelfEar non avviene in generale in un ambiente controllato come una camera anecoica o silente, ma può aver luogo ovunque. Questo comporta l'introduzione di rumore e agenti disturbanti spesso non trascurabili che rendono meno affidabile ed accurato il calcolo della risposta e che nel caso più generale possono essere solo in parte compensate da elaborazioni post-processing. Tuttavia, vogliamo sottolineare il basso costo della soluzione proposta con SelfEar in relazione a quello delle altre finora utilizzabili, in particolare di quella di misurazione in camera anecoica. Il divario che esiste tra i costi di queste due alternative è paragonabile alla differenza che esiste tra la qualità delle risposte che gli stessi metodi offrono. Nel-

lo specifico SelfEar riesce ad acquisire HRTF individualizzate, che necessitano di una ulteriore elaborazione per poter diventare individuali.

## 4.2 Funzionamento dell'App

- In prima implementazione, SelfEar è programmata per la rilevazione delle risposte all'impulso lungo il piano mediano verticale, o sagittale, mantenendo l'azimuth costante a 0, lungo cioè la direzione con il più alto contenuto di caratteristiche spettrali. Per ottenere delle HRTF il più accurate possibile non basta un'unica registrazione, anche perché si porrebbe il problema di quale punto del piano mediano scegliere, bensì sono necessari più campioni.
- L'applicazione guida l'utente attraverso una procedura atta a registrare le risposte per diverse elevazioni rispetto all'elevazione zero, ovvero il livello del canale uditivo del soggetto.
- Concretamente, la procedura consiste nel puntare queste elevazioni, chiamati bersagli, con gli speaker sempre rivolti verso la propria testa e il braccio teso, muovendolo lungo la circonferenza che disegna sul piano mediano. Nonostante il fatto che siano necessari diversi feedback a diverse altezze per ottenere funzioni di trasferimento sufficientemente buone, è attualmente in sviluppo una schermata di configurazione in cui è lasciata all'utente la scelta di quanti e quali elevazioni considerare. Le elevazioni attualmente utilizzate sono un che compongono questo sottoinsieme di quelle impiegate dal database di HRTF CIPIC [19] e sono comprese nell'intervallo  $[45, +45]$ , in quanto ritenuto un intervallo comprendente elevazioni particolarmente influenti per il contributo alle HRTF, e sono spaziate con un passo di 5.625, in accordo con le specifiche CIPIC [19].
- Al lancio dell'applicazione, all'utente viene chiesto di scegliere la posizione degli altoparlanti nel dispositivo che sta usando. Questa scelta inuenerà l'orientamento del dispositivo, dal momento che sarà necessario rivolgere gli speaker verso la testa del soggetto durante la fase di acquisizione, al ne di massimizzare la potenza del segnale ricevuto dai microfoni auricolari.
- Una volta selezionata la corretta posizione, viene mostrata all'utente la schermata principale, in cui avviene la procedura di acquisizione vera e propria.
- Successivamente alla scelta del tasto "Start" ha inizio la procedura, esemplificata dallo schema di Figura 4.3, che si compone dei seguenti passi ordinati:

1. *Raggiungimento del bersaglio*: l'attuale elevazione del dispositivo rispetto all'orizzonte,  $\phi_i$ , compare sopra all'elevazione bersaglio da raggiungere. SelfEar elabora i dati provenienti dall'accelerometro lungo i tre assi cartesiani,  $a_{x,y,z}$ , del dispositivo per calcolare  $\phi_i$  tramite la formula seguente:

$$\phi_i = \arctan\left(\frac{\pm a_y}{|a_z|}\right)$$

nel caso gli altoparlanti siano posizionati sul lato superiore o inferiore; mentre con la formula:

$$\phi_i = \arctan\left(\frac{\pm a_z}{|a_y|}\right)$$

nel caso si trovassero invece sul lato frontale o sul retro. Il segno del numeratore è:

- + per altoparlanti sul lato inferiore o sul retro;
- per altoparlanti sul lato superiore o frontale.

La sequenza delle elevazioni bersaglio spazia in ordine crescente tra gli angoli del CIPIC database equispaziati ogni  $5.625^\circ$  in un intervallo di  $[-40^\circ, 40^\circ]$ . Un beep ausiliario rende acusticamente l'informazione di differenza tra l'attuale elevazione del dispositivo e quella bersaglio, agevolando la procedura di puntamento soprattutto nei casi in cui lo schermo del dispositivo non è visibile a causa della posizione degli speaker (ad es. quando si trovano sul retro). La pausa tra un beep ed il successivo è direttamente proporzionale alla differenza tra  $\phi_i$ , e l'angolo bersaglio,  $\hat{\phi}_i$ , come mostrato nell'equazione seguente:

$$pause_i = |\phi_i - \hat{\phi}_i| \cdot k$$

dove  $i$  è un istante in cui un singolo beep termina la sua riproduzione e  $k$  è una costante che rende percepibile la pausa tra due beep consecutivi.<sup>2</sup> L'obiettivo di questo step è di avvicinarsi a  $\hat{\phi}_i$  con una precisione di  $\pm 1^\circ$ . Questo passo può venire interrotto e ripreso su richiesta dell'utente.

2. *Controllo della posizione*: una volta che  $\phi_i$  entra nel range di validità di  $\pm 2^\circ$  dal bersaglio, scatta un timer di stabilizzazione di 2 secondi; nel caso l'utente uscisse dal range per tre volte prima che il timer finisca, la procedura inizia nuovamente a partire dalla fine del passo 1.

---

<sup>2</sup>La formula restituisce un valore in millisecondi, che senza una costante moltiplicativa darebbe luogo ad una pausa troppo corta per poter essere percepita. In questa realizzazione è stato scelto un valore arbitrario di  $k = 5$ .

3. *Riproduzione dello sweep*: al termine del timer di stabilizzazione, viene riprodotto lo stimolo sonoro dagli altoparlanti del dispositivo; nel caso l'utente uscisse dal range di validá di  $\pm 2^\circ$  solamente una volta durante la riproduzione dello sweep, la procedura di ricerca di  $\hat{\phi}_i$  viene annullata.
4. *Salvataggio delle BRIR*: quando uno sweep termina con successo, il segnale audio registrato viene salvato localmente in memoria insieme all'angolo di elevazione a cui si riferisce; la procedura ritorna quindi al passo 1 con la successiva elevazione bersaglio in sequenza.
5. *Fine della sessione*: una sessione termina quando tutte le elevazioni bersaglio sono state raggiunte con successo.

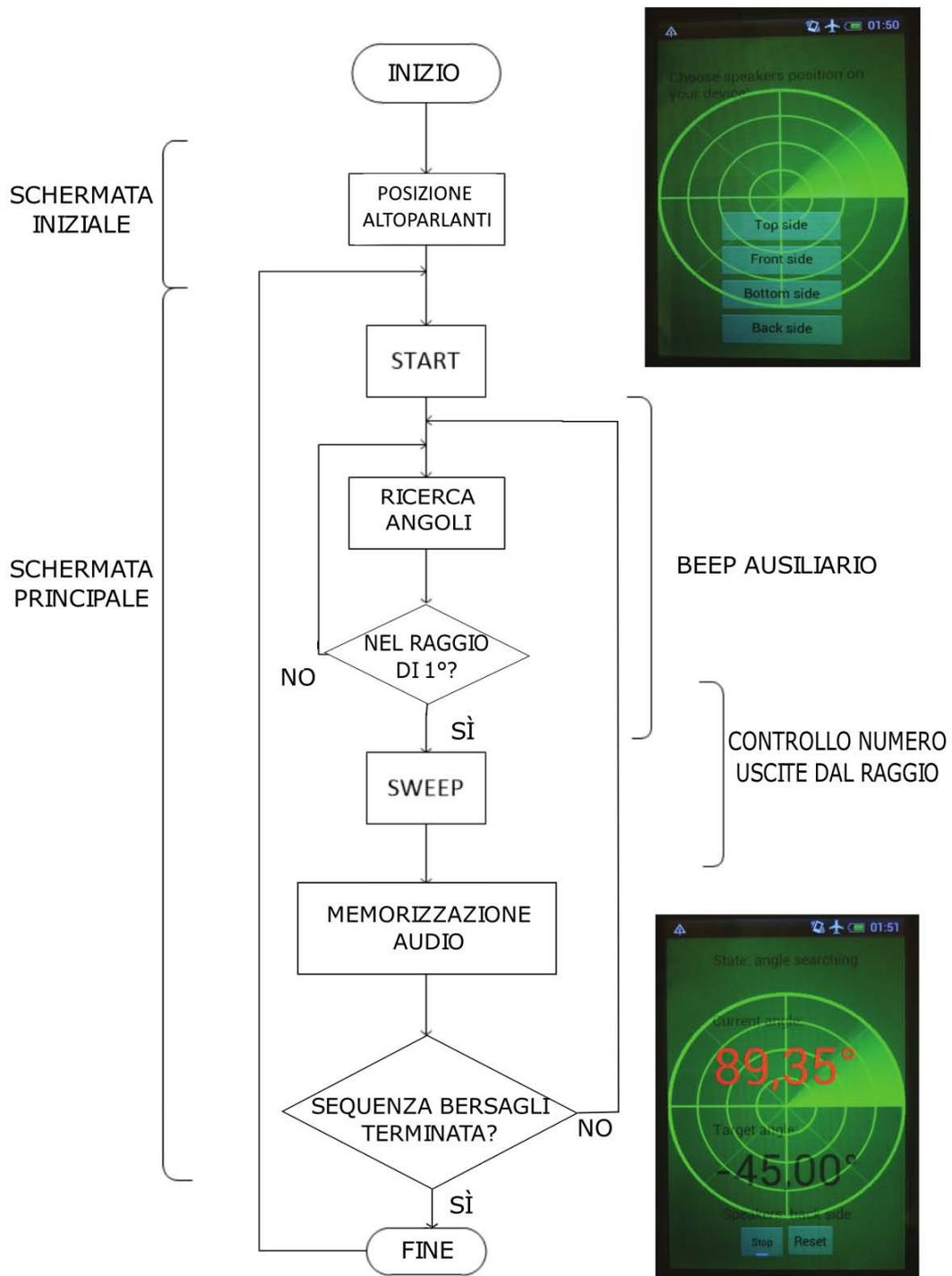


Figura 4.3: Diagramma a blocchi per la procedura di acquisizione delle risposte allo sweep.



# Capitolo 5

## Validazione del progetto SelfEar

Viene proposta di seguito una sessione di misurazione, effettuata utilizzando l'applicazione SelfEar, con l'obiettivo di acquisire le caratteristiche spettrali di un manichino KEMAR che verranno confrontate con i risultati ottenuti con una procedura professionale eseguita in ambiente anecoico. Tale confronto ci permette di discutere la qualità delle risposte acquisite mediante SelfEar, validandone l'applicabilità almeno in ambiente controllato.

Le sessioni sperimentali sono state eseguite in un ambiente non anecoico usando un manichino KEMAR minimizzando gli errori dovuti al movimento che si sarebbero manifestati con un eventuale soggetto umano. Si è posta quindi maggior concentrazione sulla direzione con il più alto contenuto di caratteristiche spettrali, ovvero la direzione frontale [39],  $\theta = 0$  e  $\phi = 0$ , fornendo un'analisi dettagliata dei segnali acquisiti nelle varie sessioni di misurazione. Infine viene riportata una valutazione qualitativa dell'applicazione SelfEar per una serie di PRIR nel piano mediano frontale.

### 5.1 Hardware

**Camera silente** Le misurazioni sono state condotte all'interno di una camera silente di dimensioni  $2 \times 2$  m. Si tratta della Sound Station Pro 45 by Puma s.r.l di Figura 5.1. In Figura 5.2 si possono osservare le caratteristiche di abbattimento sonoro e di riverbero.

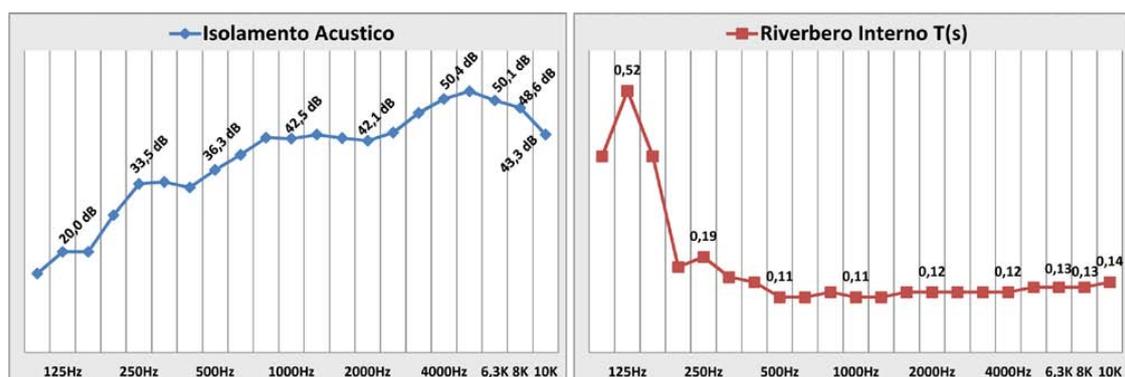
**Scheda audio** La scheda audio utilizzata per le registrazioni è la Motu 896mk3<sup>1</sup> visibile in Figura 5.3. Essa permette registrazioni fino a 192 kHz con una profondità di 24 bit, arriva ad avere sino a 28 input e 32 output (a 44.1/48 kHz) e permette operazioni stand-alone direttamente dal display incorporato. Ha una protezione contro i clip incor-

---

<sup>1</sup><http://motu.com/products/motuaudio/896mk3>



**Figura 5.1:** Cabina silente: Sound Station Pro 45



**Figura 5.2:** Caratteristiche della Sound Station Pro 45



**Figura 5.3:** *Motu 896mk3*

porata e moltissime altre features non rilevanti al fine dell'utilizzo che ne è stato fatto durante le misurazioni.

**Headset** L'headset utilizzato per le misurazioni è costituito dalle cuffiette Roland CS-10EM<sup>2</sup> visibili in Figura 5.5. Esse sono equipaggiate con due microfoni a condensatore omnidirezionali con risposta in frequenza che va da 20 Hz a 20 kHz. Le cuffiette hanno invece una risposta in frequenza che va da 15 Hz a 22 kHz. Per poter funzionare, i microfoni delle Roland CS-10EM necessitano di una alimentazione DC, questa può essere fornita dalla presa del microfono dell'attrezzatura che si usa per la registrazione o mediante l'utilizzo di un adattatore esterno. Nel nostro setup è stato utilizzato l'adattatore A3 dei microfoni binaurali Soundman OKM II Klassik solo<sup>3</sup>. Con la sua capacità di adattamento automatico di livello, la sua alta tensione di alimentazione del microfono e rumore di fondo molto basso, l'Adattatore A3 offre un significativo aumento della dinamica rispetto ad un convenzionale collegamento con una presa da microfono. Esso si collega tra il microfono e l'apparecchiatura per la registrazione fornendo un'alimentazione phantom di 6 volt ai microfoni, garantendo una risposta in frequenza da 20 Hz a 50 kHz. L'adattatore è visibile in Figura 5.5

**KEMAR** Il manichino KEMAR (Knowles Electronics Manikin for Auditory Research)<sup>4</sup> [17] visibile in Figura 5.6 è uno strumento acustico che permette misurazioni riproducibili per determinare le prestazioni di apparecchi acustici e di altri dispositivi elettroacustici, nonché la qualità di registrazioni binaurali. La testa ed il tronco si basano sulla

<sup>2</sup><https://www.roland.com/us/products/cs-10em/>

<sup>3</sup><http://www.soundman.de/it/products/>

<sup>4</sup><http://www.gras.dk/products/head-torso-simulators-kemar.html>



**Figura 5.4:** Cuffiette Roland CS-10EM



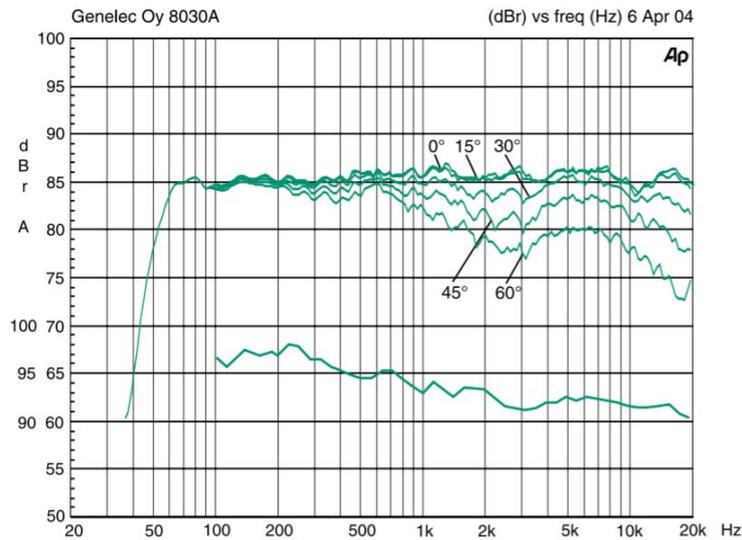
**Figura 5.5:** Adattatore A3 Soundman



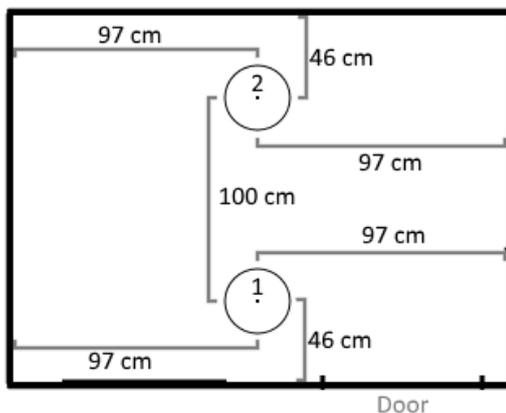
**Figura 5.6:** KEMAR

media delle dimensioni degli uomini di sesso maschile e femminile di tutto il mondo. Essa soddisfa i requisiti ANSI S3.36 / ASA58-1985 e IED 60959: 1990. La parte di orecchio esterno della KEMAR è componibile e vi sono varie configurazioni disponibili, generalmente al posto del timpano è posizionato un microfono professionale, facente sempre parte della struttura KEMAR. Tale microfono può essere posizionato all'ingresso del canale uditivo oppure alla fine, in questo secondo caso l'orecchio viene provvisto di un simulatore di canale uditivo.

**Loudspeaker** Il Loudspeaker utilizzato nel setup di misurazione è un Genelec 8030A la cui risposta in frequenza è visibile nella Figura 5.7 e va da 58 Hz a 20 kHz. Nella figura, le curve nella parte superiore mostrano le caratteristiche del Loudspeaker quando posizionato a diversi angoli di azimuth sul piano orizzontale alla distanza di un metro dalla sorgente. La curva nella parte inferiore mostra la risposta di potenza dei sistemi del loudspeaker.



**Figura 5.7:** Risposta in frequenza del Loudspeaker Genelect 8030A



**Figura 5.8:** Setup per le misurazioni. Posizioni di sorgente e ricevitore.



**Figura 5.9:** Setup di misurazione di Selfear, con selfie-stick incorporato.

**Dispositivo mobile** Si tratta di uno smartphone HTC Desire C<sup>5</sup> per il quale è stata realizzata una struttura che lo sostiene che simula un braccio umano, realizzata tramite un selfie stick;

## 5.2 Setup

La Figura 5.8 mostra l'interno della cabina silente dove sono identificabili due posizioni distanziate tra di loro di un metro e centrate all'interno della camera silente: la posizione #1 per le varie sorgenti acustiche e la posizione #2 per i vari ricevitori.

<sup>5</sup><http://www.htc.com/it/smartphones/htc-desire-c/>

Negli esperimenti sono stati usati due tipi di dispositivi di riproduzione (con definizione acronimi):

- *L* : il loudspeaker Genelec 8030A;
- *S* : lo smartphone HTC Desire C sorretto da una struttura autoprodotta che simula un braccio umano, realizzata con un selfie stick;

E due tipi di ricevitori (con definizione acronimi):

- *H* : l'headset Roland CS-10EM con due microfoni incorporati;
- *K* : due microfoni professionali G.R.A.S forniti dal simulatore di testa e torso KEMAR; nella configurazione proposta, l'orecchio destro era equipaggiato con il simulatore di canale uditivo, mentre l'orecchio sinistro ne era sprovvisto.

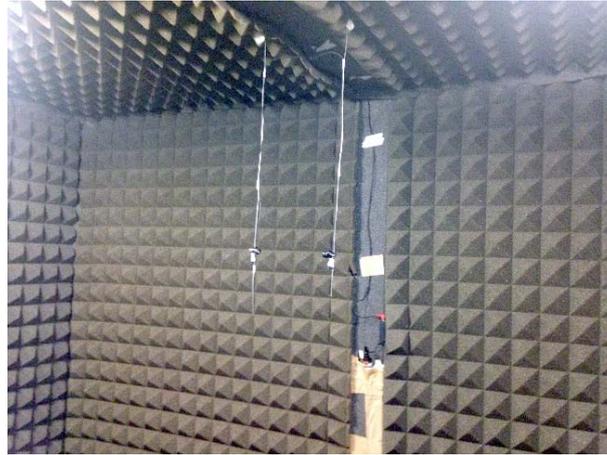
In tutti gli esperimenti, sorgenti e ricevitori, nelle rispettive posizioni, sono sempre stati posizionati alla stessa altezza dal pavimento. Il segnale sorgente utilizzato è un sine sweep logaritmico della durata di un secondo, il cui contenuto in frequenza spazia uniformemente da 20 Hz a 20 kHz. I segnali acustici sono stati registrati con il software Audacity attraverso l'interfaccia audio Motu 896 mk 3 e sono stati poi processati usando il software Matlab aggiornato alla versione 8.4.

**Calibrazione** - Il loudspeaker e lo smartphone sono stati calibrati per ottenere il miglior SNR possibile senza introdurre distorsioni. Per fare ciò è stato utilizzato un fonometro professionale con un tono di test di 500 Hz. Fonometro e loudspeaker prima e smartphone poi sono stati collocati nella camera silente all'interno della quale sono avvenute le misurazioni del fonometro. Il loudspeaker ha raggiunto i 94 dB SPL mentre lo smartphone, seppur con il volume al massimo è riuscito a raggiungere solo i 51 dB SPL.

## 5.3 Misurazioni

### 5.3.1 Diffuse field

Per il calcolo della risposta *diffuse-field* è stata realizzata una struttura apposita composta da due cavi di filo metallico agganciati al soffitto della SSP distanziati di 17.4 cm (Figura 5.10), ovvero la distanza tra i microfoni della KEMAR. Tale struttura è stata collocata in posizione #2 all'interno della camera silente, mentre in posizione #1 sono stati collocati, prima uno e poi l'altro, il loudspeaker e lo smartphone. Nella struttura realizzata sono stati collocati, prima uno e poi l'altro, le Roland e i microfoni della KEMAR. Sono state quindi acquisite le misurazioni diffuse-field per tutte le combinazioni



**Figura 5.10:** *Struttura per calcolo diffuse field con Roland*

di sorgente e ricevitore, ottenendo un totale di quattro misurazioni. Tali misurazioni permettono di acquisire le specifiche acustiche dell'ambiente e del setup.

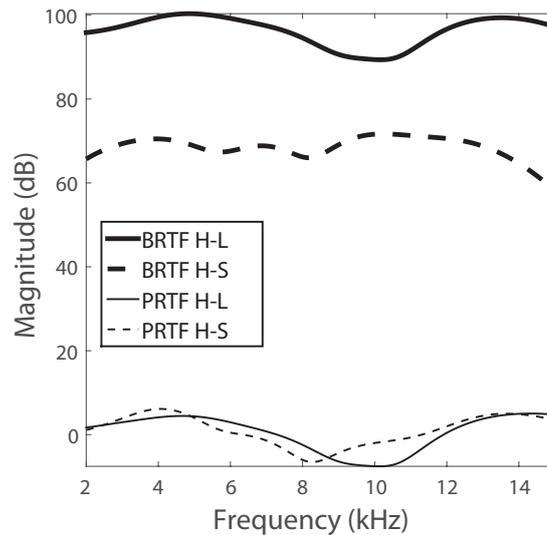
### 5.3.2 Sessioni di misurazione

**Prima sessione di misurazione** - In questa sessione sono stati posizionati all'interno della SSP, nelle rispettive posizioni (#1 e #2 della Fig. 5.8), lo speaker Genelec e la KEMAR. Come primo step sono state misurate le risposte dell'orecchio destro e sinistro. Successivamente è stata inserita nel canale uditivo destro della KEMAR la cuffietta Roland; sono state effettuate dieci misurazioni con questa configurazione riposizionando ogni volta le Roland; in tal modo è sarà possibile analizzare la variabilità introdotta nelle misurazioni dalla posizione del microfono.

**Seconda sessione di misurazione** - In questa sessione è stata posizionata in #1 (vedi Fig. 5.8) la struttura di sostegno per lo smartphone; nella posizione #2 di Fig. 5.8 è invece stata posizionata la KEMAR. Il selfie-stick manteneva lo smartphone ad una distanza di un metro dalla KEMAR e consentiva una regolazione angolare sul piano mediano. Sono state acquisite le misurazioni relative a 15 angoli tra  $-40^\circ$  e  $+40^\circ$  sul piano mediano riproponendo gli angoli presenti nel CIPIC database. Quindi, al termine della misurazione si sono ottenuti due set di 15 misurazioni per l'orecchio sinistro della KEMAR (senza cuffietta) e per il microfono destro della cuffietta.

## 5.4 Analisi

Per ogni misurazione sono stati applicati alcuni algoritmi di signal processing visibili in Appendice B sezione B.5, in particolare:



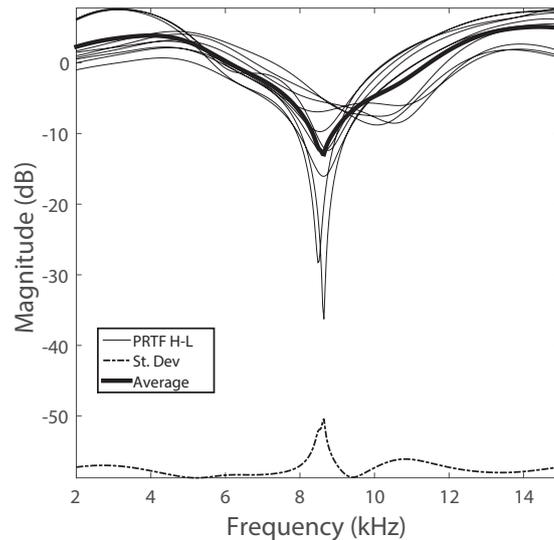
**Figura 5.11:** Confronto tra i moduli (in dB SPL) delle BRTF (linee spesse) e le relative PRTF (linee sottili) ottenute usando: come ricevitore - il microfono destro delle cuffie Roland (H); come sorgente - il loudspeaker dello smartphone (S, linee tratteggiate) e il loudspeaker Genelec (L, linee continue).

- è stato calcolato l'onset applicando la funzione di cross-correlazione con il segnale di sweep originale;
- è stata estratta la BRIR deconvolvendo la risposta allo sweep con lo sweep stesso;
- sono state rimosse riflessioni dovute dalla SSP e dalla strumentazione al suo interno, tale procedura è stata effettuata sottraendo alle BRIR le rispettive risposte di tipo *diffuse-field*.

Da questa procedura sono state estratte le HRTF. Mentre le PRTF sono state ottenute applicando ad ogni risposta impulsiva HRIR una finestra di hanning di 1 -ms (48 campioni) centrata temporalmente sul picco massimo e successivamente sono state normalizzate rispetto al valore massimo in ampiezza [40]. Tutte le PRTF normalizzate sono state filtrate con un passa banda tra 2 kHz and 15 kHz, mantenendo i picchi ed i notch dovuti alla presenza della pinna.

## 5.5 Risultati

La Figura 5.11 mostra il confronto tra i moduli in dB SPL delle BRIR estratte dalle misurazioni usando come sorgente (i) il loudspeaker Genelec (ii) lo smartphone, e come ricevitore il microfono destro delle cuffie Roland inserito nell'orecchio destro della KEMAR. Si può notare una differenza di 30 dB tra la SPL dei due loudspeaker, tale differenza porta ad un SNR minore mentre si usa lo speaker dello smartphone. Nella stessa



**Figura 5.12:** Moduli delle PRTF per i dieci riposizionamenti della cuffia Roland sul manichino KEMAR. La linea spessa rappresenta l'ampiezza media. La deviazione standard è spostata di  $-60$  dB per agevolarne la lettura.

PRTF	H-L	H-S	K-L(right)	K-L(left)
H-L	0	1.79	6.92	5.25
H-S		0	7.35	4.64
K-L(right)			0	5.47
K-L(left)				0

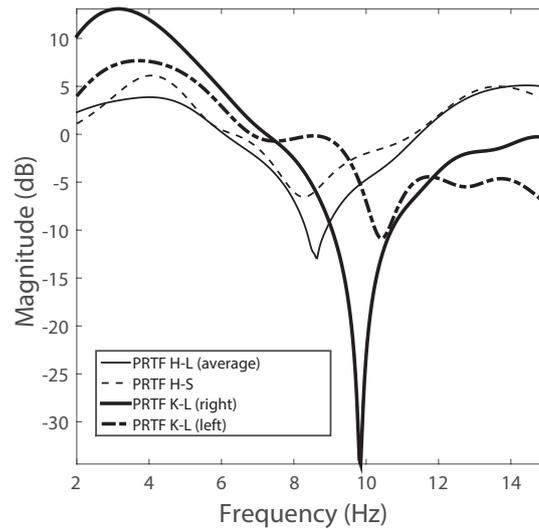
**Tabella 5.1:** Distorsione spettrale tra PRTF di Figure 5.13. Tutti i valori in dB.

figura sono mostrate le due PRTF normalizzate corrispondenti in modo da poter valutare il contributo introdotto dalla diffuse-field. Tale contributo è sensibilmente visibile per le misurazioni ottenute con lo smartphone HTC Desire C, dovuto principalmente alla distorsione non trascurabile dello speaker a basso costo integrato.

In Figura 5.12, sono riportati i moduli in dB delle PRTF dei dieci riposizionamenti insieme alla loro media. Vi è anche riportata la deviazione standard in modo da poter analizzare la variabilità introdotta nelle misurazioni dalla posizione delle cuffie/microfono. La variabilità massima la si osserva in prossimità dei notch salienti della PRTF, tra 9 e 11 kHz, che sono molto sensibili ai cambiamenti topologici tra la cuffie e la struttura dell'orecchio [41].

Le principali valutazioni sono state effettuate nella posizione frontale,  $\phi = 0$ , comparando le PRTF normalizzate in condizioni differenti. La Figura 5.13 mostra il confronto tra il modulo delle PRTF acquisite con e senza cuffiette, coinvolgendo sia la Genelec che lo smartphone.

Per le quattro PRTF è stata calcolata la distorsione spettrale media (SD) [42] tra tutte le coppie in un range di frequenze tra  $2 \text{ kHz} \leq e \leq 15 \text{ kHz}$  (i valori sono visibili in



**Figura 5.13:** Confronto tra i moduli delle seguenti PRTF: a) PRTF media della Fig.5; b) sorgente: smartphone - ricevitore: microfoni delle cuffie ; c) sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio destro della KEMAR (con canale uditivo); d) sorgente: loudspeaker Genelec - ricevitore: microfono dell'orecchio sinistro della KEMAR (senza canale uditivo).

Tabella 5.1). Questi confronti permettono diverse considerazioni:

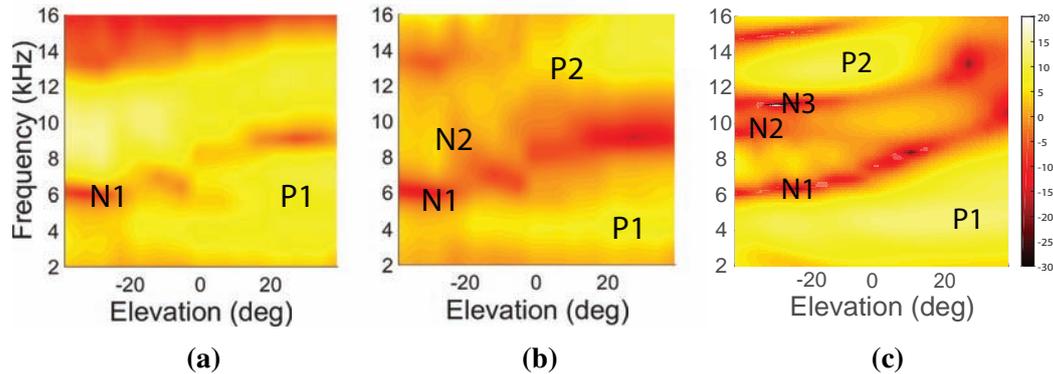
- Acustica della pinna,  $K-L_{right}$  vs.  $K-L_{left}$ : le forme diverse delle orecchie e l'acustica del canale uditivo del KEMAR differiscono notevolmente; questa differenza è visibile in tutti i confronti che coinvolgono  $K-L_{left/right}$ ;
- Loudspeaker,  $H-S_{right}$  vs.  $H-L_{right}$ : entrambi i loudspeaker introducono una distorsione spettrale trascurabile ( $< 2$  dB);
- Procedura SelfEar,  $H-S_{right}$  vs.  $K-L_{left}$ : la differenza tra l'acquisizione SelfEar delle PRTF e un sistema di misura tradizionale introducono l'errore in SD più basso tra le coppie analizzate (rimuovendo il confronto di controllo sui loudspeaker);

6

La Figura 5.14 permette di confrontare visibilmente i risultati ottenuti con la procedura SelfEar negli angoli considerati (con e senza la compensazione di tipo *diffuse-field*) e le misurazioni CIPIC nello stesso range di angoli per il soggetto 165 (KEMAR). I dati sono stati interpolati per avere una transizione visivamente più fluida.

È già noto da Christensen *et al.* [43] che sia la posizione del ricevitore, sia il suo spostamento dal punto di misura dell'HRTF ideale, cioè al centro dell'ingresso del canale uditivo, influenzano non indifferentemente i pattern di direttività dell'HRTF per

<sup>6</sup>Va notato che questo confronto è stato effettuato su orecchie diverse per ragioni pratiche. Di conseguenza i valori di SD potrebbero avere una differenza ancor più inferiore.



**Figura 5.14:** *PRTF nel piano mediano. (a) acquisizione con SelfEar senza compensazione; (b) acquisizione con SelfEar compensata con diffuse-field; (c) CIPIC KEMAR, Subject 165 - con compensazione free-field. Vengono visivamente identificati i principali picchi (P1-2) e notch (N1-3), ove presenti.*

frequenze superiori a 3 – 4 kHz. Il lavoro effettuato è in accordo con le loro misurazioni in quanto mostrano uno spostamento della frequenza del notch centrale anche di 2 kHz con una elevata variabilità del modulo nei vari riposizionamenti del microfono (vedere la deviazione standard di Fig. 5.12) con una differenza massima di 10 dB.

Gli spostamenti delle frequenze centrali di picchi e notch sono osservabili anche dalla Fig. 5.13 e sono principalmente dovuti a differenze topologiche tra il punto di osservazione, dipendente dalla posizione del microfono, e gli oggetti acustici formati dalla presenza o assenza del canale uditivo e da orecchie destra e sinistra differenti.

Avere a disposizione un ampio intervallo di elevazioni frontali permette ad ogni sistema di misura di acquisire le caratteristiche rilevanti delle PRTF [39, 40]: ovvero le due principali risonanze (P1: modo omnidirezionale, P2: modo orizzontale) e i tre principali notch (N1-3 corrispondenti alle principali riflessioni della pinna). Le PRTF provenienti dal CIPIC KEMAR (vedere Fig. 5.14(c)) contengono tali caratteristiche; in particolare P1 ha una frequenza centrale di 4 kHz e P2 di 13 kHz; inoltre N1 si muove tra 6 e 9 kHz, N3 tra 11.5 e 14 kHz con un graduale incremento in elevazione; infine N2 che inizialmente si trova a 10 kHz, progressivamente scompare man mano che si raggiunge la direzione frontale.

## 5.6 Conclusioni

L'applicazione SelfEar è in grado di acquisire P1 e N1 efficacemente, considerando sia il caso in cui le PRTF siano state compensate con la *diffuse field*, sia che non lo siano. Dal momento che l'ambiente ha un contributo non trascurabile, il confronto visivo tra Fig. 5.14(a) e (b) sottolinea l'importanza di poter estrarre con precisione le PRTF dalle

BRIR. In particolare dalla Fig. 5.14(b) si possono identificare anche P2 e, anche se in minima parte, N2. Tuttavia N3 è completamente assente, il che suggerisce la presenza di un'interferenza introdotta dalle cuffie nella concha della pinna. Seguendo il modello risonanza-più-riflessione per le PRTF [40, 42] possiamo ipotizzare che l'assenza di riflessi nella concha sia dovuta alla presenza della cuffia; inoltre il volume della concha in questa condizione è drasticamente ridotto, producendo così variazioni nei modi di risonanza della struttura della pinna [41]. Inoltre, comparando i valori della distorsione spettrale tra  $H - S$  vs.  $K - L_{left}$  si ha una differenza di 4.64 dB che suggerisce una buona affidabilità paragonabile al metodo di personalizzazione in [42] (valori di SD tra 4 e 8 dB) e allo stato dell'arte delle simulazioni numeriche in [41] (valori di SD tra 2.5 e 5.5 dB).

Vale la pena notare che i parametri di notch e picchi, cioè la frequenza centrale, il guadagno e la banda, possono essere direttamente calcolati dalle PRTF disponibili. Queste caratteristiche spettrali possono fornire i parametri a modelli di PRTF sintetiche e/o, seguendo un approccio di modellazione strutturale mista [16]), a procedura di selezione delle HRTF. Infine non vi è nulla che impedisca l'uso diretto di PRTF estratte da SelfEar per il rendering di audio binaurale. È comunque inopinabile che sarà necessaria una valutazione psico-acustica con soggetti umani per confermare l'affidabilità delle tecnologie del progetto SelfEar con l'obiettivo di creare sorgenti sonore virtuali esternalizzate fornendo HRIR efficaci ed individuali.

# Capitolo 6

## Interazione uomo-macchina in Selfear

Dal capitolo precedente si è concluso che l'App SelfEar permette l'individuazione di HRTF soggettive (secondo ad esempio con formulazione a modelli strutturali misti) in ambiente controllato, utilizzando come soggetto un manichino KEMAR. Resta da validare l'applicazione in altri tipi di ambiente con soggetti umani. In questo capitolo si descrivono delle sessioni sperimentali effettuate in quattro ambienti, controllato, domestico, riverberante ed esterno, al fine di testare l'applicazione nelle più diverse configurazioni d'uso, anche fuori da un ambiente di laboratorio. Viene qui proposta una procedura per il calcolo della componente diffuse field per l'estrazione dei parametri della PRTF.

### 6.1 Hardware

**Camera silente** Le misurazioni in ambiente controllato sono state condotte all'interno di una camera silente di dimensioni  $2 \times 2$  m. Si tratta della stessa camera silente descritta nella sessione 5.1

**Scheda audio** La scheda audio utilizzata per le registrazioni è la Fast Track Pro della M-Audio<sup>1</sup> visibile in Figura 6.1. Si tratta di una scheda audio progettata per fornire un'interfaccia audio e MIDI professionale e portatile per computer laptop o desktop. Essa permette registrazioni fino a 96 kHz con una profondità di 24 bit, possiede 2 linee di input e fino a 4 di output.

**Headset** L'headset utilizzato per le misurazioni è costituito dalle cuffiette Roland CS-10EM, lo stesso descritto nella sessione 5.1

---

<sup>1</sup><http://www.m-audio.com/>



**Figura 6.1:** Scheda audio Fast Track Pro

**Dispositivi mobile** Le misurazioni sono state effettuate utilizzando due smartphone:

- **Asus Zenfone 2** -Smartphone di fascia media con speaker posteriore posizionato in basso spostato a sinistra (guardando lo speaker). Aggiornato alla versione Android 5.0.1
- **Wiko Oggy** - Smartphone di fascia bassa con speaker posteriore posizionato in basso centrale. Aggiornato alla versione Android 4.2.

Tra i risultati del capitolo vi è un confronto fatto con l'HTC desire C del capitolo 5.

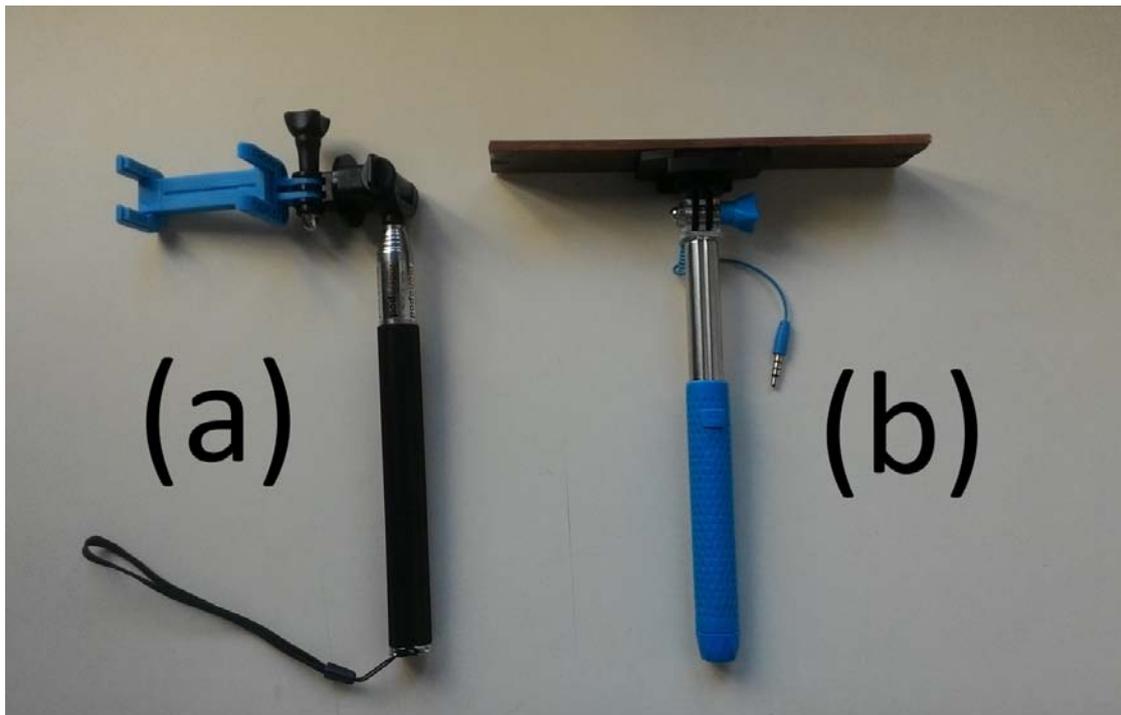
## 6.2 Misurazioni

Durante la sessione sperimentale si definiscono due misurazioni diverse: la prima relativa all'acquisizione della diffuse field, la seconda relativa all'esecuzione dell'App SelfEar. Tali misurazioni sono state effettuate in quattro ambienti:

- controllato
- domestico
- riverberante
- esterno

### 6.2.1 Setup

Per quanto riguarda l'acquisizione delle diffuse field è stato realizzato un supporto per le Roland (Figura ??), costituito da un selfie stick a cui è stato fissato un pezzo di compensato su cui adagiare le Roland, alla distanza esatta di 17,4cm, con il microfono che volge verso l'esterno come se fossero realmente indossate da qualcuno. Per quanto riguarda



**Figura 6.2:** *Setup: (a) supporto per lo smartphone; (b) supporto per le Roland;*

L'esecuzione di SelfEar è stato realizzato un supporto per lo smartphone combinando dei pezzi di un selfie stick con i pezzi di una action cam per ottenere l'oggetto di Figura ???. In esso viene collocato lo smartphone che viene impugnato dal soggetto durante l'esecuzione dell'App.

### 6.2.2 Sessioni sperimentali

**Acquisizione diffuse field** per l'acquisizione delle diffuse field si è ideata una procedura che permettesse all'utente di eseguire operazioni semplici ma allo stesso tempo efficaci. L'utente, una volta sistemate le cuffiette sull'apposita struttura, deve, tenendo la struttura in mano di fronte a se con il braccio disteso, riprodurre il segnale di sweep dallo smartphone che terrà con lo speaker rivolto verso le cuffie appena di fronte a se, alla stessa altezza delle cuffie (Figura ??). Registrando il segnale, le cuffie acquisiscono anche tutto ciò che concerne l'ambiente circostante e l'utente nell'ambiente. Rimane da validare se questo principio adottato eguaglia il calcolo delle diffuse field effettuato nel capitolo precedente in un ambiente controllato senza perturbazione dell'utente. Per la validazione sono state eseguite le seguenti prove:

1. All'interno della camera silente sono stati posti su due trepiedi la struttura che sostiene le Roland e lo smartphone posizionato in modo tale che lo speaker punti verso le cuffie e che entrambi siano alla stessa altezza. Sono state effettuate delle



**Figura 6.3:** *Procedura di acquisizione di diffuse field*

registrazioni del segnale di sweep riprodotto dallo smartphone e registrato dai microfoni delle Roland. L'esperimento è stato ripetuto scambiando smartphone con headset.

2. è stata effettuata una procedura completa di calcolo diffuse field secondo il procedimento e sono stati confrontati i risultati con le diffuse field acquisite nel capitolo precedente in camera silente.

Per ogni ambiente sono state acquisite 5 registrazioni di diffuse field.

**Esecuzione dell'App** Per l'esecuzione dell'applicazione è stato necessario realizzare una procedura tramite Pure Data che permettesse la registrazione dei segnali di sweep alle varie elevazioni in maniera automatica. È quindi stata modificata l'applicazione inserendogli una procedura di comunicazione tramite OSC con Pure Data. La patch realizzata con Pure Data è visibile in Figura ?? e si occupa semplicemente di acquisire le informazioni guidate dai comandi inviati dall'applicazione, in modo tale da registrare per ognuna delle elevazioni solo il segnale di sweep corretto nel corretto intervalli di

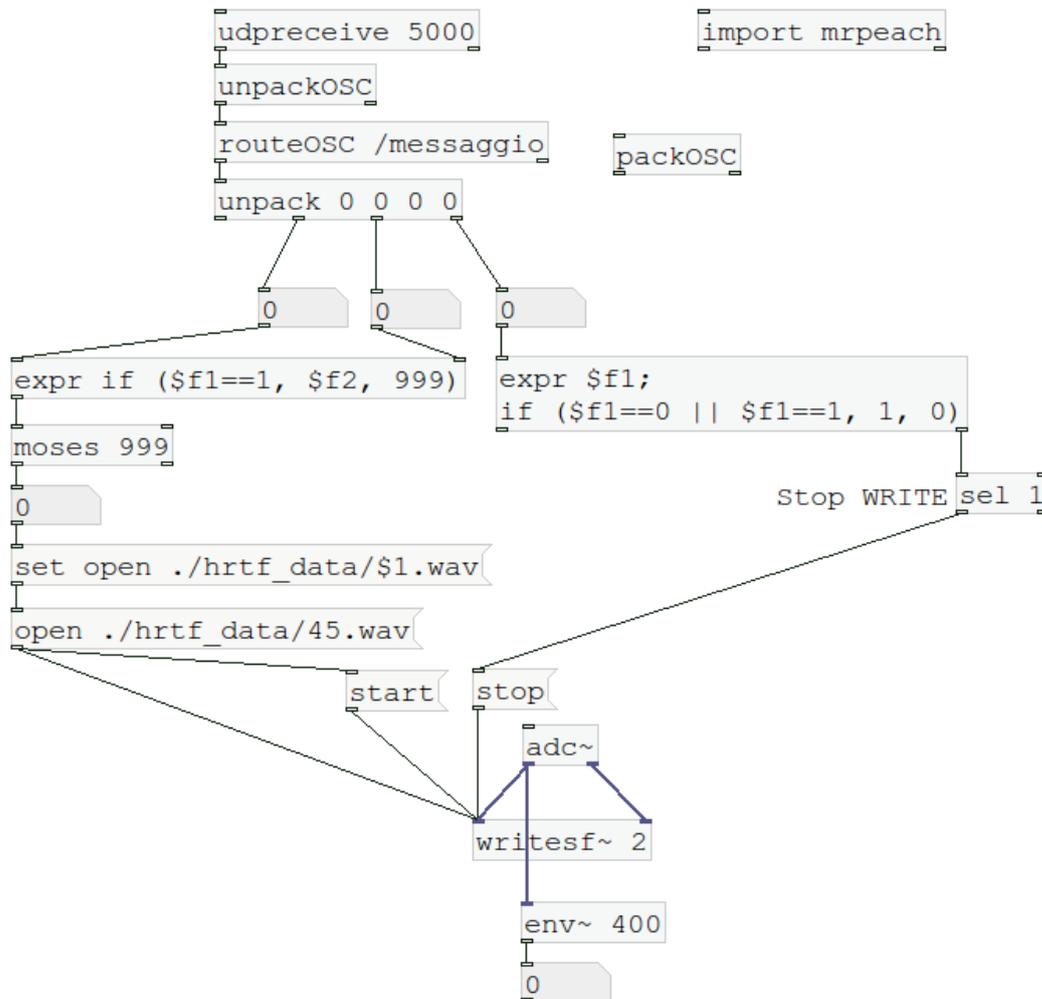
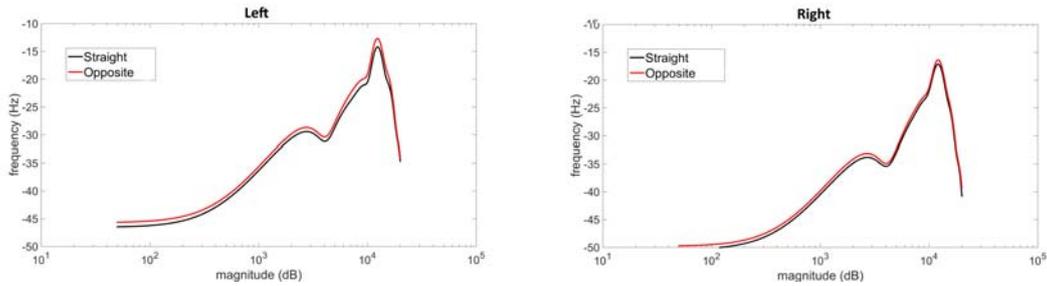


Figura 6.4: Patch per la registrazione dei segnali di SelfEar

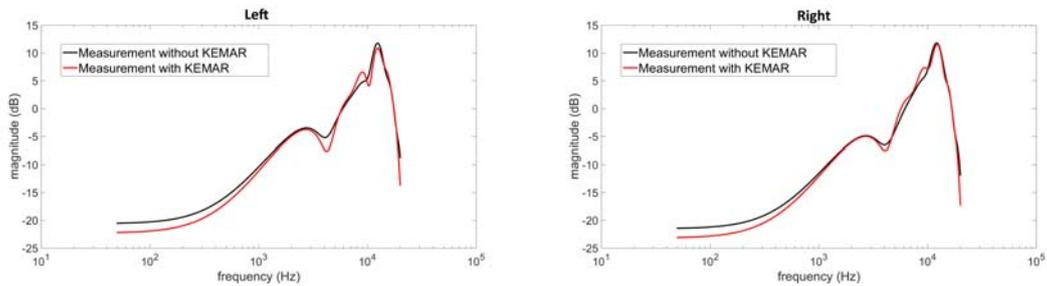
tempo senza inserire porzioni di segnale non utili, ma anzi potenzialmente dannose), per ognuna delle elevazioni. L'output finale è composto da una serie di file audio relativi a tutte le elevazioni misurate. Per ogni ambiente è stata eseguita la procedura sia con lo smartphone Zenfone 2, sia con il Wiko Iggy.

### 6.2.3 Analisi

Una volta ottenute le cinque registrazioni di diffuse field, attraverso la procedura *df\_variation.m* è stata effettuata la media per eliminare contributi rumorosi transitori. Attraverso le procedure *principe\_recip*, *kemar\_contribute* e *subject\_contribute* vengono rispettivamente mostrati il confronto tra le registrazioni del punto 1 della sessione sperimentale per l'acquisizione della diffuse field, e l'interferenza della KEMAR e di un soggetto nelle regi-



**Figura 6.5:** Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, in una posizione (straight) e nella posizione opposta (opposit) per entrambi canali (left e right). Smartphone: Zenfone 2



**Figura 6.6:** Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, con la presenza della KEMAR e senza. Smartphone: Zenfone

strazioni del punto 2. La procedura *Ambient.m* invece calcola la PRTF (sinistra) e mostra per l'ambiente scelto:

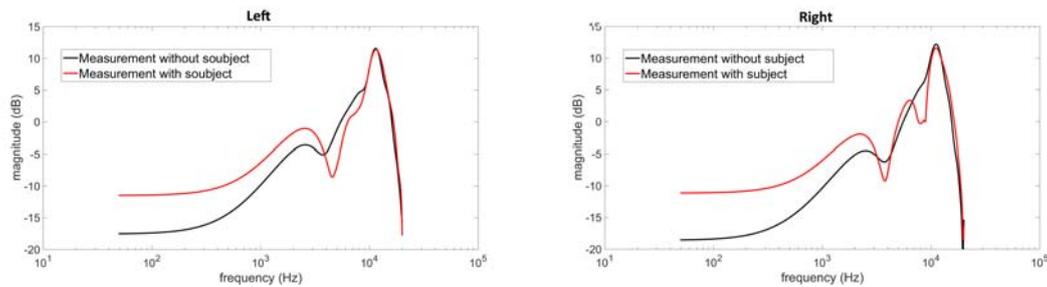
- la BRTF: il segnale elaborato non compensato
- la PRTF: il segnale elaborato compensato con contributo diffuse field

Le procedure di elaborazione per ottenere la funzione di trasferimento sono le stesse adottate nel capito precedente. Tutte le procedure qui elencate sono visibili in Appendice B.

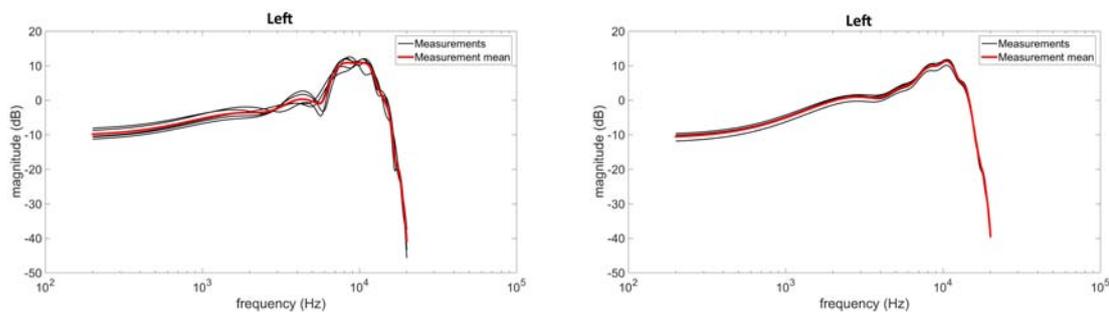
Nelle Figure 6.8 e 6.9 vengono riportati i grafici relativi alle risposte diffuse field (solo per orecchio sinistro) acquisite mediante smartphone wiko in due casi differenti:

1. smartphone attaccato al corpo
2. smartphone distanziato di 10cm dal corpo

Si può osservare come questa piccola variazione porti a risultati visibilmente differenti.



**Figura 6.7:** Medie delle risposte diffuse field acquisite in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto e senza. Smartphone: Zenfone



**Figura 6.8:** Misurazioni effettuate in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto. Smartphone: Wiko. Smartphone attaccato al corpo

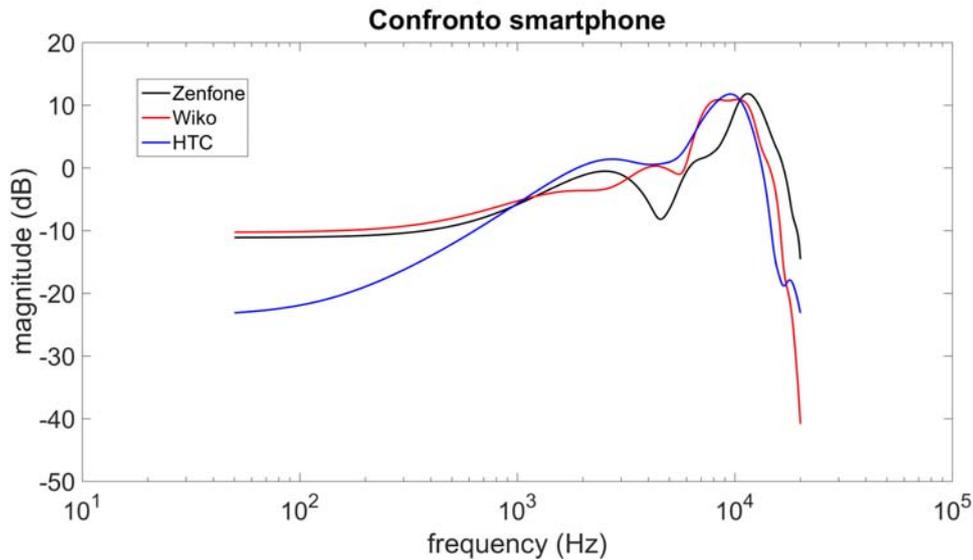
**Figura 6.9:** Misurazioni effettuate in camera silente con cuffie e smartphone su supporti, con la presenza di un soggetto. Smartphone: Wiko. Distanza dello smartphone dal corpo: 10 cm

### 6.3 Risultati

In Figura 6.5 sono visibili i risultati della misurazione 1 della procedura di acquisizione diffuse field, le curve plottate corrispondono alla media di cinque registrazioni acquisite. Per tale misurazione è stato utilizzato come smartphone solo lo Zenfone. In Figura 6.6 è visibile il contributo che apporta la KEMAR, più visibile nella figura di sinistra e descrivibile come un pronunciamento del notch posto a circa 4 kHz e uno a 9 kHz. In Figura 6.7 è visibile il contributo che apporta la presenza di un soggetto che effettua l'operazione di misurazione (quindi tenendo il cellulare di fronte al corpo e le cuffiette sull'apposita struttura in mano col braccio disteso). Il braccio con cui si stavano tenendo le cuffiette era il braccio destro. Anche in questo caso si osserva che l'apporto consiste nel pronunciamento dei notch posti a circa 4 e 9 kHz.

In Figura 6.10 vengono comparate le risposte diffuse field registrate dai due cellulari. Viene aggiunta anche la risposta acquisita dallo smartphone HTC Desire C nel Cap 5. Di queste tre curve viene calcolata anche la distorsione spettrale visibile in Tabella 6.1

In appendice A sezione A.1 sono visibili per ogni ambiente e per ogni smartpho-



**Figura 6.10:** confronto tra le acquisizioni di tre smartphone. L'acquisizione dello smartphone HTC è quella ottenuta nel cap. 5 (quindi ci si aspetta essere migliore delle altre)

	Zenfone	Wiko	HTC
Zenfone	0	5.04	7.43
Wiko		0	6.46
HTC			0

**Tabella 6.1:** Distorsione spettrale delle risposte diffuse field acquisite mediante smartphone.

ne: (i) l'acquisizione effettuata da SelfEar (BRTF) e (ii) la stessa compensata mediante risposta diffuse field (PRTF). I dati sono stati interpolati per avere una visualizzazione più fluida. In Figura 6.11 vengono mostrati i grafici relativi all'acquisizione in ambiente controllato. In appendice A Vengono anche mostrati due grafici relativi alle acquisizione della risposta diffuse field dei due smartphone, in tal caso si tiene nota del fatto che non è stata tenuta in considerazione la posizione dello smartphone durante l'acquisizione in quanto durante le registrazioni non erano ancora emerse le differenze evidenziate da Figura 6.8 e 6.9 dovute alla distanza tra utente e smartphone. Non sono stati condotti ulteriori esperimenti di approfondimenti rispetto a questo fenomeno che apparentemente sembra responsabile della presenza dei notch a circa 4 kHz e 9 kHz nella risposta diffuse field. Tale notch si osserva essere sempre presente invece nelle BRTF.

Il tempo di acquisizione di ogni misurazione è dell'ordine dei minuti, l'esecuzione completa di SelfEar ha quindi una durata dello stesso ordine.

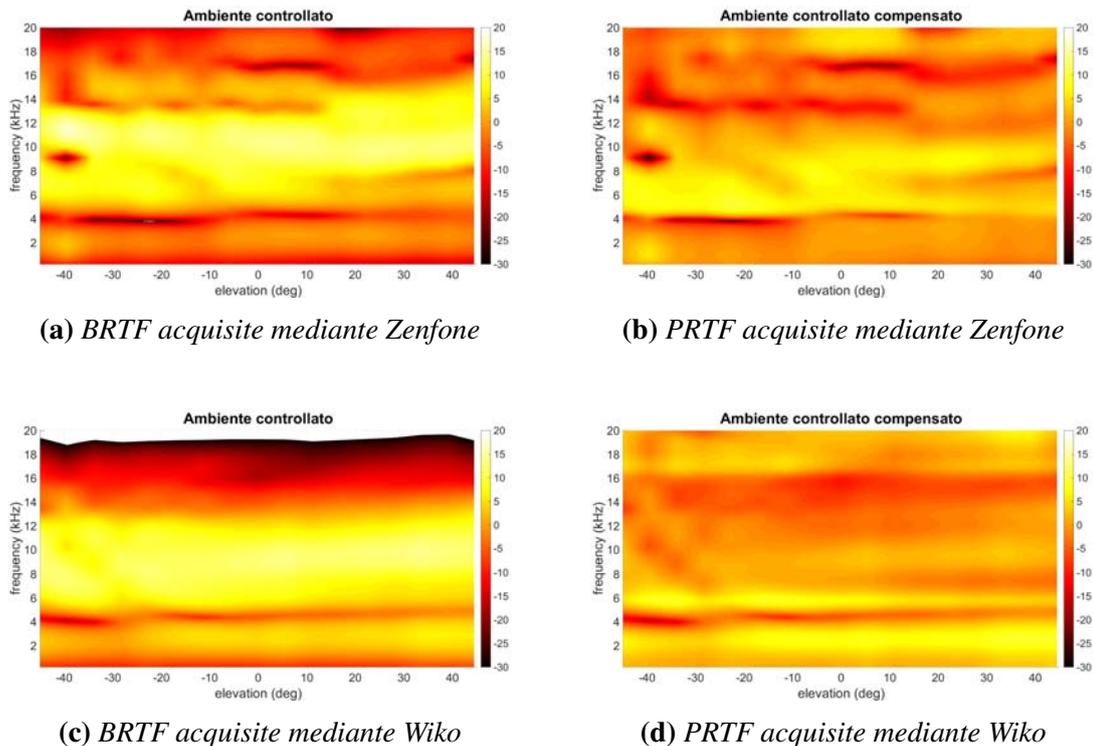


Figura 6.11: Risultati relativi ad ambiente controllato

## 6.4 Discussione

Dai grafici di Figura 6.5, è stato validato lubrificamente il principio di reciprocità su cui si basa l'acquisizione diffuse field di questa procedura, cioè sul fatto che l'inter-scambio tra cuffiette e smartphone non comporta l'introduzione di alcuna distorsione.

Dai grafici in Figura 6.6, 6.7 si osserva che le variazioni introdotte dalla KEMAR sono inferiori rispetto a quelle introdotte da un soggetto, tale effetto è spiegabile dal fatto che le misurazioni effettuate da un soggetto sono influenzate dai movimenti del soggetto stesso e dal modo in cui esso tiene il cellulare per la misurazione. Tali effetti sono maggiormente visibili nelle Figure 6.8 e 6.9 dove si osserva come una distanza di soli 10cm tra due posizioni del cellulare dia luogo ad un notch rilevante nella risposta misurata.

Le principali valutazioni sono state effettuate nella posizione frontale,  $\phi = 0$ , confrontando le PRTF normalizzate acquisite in ambienti differenti.

Le Figure A.1(riportata in Figura 6.11), A.4, A.7, A.10 permettono di confrontare visibilmente i risultati ottenuti con la procedura SelfEar negli angoli considerati (con e senza la compensazione di tipo *diffuse-field*) di tutti e quattro gli ambienti, mettendo in risalto gli elementi che definiscono la qualità delle PRTF misurate. In particolare si osserva come la compensazione tramite diffuse field permetta un avvicinamento ad una risposta calcolata con strumentazione professionale come quella mostrata in Figura 5.14

deg	notch 1	notch 2	notch 3
-45	8708,06	10775,17	16860,24
-33,75	8087,49	10180,70	15308,87
-22,5	7668,80	9374,69	11963,82
-11,25	7111,50	8637,62	10114,18
0	6332,72	8283,71	9602,73
11,25	6221,66	7994,88	9230,42
22,5	5892,18	7735,30	9076,68
33,75	5686,92	7404,36	9085,76
45	5310,90	7187,64	9983,44

**Tabella 6.2:** *Frequenze centrali dei notch alle varie angolazioni*

(c). Tuttavia nella maggioranza delle PRTF calcolate non si riesce ad avere una distinzione chiara delle caratteristiche rilevanti delle PRTF (tre notch e due picchi [39, 40]), si individua solo una zona tra 8 e 16 kHz in cui sono presenti i notch. Purtroppo non si ha un riscontro con una PRTF calcolata professionalmente e non si ha modo di affermare se tale zona individuata dalla procedura corrisponda nella realtà. Si hanno a disposizione i valori provenienti da una modellazione di sintesi condotta negli anni precedenti sul soggetto [42], visibili in Tabella 6.2 che confermano la presenza dei notch nella banda individuata, tali valori però non sono completamente affidabili non essendo completamente affidabile la procedura di sintesi applicata.

Per quanto riguarda il confronto tra gli ambienti si ha che le caratteristiche sono più pronunciate nella banda individuata nei casi (i) controllato e (ii) domestico. Lo smartphone Zenfone si rivela in questi due casi migliore rispetto al Wiko poiché permette una miglior distinzione dei picchi e dei notch.

Come ultima considerazione, ma non meno rilevante delle precedenti, si evidenzia la presenza di un notch fisso a circa 4 kHz in tutte le misurazioni. Non si hanno conferme sull'entità di tale caratteristica vista la natura esplorativa di questo ultimo lavoro, osservando però il fatto che esso è sempre presente nelle misurazioni BRTF e nella risposta diffuse field acquisita con lo smartphone attaccato al corpo si può ipotizzare che tale notch sia dovuto a delle caratteristiche del soggetto. Uno studio più approfondito potrebbe dar conferma di questa ipotesi o comunque definire l'identità di tale notch.

Infine si conclude affermando che la procedura descritta in questo capitolo migliora le acquisizioni effettuate da SelfEar in ambiente non controllato ampliando le possibilità dell'applicazione, tuttavia non permette di acquisire delle PRTF direttamente utilizzabili. I risultati ottenuti sono comunque incoraggianti e suggeriscono un miglioramento dell'acquisizione di PRTF mirate all'uso esclusivo all'interno di una modellazione strutturale mista.

# Capitolo 7

## Trasparentizzazione headset

Come è stato precedentemente esposto nel paragrafo 3.1.1, un mixer AAR è composto da un filtro compensativo che ha lo scopo di ripristinare le caratteristiche di un canale uditivo occluso da un headset alla condizione in cui esso invece è aperto. Avendo a disposizione un microfono miniaturizzato si può, utente per utente, misurare le condizioni a canale aperto e utilizzare le informazioni ottenute dal confronto con la misurazione a canale chiuso per ottenere il filtro compensativo, come viene effettuato in [4]. Non avendo a disposizione tale strumentazione diventa necessario trovare soluzioni diverse e meno invasive. Tale capitolo esplora le possibilità di utilizzo di una procedura di auto-calibrazione che si propone di essere accessibile per un utente medio nella definizione di un filtro compensativo personale, applicabile in campo mobile. Sono state effettuate delle sessioni sperimentali psico-acustiche su 11 soggetti focalizzandosi su di un numero limitato di parametri selezionabili dall'utente per la definizione di un filtro compensativo direttamente utilizzabile in contesto mobile AAR.

### 7.1 Hardware

**Loudspeaker** Il loudspeaker utilizzato è la cassa attiva di marca Energy modello 8EP200PWD avente le seguenti caratteristiche:

range frequenze (-3dB)	65Hz - 18kHz
range frequenze (-10dB)	55Hz - 18kHz
Max SPL	110 dB
potenza di uscita	150W (bassi) + 50W (tweeter)

**Tabella 7.1:** *Caratteristiche del loudspeaker Energy*

**Headset** L'headset utilizzato per le misurazioni è costituito dalle cuffiette Roland CS-10EM, le stesse descritte nella sessione 5.1

**Scheda audio** La scheda audio utilizzata è la Fast Track Pro della M-audio già descritta nella sessione 6.1

## 7.2 Elaborazione del segnale

### 7.2.1 Segnali

Il segnale utilizzato in questa procedura è uno sweep logaritmico che copre uniformemente in 3s tutte le frequenze tra 20Hz e 20kHz, cioè tutte le frequenze udibili dall'uomo. Durante la procedura tale segnale viene opportunamente filtrato per renderlo muto al di fuori di determinate bande, si ottengono quindi tre nuovi segnali di sweep in cui la durata è la stessa dello sweep originale, ma dove sono stati inseriti dei silenzi. Le tre bande in cui viene diviso sono [0, 2500], [2500, 6000], [6000, 20000]. Questa operazione è stata effettuata per impedire che ogni filtro applicato non andasse ad influire su frequenze esterne al range di interesse di applicazione. Come discusso in 3.2.2 si ha che in un canale aperto la prima risonanza si trova a circa 3 kHz, mentre in un canale chiuso si sposta a circa 6kHz. Alla risposta a canale chiuso va quindi applicato un filtro nella banda [2500, 6000] Hz per poter ripristinare la risonanza e un filtro nella banda [6000, 20000] Hz per eliminare la risonanza a canale chiuso.

### 7.2.2 Pure Data

Pure Data è un ambiente di programmazione grafica in tempo reale per processare audio e video. L'interfaccia con cui il musicista-programmatore parla con PD è grafica, quindi non c'è la necessità di scrivere codice in un editor di testo, ma si realizzano delle patch combinando fra loro vari tipi di oggetti grafici, nonostante ciò si tratta comunque di un linguaggio di programmazione perché consente di realizzare algoritmi più o meno complessi come tutti gli altri linguaggi. Nel gergo dei sintetizzatori analogici una patch rappresentava l'insieme dei collegamenti fra i suoi moduli. PD mutua questo concetto: attraverso una patch si definisce graficamente l'ordine con cui i vari oggetti sono collegati fra loro. Gli algoritmi vengono creati selezionando una serie di entità grafiche all'interno di una finestra, detta patch window. Il funzionamento di PD avviene in tempo reale, quindi gli algoritmi sono interattivi e i parametri possono essere modificati durante l'esecuzione. È anche possibile cambiare la struttura stessa di tali algoritmi mentre sono attivi, aggiungendo o rimuovendo moduli in modo semplice e intuitivo. Nasce con la funzione di creare applicazioni audio e, da qualche tempo, anche video. Pure Data è un software libero, i suoi sorgenti possono essere scaricati, studiati, modificati e redistribuiti da chiunque. Inoltre Pure Data è multiplatforma, quindi gira sui sistemi operativi

più comuni: GNU/Linux, Microsoft Windows, Apple Mac OS X, FreeBSD. La sua nascita risale al 1996 ideato e scritto da Miller Puckette, lo stesso che a metà degli anni '80 aveva sviluppato Max, in seguito divenuto un software commerciale. PD riprende le idee e i concetti di quest'ultimo, pur basandosi su una filosofia più libera e aperta che ne fa un sistema più dinamico e dalle prospettive future più interessanti.

### 7.2.3 Filtri e implementazione

I filtri utilizzati, come riportato nella sezione 3.1.1, sono: un filtro passa alto, un filtro peak e un filtro notch. Di seguito viene descritto come realizzarli in PD.

**Passa alto** - L'oggetto per realizzare un filtro passa alto è lo *[hip]*. Esso possiede una inlet (punto di input) per il segnale ed una inlet per la frequenza di taglio. Le frequenze sotto la frequenza di taglio vengono attenuate, essendo un filtro del primo ordine, di  $20dB/dec$  [44].

**Filtro peak** - Per realizzare il filtro peak del secondo ordine si sono usati in cascata due filtri *[vcf]* (voltage controlled filter) del primo ordine. Il filtro vcf è un filtro passa banda la cui frequenza centrale può cambiare continuamente nel tempo. Possiede tre inlet: la prima prende in ingresso il segnale, la seconda prende in ingresso la frequenza centrale e l'ultima prende in ingresso il parametro Q [44].

**filtro notch** - Per realizzare il filtro notch è stato riusato l'oggetto *[vcf]*. Per ottenere il notch è bastato operare una differenza tra il segnale ed il segnale filtrato tramite un filtro peak (realizzato come discusso precedentemente). I parametri in ingresso risultano quindi uguali a quelli del filtro peak.

## 7.3 Protocollo di comunicazione: OSC

OpenSound Control (OSC) è stato creato dal CNMAT (Berkeley University) nel 1997 ed è un protocollo open source per la comunicazione tra computer, sintetizzatori audio e altri dispositivi multimediali.

OSC è progettato per supportare un'architettura client/server e i dati, che vengono trasmessi in unità, sono denominati pacchetti. OSC è un protocollo di alto livello: non si occupa cioè del meccanismo di trasmissione dei dati. La sola rete che trasporta un pacchetto OSC è responsabile della consegna sia del contenuto che del dato riguardante la sua grandezza. Un pacchetto OSC può essere rappresentato in un datagramma come un protocollo di rete UDP. Tutto ciò che manda (una macchina, un'applicazione

ecc.) messaggi OSC (sotto forma di pacchetti) è un client mentre tutto ciò che riceve (una macchina, un'applicazione ecc.) messaggi OSC (sotto forma di pacchetti) è un server. L'OSC è stato un'evoluzione diretta del MIDI (Musical Instrument Digital Interface), ovvero il protocollo standard, utilizzato fin dagli anni '80, per l'interazione degli strumenti musicali elettronici. La pressante domanda di riduzione dei costi, maggiore affidabilità e maggiore accessibilità a tutti gli utenti ha portato OSC a svilupparsi a tal punto da eguagliare e poi superare l'uso del MIDI. L'OSC, infatti, presenta molti vantaggi rispetto al MIDI tra i quali una maggiore velocità dei dati di trasmissione e una risoluzione notevolmente superiore rispetto agli 8 bit concessi da MIDI.

Nel nostro caso, Matlab viene definito come client e Patch 1 di Pure Data come server. Qui di seguito una parte di codice per l'invio di un messaggio OSC da parte di Matlab.

```
socket = pnet('udpsocket', 3455);
pnet(socket, 'udpconnect', '127.0.0.1', 3456);
pd_packet = [single(1), single(F0HP), ...
            single(volume), single(delay), single(recOn)];
[~, datastring] = cstruct(pd_packet);
[~, zerostring] = cstruct(single(0));
string = char(['/pd_packet' zerostring(1:2) ',ffffffff' ...
            zerostring(1:2) datastring]);
pnet(socket, 'write', string);
pnet(socket, 'writepacket');
pnet('closeall');
```

Gli elementi fondamentali sono la funzione `pnet` e la funzione `cstruct` che si occupano rispettivamente della connessione e della creazione del pacchetto.

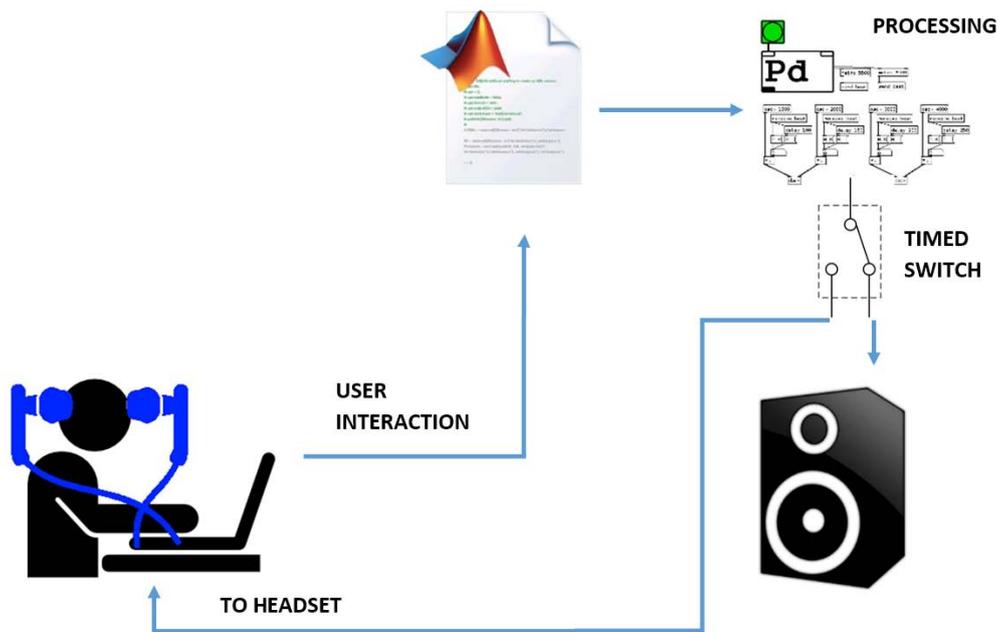
Nella Figura 7.8 si vede invece, contornato dal colore nero, il blocco relativo alla ricezione del messaggio OSC da parte di Pure Data.

## 7.4 Procedura guidata

La procedura guidata è composta da due patch scritte in Pure Data e tante schermate di una GUI generale grafica sviluppata in Matlab che guidano l'utente nella scelta di determinati parametri per la definizione dei tre filtri per la trasparentizzazione dell'headset.

La metodologia adottata per il controllo adattivo dell'utente consiste nel far percepire all'utente le differenze tra

- riproduzione da loudspeaker dello sweep originale



**Figura 7.1:** Collegamenti tra i vari componenti della procedura

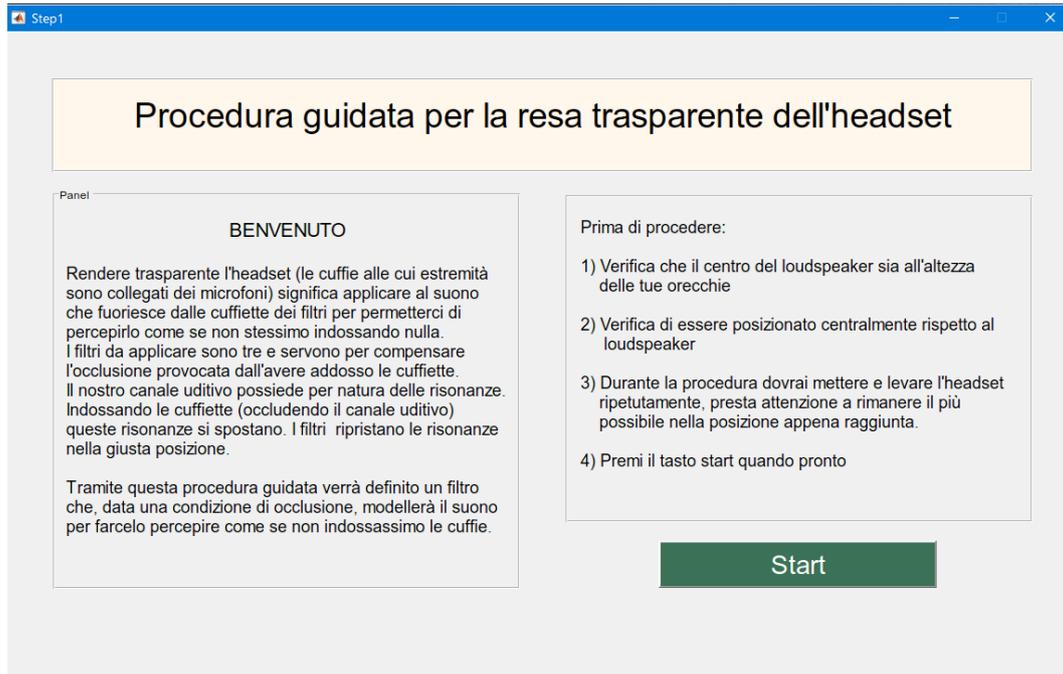
- riproduzione tramite headset della registrazione dello sweep (ottenuta in fase di setup dai microfoni dell'headset stesso indossato dal soggetto) a cui viene applicato il filtro che l'utente sta definendo

in modo da poter scegliere i parametri per cercare di rendere i due suoni il più identici possibile.

I parametri impostabili dall'utente sono concordi con i parametri che i blocchi di Pure Data relativi ad ogni filtro prendono in input, e sono i seguenti:

- Filtro passa alto: Frequenza di taglio;
- Filtro peak: Frequenza centrale, banda;
- Filtro notch: Frequenza centrale, banda;

Il principio di funzionamento segue lo schema in Figura 7.1: l'utente interagisce con una GUI di matlab all'interno della quale gli è possibile modificare tutti i parametri del filtro che vuole realizzare. Avviando la riproduzione (sempre dalla stessa view) Matlab comunica a Pure Data via protocollo OSC i dati inseriti dall'utente, Pure Data si occupa dunque della realizzazione del filtro e della riproduzione alternata tra loudspeaker e headset. L'utente può continuare a modificare i parametri del filtro in riproduzione, che in real-time modificano la riproduzione stessa permettendo all'utente di percepire istantaneamente le modifiche apportate. La riproduzione avviene nel seguente modo: il segnale viene riprodotto prima dal loudspeaker e poi, con un delay costante intercalato tra i due, dall'headset, ciclicamente.



**Figura 7.2:** *View 1*

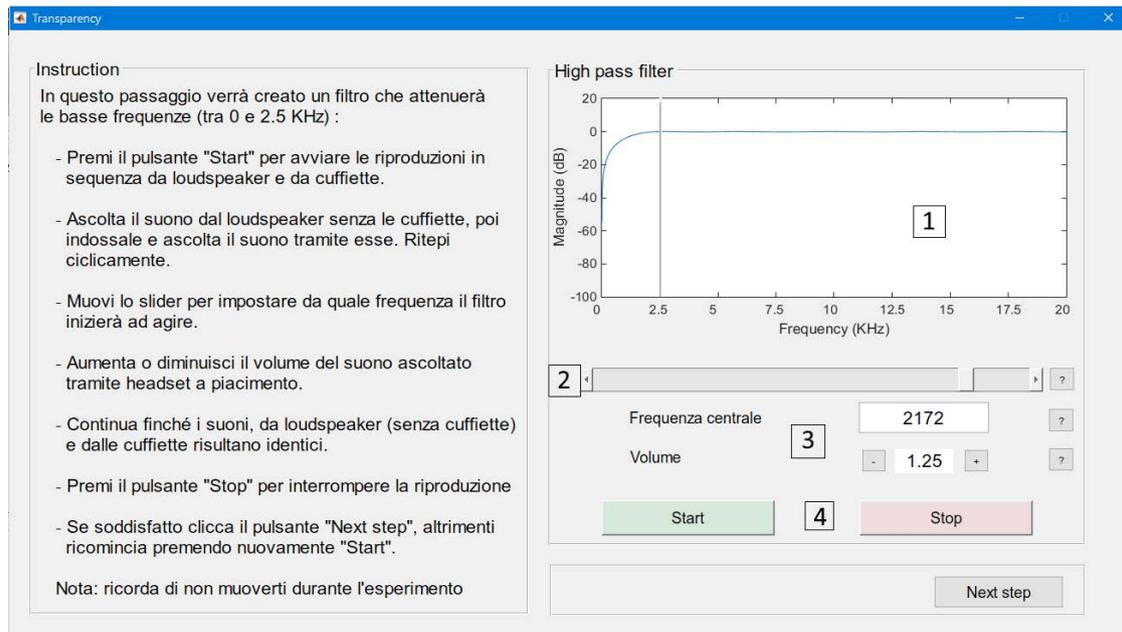
### 7.4.1 GUI di Matlab

Una prima realizzazione della procedura prevedeva una sola view in cui si potevano modificare tutti i parametri dei tre filtri contemporaneamente (vedi Figura ), giudicata troppo complessa per un utente standard si è deciso di realizzare una seconda versione in cui i filtri venissero definiti in schermate separate ed indipendenti. Si sono quindi realizzate cinque schermate:

- view 1: finestra che introduce l'utente alla procedura, da le indicazioni di base per la buona riuscita della procedura.
- view 2: finestra per la definizione del filtro passa alto.
- view 3: finestra per la definizione del filtro peak.
- view 4: finestra per la definizione del filtro notch.
- view 5: finestra di riepilogo per l'ascolto completo di quanto realizzato.

Lo sweep scomposto in bande, una volta terminata la procedura, viene ricomposto e nella view 5 viene fatto ascoltare all'utente il segnale completo compensato.

Un ultimo parametro modificabile dall'utente per ogni banda è il gain del segnale in cuffia.



**Figura 7.3:** View 2

**View 1** La prima view che l'utente visualizza è quella in Figura 7.2. Si osserva subito che tale view semplicemente descrive all'utente ciò che si sta per fare e lo informa rispetto alla postura e posizione da mantenere durante l'esecuzione della procedura guidata. Cliccando il pulsante verde "Start" si dà inizio alla procedura visualizzando view 2.

**View 2** La seconda view è quella relativa al filtro passa alto, visibile in Figura 7.3. Nella parte di sinistra vengono indicate le istruzioni per svolgere la procedura, nella parte di destra invece si identificano una serie di elementi suddivisi in due pannelli. Nel primo pannello, andando dall'alto verso il basso si incontrano:

1. un grafico che mostra la risposta in frequenza del filtro che si sta definendo. Essa è realizzata mediante lo script `HPfilter.m` (visibile in appendice B sezione B.4.1) dove viene usata la funzione `fdesign` di Matlab, avendo avuto cura di fissare tutti i parametri per la progettazione del filtro ad eccezione della frequenza di taglio che viene impostata dall'utente. Tale risposta non corrisponde esattamente a quella che viene realizzata da Pure Data e applicata allo sweep in riproduzione (questo è dovuto al fatto che Matlab permette una progettazione più accurata mettendo a disposizione molti più parametri per definire il filtro), è però sufficiente per dare una rappresentazione dell'azione del filtro che si sta realizzando, obiettivo che tale grafico si propone. Sul grafico è inserita una barra a 2500 Hz che divide il range di azione del filtro (a sinistra della barra) dal range in cui invece non comporta modifiche.

2. uno slider movibile dall'utente. Il valore minimo impostabile è 0, mentre il valore massimo è 2500. Il passo è di 10 se premuto nei pulsanti alle estremità, 100 se premuto nel range di movimento dello slider. Il valore che l'utente seleziona corrisponde alla frequenza di taglio del filtro passa alto ed è visibile nella textbox sotto lo slider, all'interno della quale può essere direttamente scritto il valore che si intende far assumere alla frequenza di riferimento, ovviamente sono inseriti dei controlli tali per cui non si possano sfiorare i limiti del range associato al filtro.
3. due tasti per aumentare o diminuire il volume del suono ascoltato in cuffia. Quando si accede alla view il volume è preimpostato al valore 1, tale valore viene moltiplicato in Pure Data per il segnale da riprodurre in cuffia (quindi il volume ad 1 lascia inalterato il segnale, a 2 raddoppia l'intensità del segnale). Il volume minimo è 0, il volume massimo è 5 ed il passo è di 0.25.
4. i pulsanti start e stop che avviano e stoppano rispettivamente la riproduzione audio alternata dei segnali su loudspeaker e headset.

Vicino ad ognuno degli elementi sopra descritti si trova un pulsante con un punto di domanda, premendo tale pulsante si apre una finestra che spiega a cosa serve l'elemento a cui è riferito.

Nel secondo blocco si incontra solo un pulsante "Next" che quando cliccato svolge in ordine le seguenti operazioni:

- salva i parametri del filtro su di un file .mat;
- comunica a Pure Data che è terminata l'acquisizione del filtro;
- richiama la view successiva.

Pure Data, una volta ricevuto il messaggio inviatogli da Matlab, salva su disco un file .wav contenente lo sweep (con la sola banda di interesse) compensato con il filtro realizzato.

**View 3** La terza view è quella relativa al filtro peak, ed è molto simile alla precedente nella sua struttura, le differenze sono le seguenti:

1. Il grafico viene realizzato mediante lo script Pfilter.m (visibile in appendice B sezione B.4.2). La banda in cui il filtro andrà ad agire viene identificata da due delimitatori visibili sul grafico.
2. Gli slider sono due, uno relativo alla frequenza centrale, l'altro relativo alla banda. Il primo ha un range di [2500, 6000] con passo di 10 e/o 100 a seconda di dove si

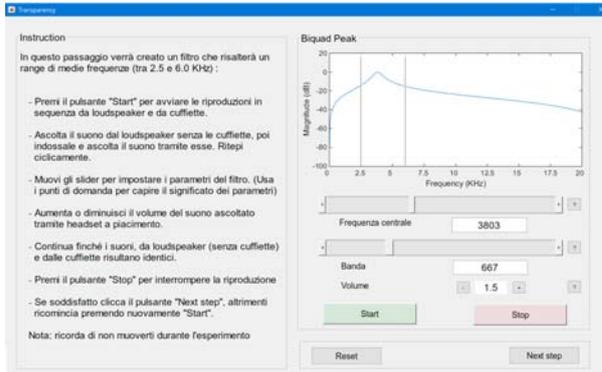


Figura 7.4: View 3

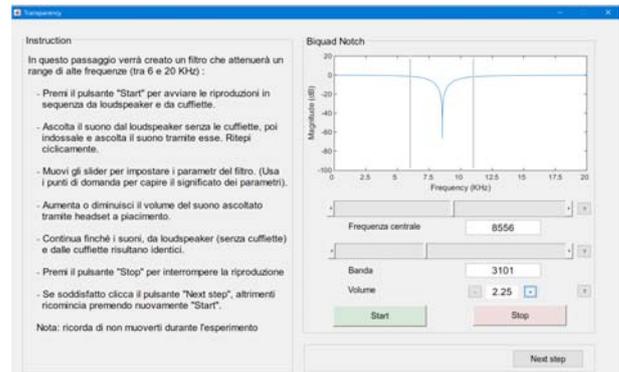


Figura 7.5: View 4

clicca per aumentare/diminuire il valore. Il secondo ha un range di  $[0, 2500]$  con stesso passo del precedente.

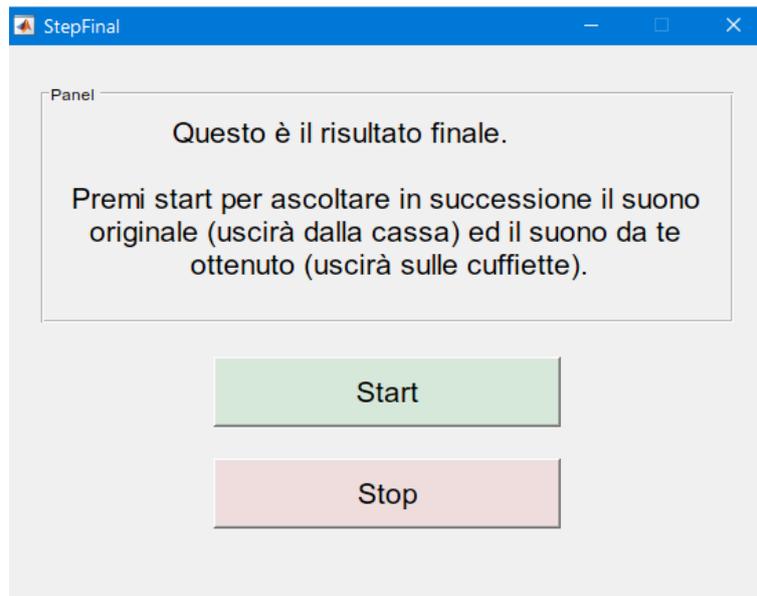
Una nota importante va spesa nella comunicazione tra Matlab e Pure Data, in questo caso infatti, come già visto, si ha che il blocco per la realizzazione del filtro non prende in input il parametro banda ma il parametro  $Q$ , viene quindi realizzata la conversione  $Q = Fc/B$  dove  $Fc$  è la frequenza centrale e  $B$  è la banda.

**View 4** La quarta view è quella relativa al filtro notch, ed è molto simile alla seconda view nella sua struttura, le differenze sono le seguenti:

1. Il grafico viene realizzato mediante lo script `Nfilter.m` (visibile in appendice B sezione B.4.3). La banda in cui il filtro andrà ad agire viene identificata da un delimitatore visibile sul grafico (a sinistra di esso il filtro agisce, a destra no).
2. Gli slider sono due, uno relativo alla frequenza centrale, l'altro relativo alla banda. Il primo ha un range di  $[6000, 20000]$  con passo di 10 e/o 100 a seconda di dove si clicca per aumentare/diminuire il valore. Il secondo ha un range di  $[0, 8000]$  con stesso passo del precedente.

Come nel caso precedente, anche in questo il blocco per la realizzazione del filtro non prende in input il parametro banda ma il parametro  $Q$ , viene quindi realizzata la conversione  $Q = Fc/B$  dove  $Fc$  è la frequenza centrale e  $B$  è la banda.

**View 5** La quinta view termina la procedura guidata. Presenta due pulsanti "start" e "stop" che rispettivamente avviano e fermano la riproduzione acustica tramite l'invio di un segnale OSC a Pure Data che, leggendo i file .wav relativi ai blocchi per bande dello sweep salvati dalle varie view, ne effettua un overlap and add nel tempo ottenendo un unico file che riproduce tramite headset alternatamente con lo sweep originale riprodotto tramite loudspeaker.



**Figura 7.6:** View 5

## 7.4.2 Patch di Pure Data

Le patch di Pure data sono due: *osc\_receiver\_reproduction\_original.pd*, dove Figura 7.9 contiene la parte relativa al solo filtro peak, e *osc\_receiver\_reproduction\_filtered.pd* dove Figura 7.8 contiene, in corrispondenza all'immagine precedente, la parte relativa al solo filtro peak. Sono state realizzate due patch in modo da poter scegliere il dispositivo di uscita per ognuna delle due patch. Una di essa infatti ha il compito di riprodurre il suono tramite cuffie, l'altra tramite loudspeaker.

**patch 1: *osc\_receiver\_reproduction\_original.pd*** - Questa patch si occupa della riproduzione tramite loudspeaker della banda dello sweep relativo al filtro che si vuole applicare. Si evidenziano tre blocchi procedurali (in Figura 7.8 contraddistinti dai colori nero, rosso, e verde).

1. Il primo blocco (evidenziato dal numero 1 e contornato di nero), si occupa della ricezione del messaggio OSG inviato da Matlab contenente i parametri del filtro e alcuni segnali di controllo, uno dei quali avvia o stoppa la procedura della patch.
2. Il secondo blocco in ordine di esecuzione (evidenziato dal numero 2 e contornato di rosso) si occupa dell'apertura e della riproduzione del file sweep.
3. L'ultimo blocco (evidenziato dal numero 3 e contornato di giallo) ha il compito di inviare un messaggio a Patch 2 per inviargli i parametri del filtro ricevuti da matlab e altri segnali di controllo tra cui un segnale per la sincronizzazione.

**patch 2: osc\_receiver\_reproduction\_filtered.pd** - Questa patch si occupa della riproduzione in cuffia della banda dello sweep relativo al filtro che si sta applicando, filtrato con il filtro stesso. Si evidenziano cinque blocchi procedurali (in Figura 7.9 contraddistinti dai colori nero, rosso, verde, giallo e blu).

1. Il primo blocco (evidenziato dal numero 1 e contornato di nero) si occupa della ricezione di un messaggio inviato dalla patch 1 contenente i parametri del filtro e alcuni segnali di controllo, uno dei quali avvia o stoppa la procedura della patch.
2. Il secondo blocco in ordine di esecuzione (evidenziato dal numero 2 e contornato di rosso) si occupa dell'apertura del file sweep da riprodurre.
3. Segue il terzo blocco (evidenziato dal numero 3 e contornato di giallo) che attende la riproduzione dello sweep e poi si occupa di inviare un messaggio alla patch 1 per farla partire.
4. Mentre il terzo blocco è in attesa viene eseguito il quarto blocco (evidenziato dal numero 4 e contornato di verde) che si occupa della realizzazione del filtro secondo i parametri ed il volume impostati dall'utente.
5. L'ultimo blocco (evidenziato dal numero 5 e contornato di blu) viene eseguito solo quando l'utente clicca sul pulsante "Next" della view di Matlab, qui viene registrato e salvato su disco lo sweep filtrato.

## 7.5 Sessioni sperimentali

Le sessioni sperimentali condotte sono state due: la prima è stata condotta come test pilota, la seconda invece produce dei dati rilevanti che verranno analizzati nella sezione "Risultati". Le principali differenze tra la prima sessione e la seconda sono:

- *Il segnale di test*: nella prima sessione è l'intero sweep, mentre nella seconda è lo sweep con banda diversa per ogni view.
- *Il volume*: nella prima sessione il volume influisce solo sul suono ascoltato senza alterare il segnale, nella seconda invece il volume fornisce un gain al segnale alterandolo in ampiezza sempre per bande.

### 7.5.1 Prima sessione sperimentale

**Preparazione soggetto** - Nella fase di setup viene chiesto all'utente di sedersi di fronte al loudspeaker ad un metro di distanza da esso, regolando lo sgabello in modo che, da

seduto, l'ingresso del canale uditivo ed il centro del cono del loudspeaker siano alla stessa altezza. Vengono fatti indossare i microfoni miniaturizzati Soundman e viene effettuata la registrazione di un segnale di sweep. Successivamente viene fatto indossare l'headset al soggetto, si procede in ultima con la registrazione del segnale di sweep che diventa il segnale di test per il soggetto.

**Esecuzione della procedura guidata** - Inalterando la posizione del soggetto raggiunta nella fase di setup viene lanciata la procedura guidata. Il computer sulla quale essa viene visualizzata si trova su di un tavolino posto sotto il loudspeaker, in tal modo non vi sono influenze col suono riprodotto. Per le view 2, 3 e 4 viene chiesto al soggetto di scegliere i valori dei parametri del filtro da realizzare relativo alla view. La scelta viene fatta per tentativi: il soggetto leva l'headset e clicca sul pulsante "start" della view, quindi ascolta il suono riprodotto dal loudspeaker, inserisce poi l'headset ed ascolta il suono riprodotto da questi. L'obiettivo del soggetto è rendere quanto più possibile simili i due suoni ascoltati, per fare ciò può agire sugli slider che in tempo reale modificano i parametri del filtro applicato al suono riprodotto tramite headset, e sul volume del suono sempre riprodotto tramite headset. Quando il soggetto reputa che i due suoni siano simili, premendo il tasto "next" procede alla view successiva. Una volta raggiunta l'ultima view del filtro notch, verrà proposto lo sweep completo dal loudspeaker e successivamente lo sweep registrato con l'headset e con l'applicazione dei filtri risultanti dalla procedura di autocalibrazione.

## 7.5.2 Seconda sessione sperimentale

**Setup** - Il Setup di questa sessione è lo stesso effettuato per la prima sessione, con l'aggiunta che il segnale di test viene elaborato mediante lo script *preProcessing.m* (visibile in appendice B sezione B.2) che divide il segnale in tre segnali, ognuno dei quali filtrato con un filtro Butterworth dell'ottavo ordine per lasciare inalterata la sola banda di interesse.

**Esecuzione della procedura guidata** - Questa esecuzione è la stessa usata per la prima sessione sperimentale: Viene lasciata inalterata la posizione del soggetto raggiunta nella fase di setup e viene lanciata la procedura guidata. Per le view 2, 3 e 4 viene chiesto al soggetto di scegliere i valori dei parametri del filtro da realizzare relativo alla view sempre per tentativi. Quando il soggetto reputa che i due suoni siano simili, premendo il tasto "next" procede alla view successiva. Una volta raggiunta l'ultima view del filtro notch, verrà proposto lo sweep completo dal loudspeaker e successivamente lo sweep registrato con l'headset e con l'applicazione dei filtri risultanti dalla procedura

ID	Sessione	Sesso	Età
1	1	F	25
2	1	M	16
3	1	M	18
4	1	M	21
5	1	M	31
6	2	M	26
7	2	M	25
8	2	M	26
9	2	M	26
10	2	F	25
11	2	M	26

**Tabella 7.2:** *Soggetti*

di autocalibrazione. Infine, per ottenere il segnale completo a partire dalle tre bande si lancia lo script *postProcessing.m* (visibile in appendice B sezione B.3).

Tutte e due le procedure hanno una durata di circa 40 minuti ciascuna.

## 7.6 Soggetti

I soggetti testati sono di sesso sia femminile che maschile. La loro età varia tra i 16 e i 31 anni. In Tabella 7.2 vi è la lista dei soggetti che hanno partecipato alle due sessioni sperimentali, nella seconda colonna si identifica con “1” chi ha partecipato alla prima sessione, con “2” chi invece ha partecipato alla seconda. Restano esclusi tre soggetti che sono hanno contribuito con le loro misurazioni ad allestire le procedure di setup e di esecuzione della procedura guidata, i cui risultati sono incompleti e/o non comparabili con gli altri risultati ottenuti.

## 7.7 Risultati

Mentre la sessione pilota ha messo in luce alcune criticità riguardanti il protocollo e gli stimoli sonori ad ogni fase della procedura, la seconda sessione sperimentale ha portato a dei risultati rilevanti. Più precisamente si vogliono osservare i parametri dei filtri scelti dai soggetti e si vuole confrontare il suono filtrato mediante i filtri realizzati con il suono registrato dai SoundMan all’interno del canale uditivo (in condizione aperto). In Tabella 7.3 si osservano i parametri dei filtri scelti dai soggetti.

Si denota una variabilità molto elevata per quanto riguarda la frequenza di taglio del filtro passa basso. Il range in cui invece viene impostato il picco è più limitato, non scende sotto i 2.9kHz e non supera i 5.2. La banda, eccetto due outlier è sempre sopra

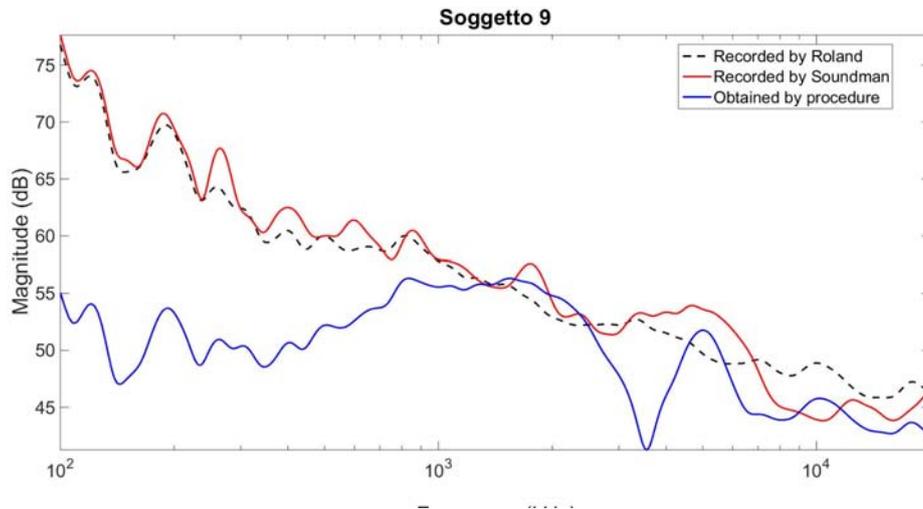
	Sogg. 6	Sogg. 7	Sogg. 8	Sogg. 9	Sogg. 10	Sogg. 11
Passa alto F0	747	1111	299	2102	1800	500
Passa alto Vol	2.25	1.5	0.5	1.75	2.25	1.5
Filtro peak F0	4791	2983	5814	5106	5199	3920
Filtro peak BW	266	1130	1520	762	1859	1265
Filtro peak Vol	1.25	2.75	1.5	1.75	2.25	1.25
Filtro notch F0	9597	8283	11000	8076	9201	8041
Filtro notch BW	2226	2436	3793	4791	3793	3892
Filtro notch Vol	1.5	3.5	5	2	2.75	1.75

**Tabella 7.3:** Risultati - parametri dei filtri

i 1000 Hz. Per quanto riguarda la frequenza centrale del notch, essa non scende sotto 8.0 kHz e non supera gli 11 kHz, la banda è invece compresa tra i 2.2 kHz e i 4.8 kHz. Si nota come venga abbastanza rispettata la teoria del tubo semiaperto affrontata nella sezione 3.2.2, la quale prevede che la frequenza centrale del picco raddoppi, e nel nostro caso si traduce nell'aver il notch collocato ad una frequenza centrale doppia di quella di dove viene collocato il picco. I volumi, seppur diversi tra i soggetti, rimangono abbastanza coerenti su singolo soggetto (eccetto che per il soggetto 8 che ha una variabilità massima).

Per quanto riguarda i risultati grafici, visibili in appendice A sezione A.2, si ha che per ogni grafico vengono visualizzate tre curve, la prima in nero tratteggiata rappresenta il suono registrato dall'headset in fase di setup; la seconda in rosso rappresenta il suono registrato dai microfoni miniaturizzati Soundman in fase di setup; la terza curva corrisponde alla prima a cui viene applicato il filtro realizzato mediante la procedura di auto-calibrazione. Quando le due curve, rossa e blu, diventano uguali, si ha raggiunto l'obiettivo target.

Per dare una caratterizzazione alla bontà di un grafico osserviamo come esempio i risultati del Soggetto 9, visibili in Figura 7.7. Egli ha fatto agire il filtro passa alto a circa 2 kHz, si vede infatti dal grafico come la curva blu sia attenuata prima dei 2 kHz, nonostante sia stato dato un guadagno positivo (maggiore di uno) al segnale (tramite il volume). Si osserva poi dalla curva rossa e nera che il picco dovrebbe trovarsi intorno ai 5 kHz, esattamente dove il soggetto 9 lo ha posizionato. Si denota però un notch a circa 3.5 kHz, si tratta di un artefatto introdotto dall'overlap dei segnali sweep filtrati per bande e riconducibile al fatto che al picco non è stata data una banda sufficientemente larga. Sempre dalle curve rossa e blu si evidenzia la posizione del notch a circa 9 kHz con una larghezza di banda di circa 4 kHz, il Soggetto 9 l'ha posizionato a 8 kHz con una larghezza di banda di poco meno di 5 kHz. Si denota poi come il volume sia stato



**Figura 7.7:** Risultati del Soggetto 9

impostato un poco basso nel picco e più corretto nel notch. Un ultimo dato rilevante deriva dal fatto che la maggior parte dei soggetti ha, a procedura conclusa, confermato il proprio lavoro di scelta dei parametri dicendo informalmente che i suoni ascoltati nell'ultima view erano simili. Non è stata condotta nessuna indagine approfondita sulla percezione di similarità tra i due segnali.

In Tabella 7.4 è mostrato il confronto tra le frequenze e le bande dei picchi e dei notch identificati dai soggetti, con quelli invece che si identificano dai grafici (dalla registrazione con i microfoni miniaturizzati Soundman). I grafici relativi a tutti i soggetti sono visibili in appendice A sezione A.2.

## 7.8 Conclusioni

I risultati ottenuti in questa sessione sperimentale non sono sufficienti per validare o meno la procedura guidata, ma favoriscono delle riflessioni riguardo ad alcuni accorgimenti, ed alcuni modi per proseguire ed approfondire questo tema di ricerca. Essendo esperimenti psicoacustici, essi si basano sulla percezione di ogni singolo soggetto, e questa ha variabilità soggettiva di difficile analisi senza un'estesa sessione sperimentale. Una prima considerazione sta nel fatto che alcuni soggetti sono riusciti a posizionare abbastanza correttamente il picco ed il notch, seppur con bande tendenzialmente non combacianti con quelle visibili dalla registrazione fatta con i Soundman. Un altro punto importante risiede nella percezione delle basse frequenze. Si evidenzia che quasi tutti i soggetti hanno un volume più basso (per via del volume e/o dell'azione del passa alto) di quello che dovrebbe essere, tale situazione è probabilmente associata ad un effetto dovuto alla conduzione ossea (vista nella sezione 3.2.3) che tende ad aumentare l'energia

		Filtro Peak		Filtro Notch	
		$F0$	$Bw$	$F0$	$Bw$
Sogg. 6	<i>Percepiti</i>	4791	266	9597	2226
	<i>Originali</i>	-	-	9300	2000
Sogg. 7	<i>Percepiti</i>	2983	1130	8283	2436
	<i>Originali</i>	2500	2000	9200	2500
Sogg 8	<i>Percepiti</i>	5814	1520	11000	3793
	<i>Originali</i>	4700	1000	8500	4000
Sogg 9	<i>Percepiti</i>	5106	762	8076	4791
	<i>Originali</i>	5000	3500	9000	4000
Sogg 10	<i>Percepiti</i>	5199	1859	9201	3793
	<i>Originali</i>	4900	2500	10300	1000
Sogg. 11	<i>Percepiti</i>	3920	1265	8041	3892
	<i>Originali</i>	-	-	-	-

**Tabella 7.4:** Confronto tra parametri trovati dai soggetti e approssimazioni dei parametri originali. Il trattino indica che non è stato possibile quantificare un dato valore.

delle basse frequenze. La terza considerazione si basa sul processo di overlap delle tre bande dello sweep estratte, nel nostro caso infatti si vede che lasciando piena libertà al soggetto si possono verificare degli artefatti (come visto accadere per un notch nel grafico del Soggetto 9). Una soluzione è quella di limitare le scelte del soggetto in modo da limitare il più possibile il formarsi di queste distorsioni, limitando ad esempio la scelta della banda inferiormente (ad esempio al valore 1000) si dovrebbe avere un allargamento del filtro peak sufficiente eliminazione dei notch alle bande laterali nella banda di applicazione. Un'altra soluzione è quella di studiare un nuovo modo per lavorare sulle bande dello sweep, applicando ad esempio i filtri in parallelo su una copia dello sweep originale, con conseguente ridefinizione del processo di overlap.

Infine la percezione soggettiva potrebbe avvalersi di una nuova sperimentazione della metodologia MUSHRA<sup>1</sup> (MUltiple Stimuli with Hidden Reference and Anchor) in cui si possono far ascoltare tramite headset diversi segnali di sweep al soggetto, uno dei quali è quello realizzato con la procedura, mentre gli altri contengono delle variazioni come ad esempio l'assenza del picco o del notch o del filtro passa alto. Il soggetto deve poi indicare un indice di similarità con lo sweep riprodotto dal loudspeaker. L'esperi-

<sup>1</sup> Si tratta di una metodologia per la valutazione soggettiva della qualità acustica, per valutare la qualità percepita dall'output di un algoritmo di compressione acustica senza la perdita di dati.

---

mento ha l'obiettivo di valutare la bontà del filtro compensativo realizzato dal soggetto e delle singole componenti definite nelle tre viste fasi.

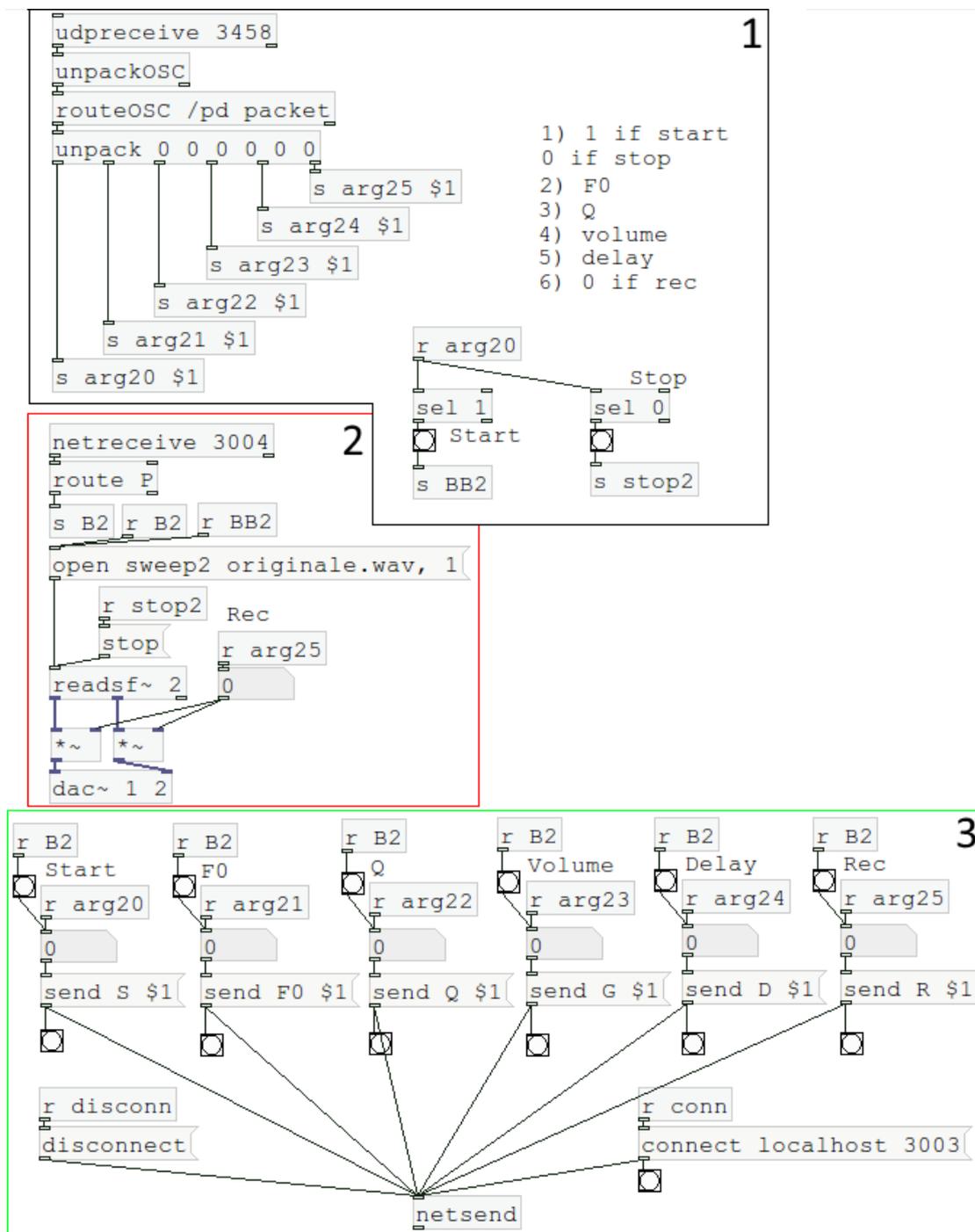


Figura 7.8: Patch 1

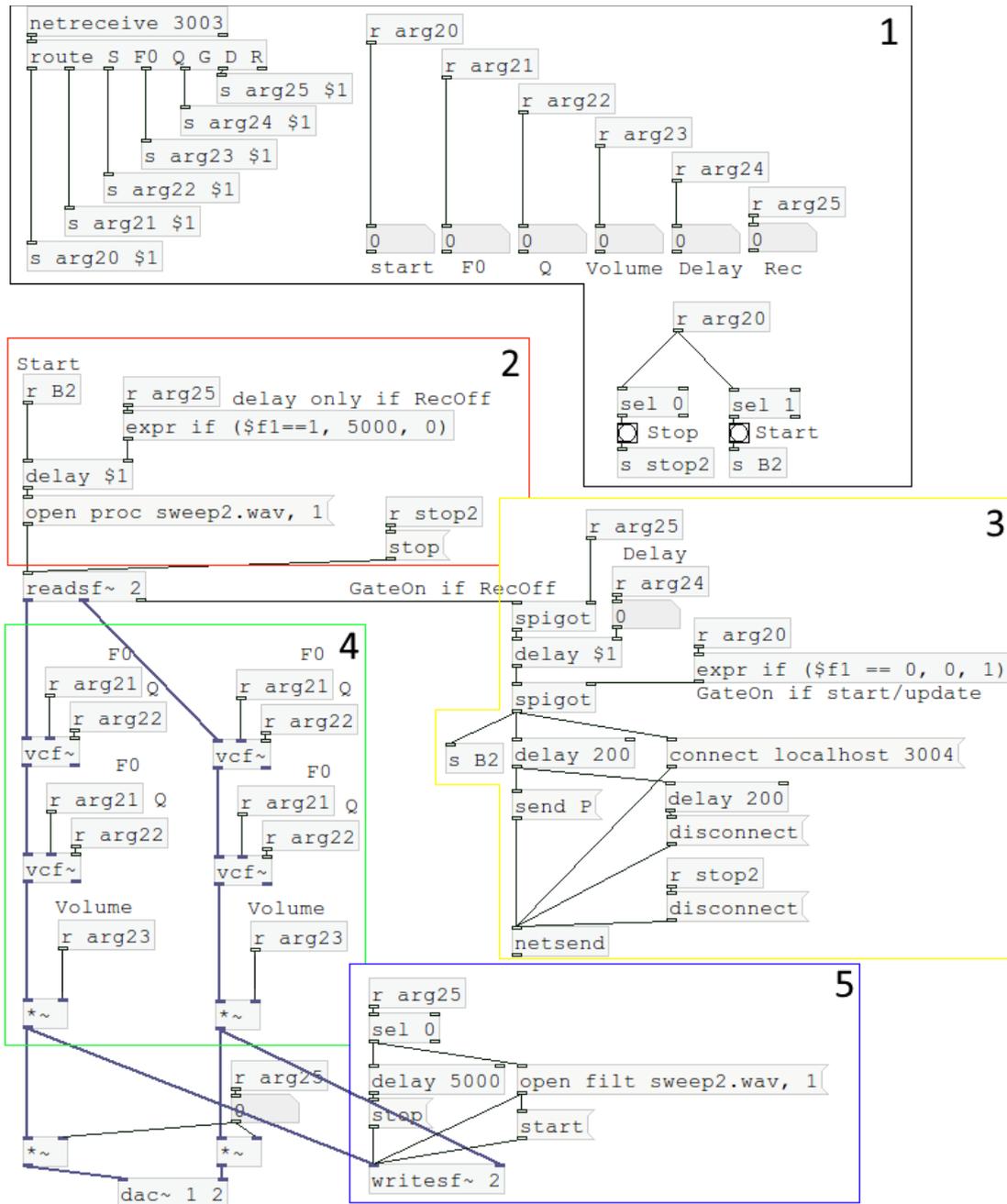


Figura 7.9: Patch 2



# Capitolo 8

## Conclusioni

In questa tesi sono stati affrontati tre aspetti legati al campo della realtà aumentata acustica. Partendo dal problema della personalizzazione dei parametri coinvolti nel rendering spaziale del suono, sono state esplorate delle soluzioni per l'acquisizione di tali parametri in modo semplice ed economico. L'applicazione SelfEar, sviluppata per dispositivi Android, è stata validata sperimentalmente in laboratorio conducendo ai seguenti risultati:

1. L'applicazione permette di individuare correttamente due dei cinque parametri fondamentali che caratterizzano l'acustica della pinna di un soggetto (la risonanza P1 ed il notch N1) [39, 40].
2. L'individuazione migliora nel caso le misurazioni vengano compensate con una risposta diffuse field.
3. L'applicazione non permette l'acquisizione di PRTF direttamente utilizzabili.

Con i parametri individuati da SelfEar è possibile ricercare all'interno di un database le PRTF o HRTF di un soggetto avente i parametri più simili possibili a quelli di ricerca. In tal modo è possibile inserire l'applicazione in una procedura strutturale mista che permette un tale meccanismo di selezione. Il Cap 6 espone i risultati di interazione dell'applicazione in vari ambienti e si occupa della progettazione e realizzazione di una procedura veloce e semplice per l'acquisizione di una risposta diffuse field da applicare successivamente alle registrazioni effettuate dall'applicazione in modo da compensare il contributo ambientale. Quest'ultima operazione si rivela indispensabile nel caso di ambienti diversi da quello controllato poiché permette di individuare anche uno solo dei cinque parametri rilevanti per l'acustica dell'orecchio.

Le acquisizioni che si ottengono mediante SelfEar spaziano nel piano mediano e permettono la definizione della PRTF. Per ottenere la HRTF dovrebbero essere possibili

acquisizioni anche lungo il piano orizzontale, un possibile sviluppo futuro dell'applicazione potrebbe quindi prevedere una ridefinizione della griglia di acquisizione permettendone al di fuori del solo piano mediano, con l'eventuale utilizzo di tecniche di head-pose estimation per dare robustezza alla procedura di acquisizione.

Nello schema di realizzazione di un sistema di realtà aumentata, una volta ottenute le caratteristiche soggettive di un utente necessarie alla modellazione del suono, sono state affrontate le problematiche legate all'uso di un headset. Questa tesi affronta quindi il problema della compensazione delle variazioni che il canale uditivo porta alla propagazione del suono nel caso in cui da aperto passi a chiuso. È stata definita una procedura di auto-regolazione per la definizione di un filtro compensativo guidando l'utente per passi ed è stata effettuata una prima validazione su un campione di 11 soggetti. Le sessioni sperimentali forniscono risultati incoraggianti che mostrano come gli utenti riescano nella media ad individuare caratteristiche simili a quelle misurate per quanto riguarda i parametri del filtro compensativo. I soggetti hanno poi espresso, anche se informalmente, un giudizio positivo sulla qualità dell'auto-regolazione che hanno effettuato dimostrando che soggettività, esperienza e risolutezza d'ascolto giocano un ruolo fondamentale in esperimenti, come questo, di psico-acustica. Ancora molti passi sono da compiere per giungere al risultato finale di questo progetto:

- validazione della possibilità di utilizzare i risultati acquisiti da SelfEar in ambienti non controllati con un modello strutturale misto come quello sopra descritto;
- Va condotta una statistica più consistente per la procedura di auto-regolazione

Infine per concludere lo schema di realizzazione di un sistema di realtà aumentata è necessaria una procedura che permetta di acquisire informazioni sull'ambiente in cui viene eseguita SelfEar. In tal modo si acquisirebbero i parametri necessari sia per spazializzare il suono, sia per inserire sorgenti sonore nella scena acustica in cui si è immersi.

# **Appendici**

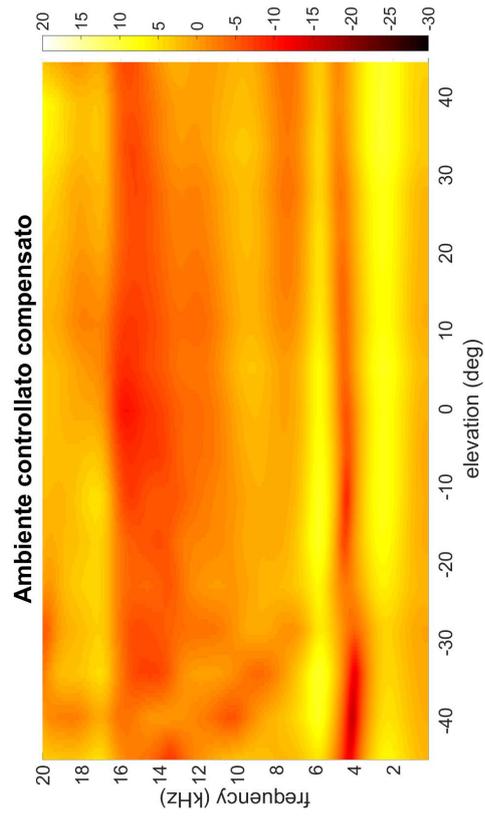
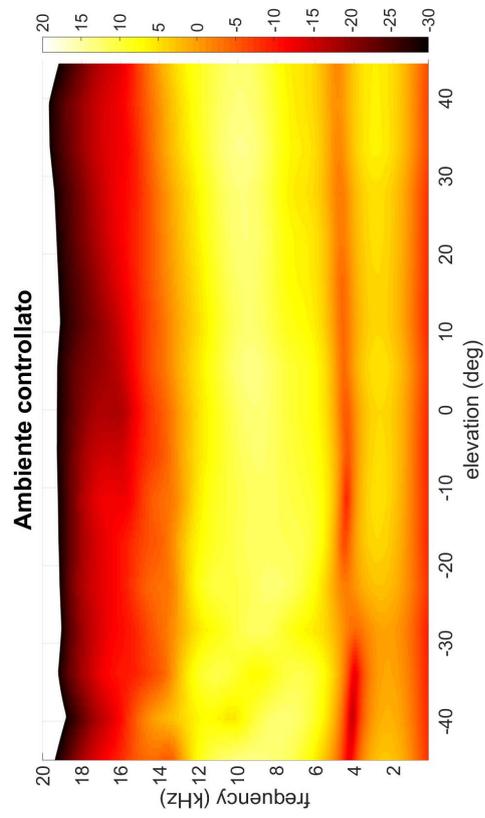
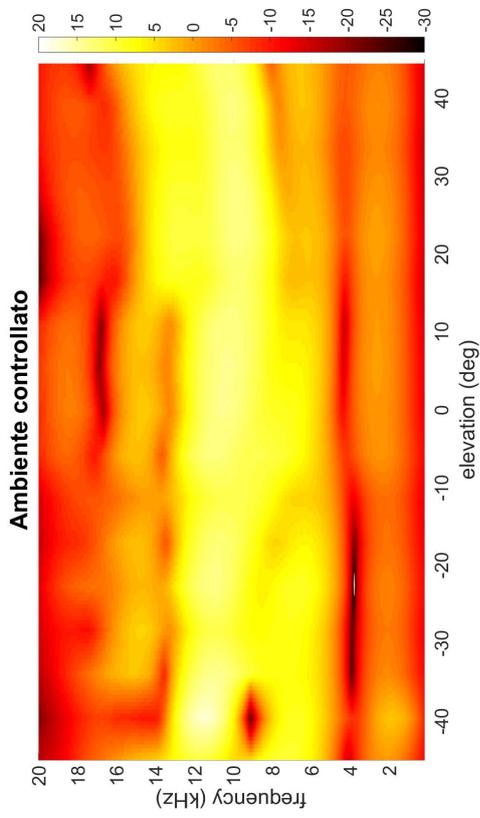
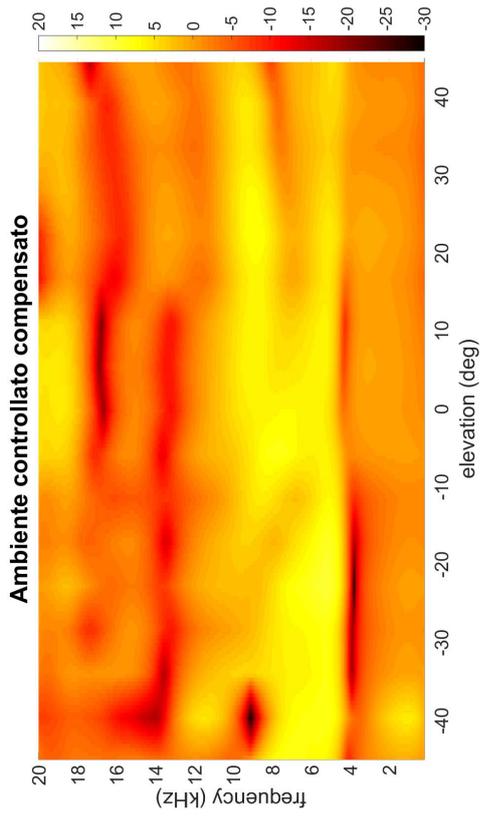


# **Appendice A**

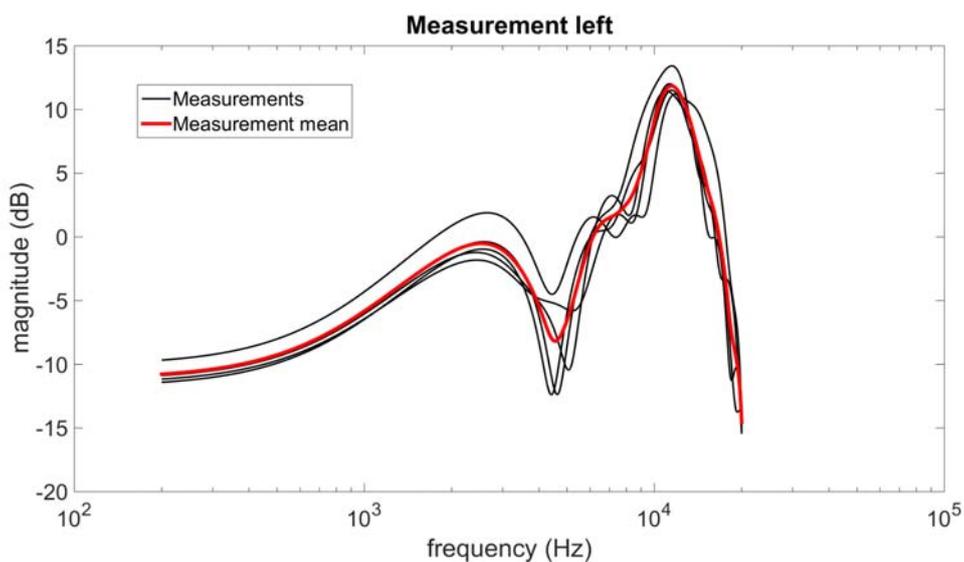
## **Risultati**

### **A.1 Risultati Cap 6**

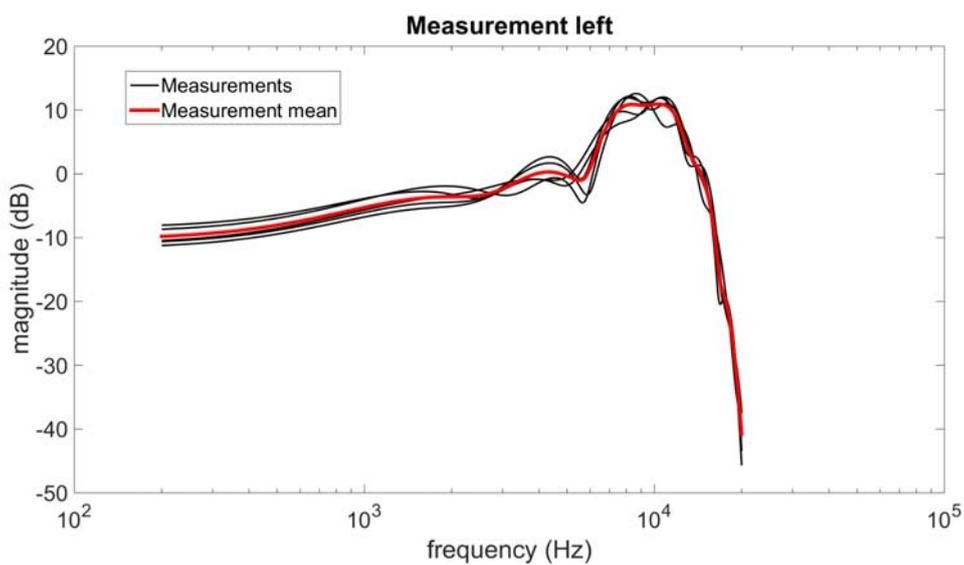




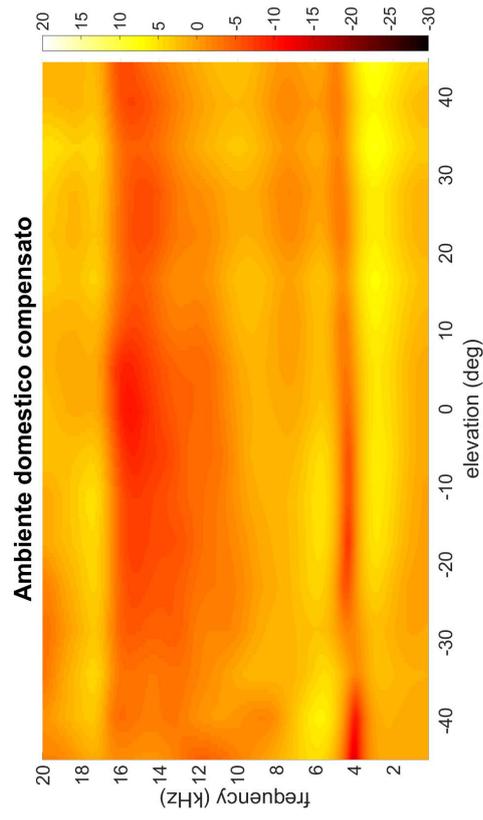
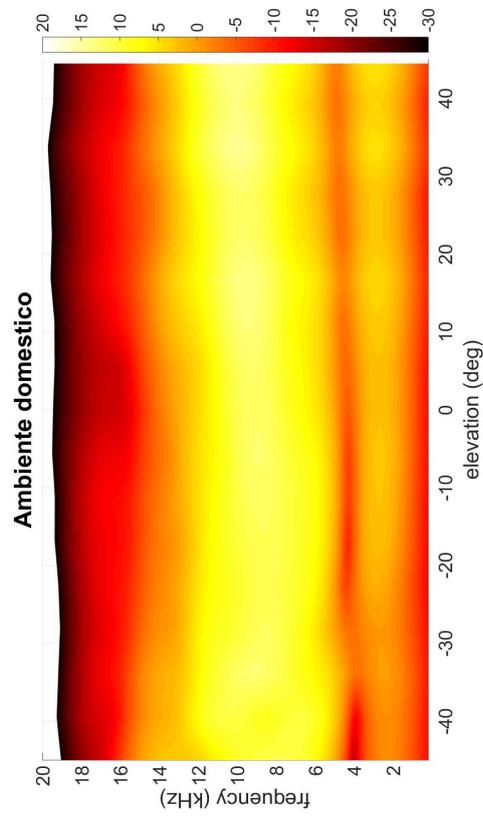
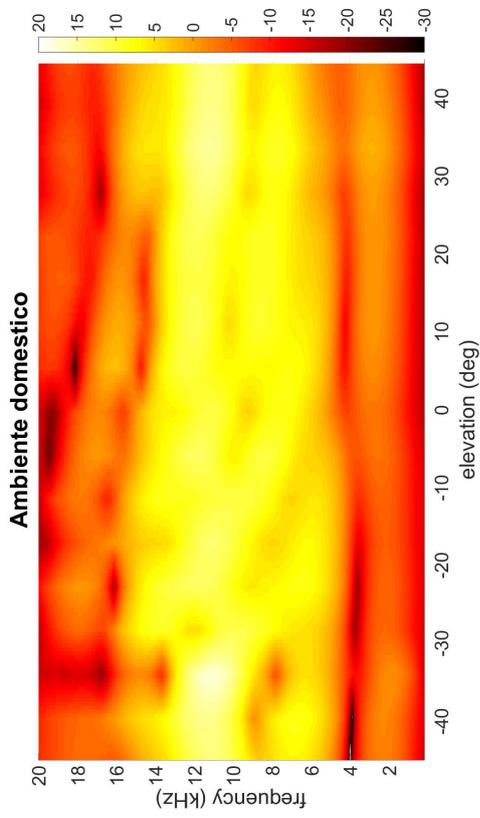
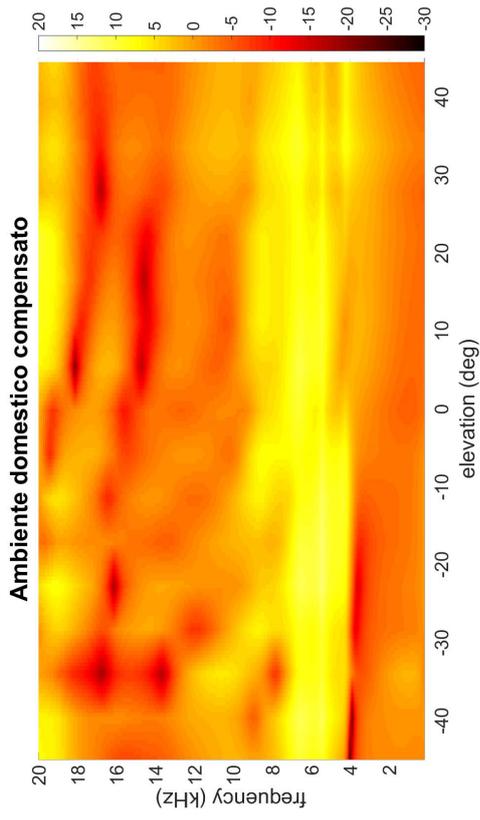
**Figura A.1: Risultati relativi ad ambiente controllato**



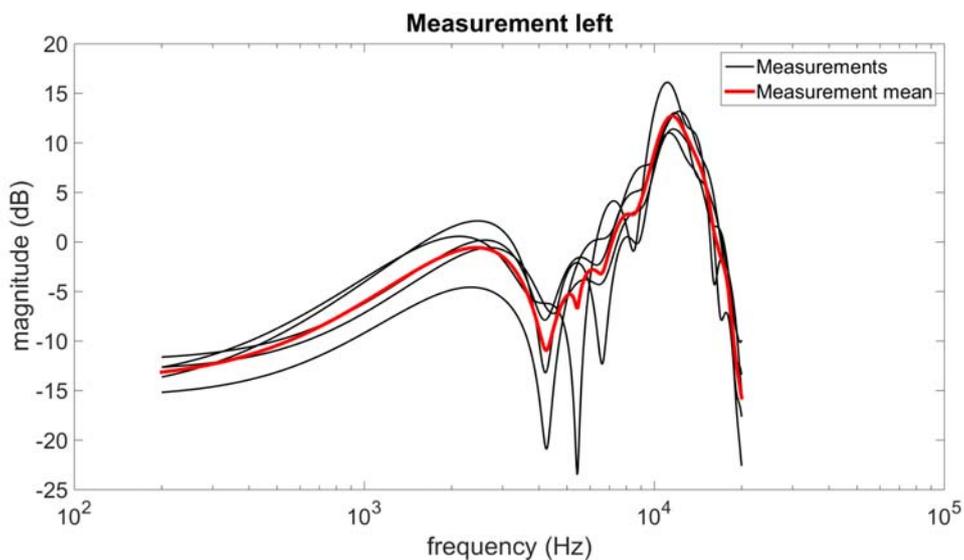
**Figura A.2:** Misurazione diffuse field acquisite in ambiente controllato con Zenfone



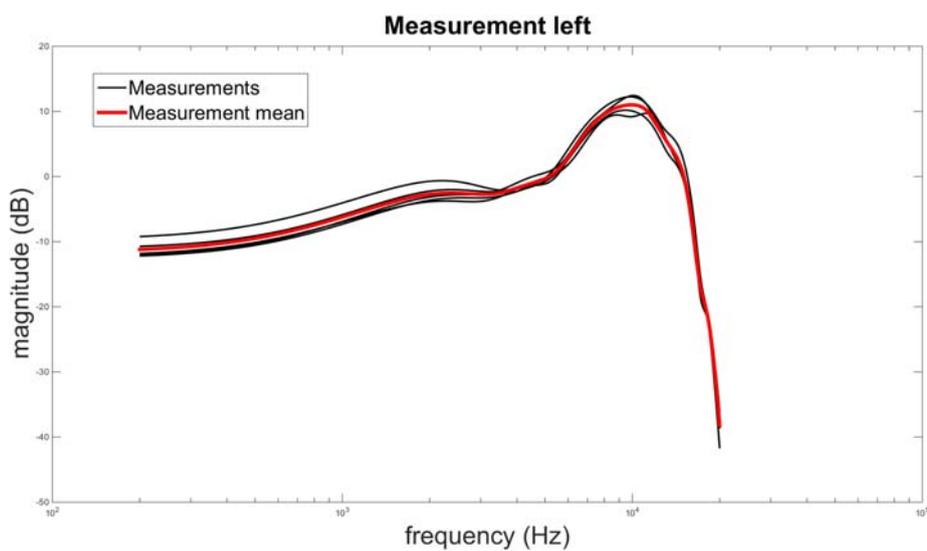
**Figura A.3:** Misurazione diffuse field acquisite in ambiente controllato con Wiko



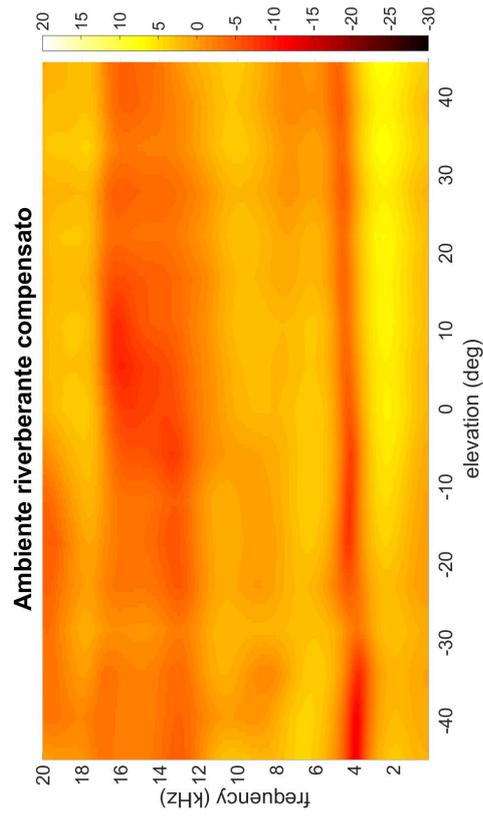
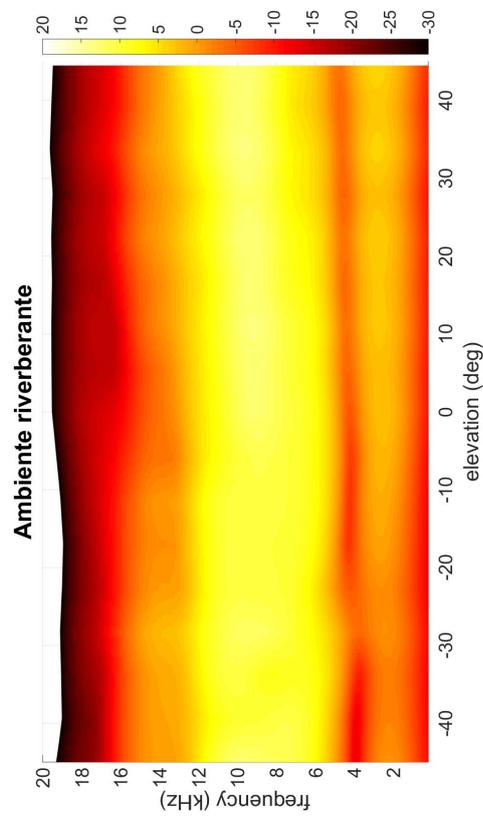
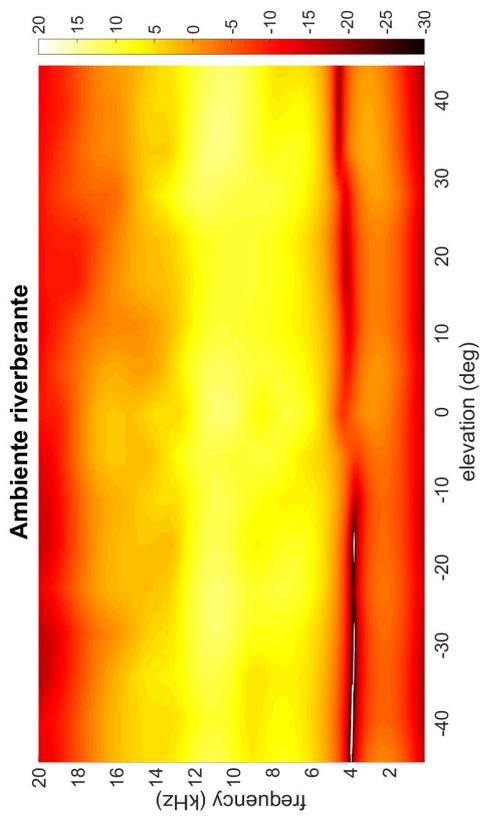
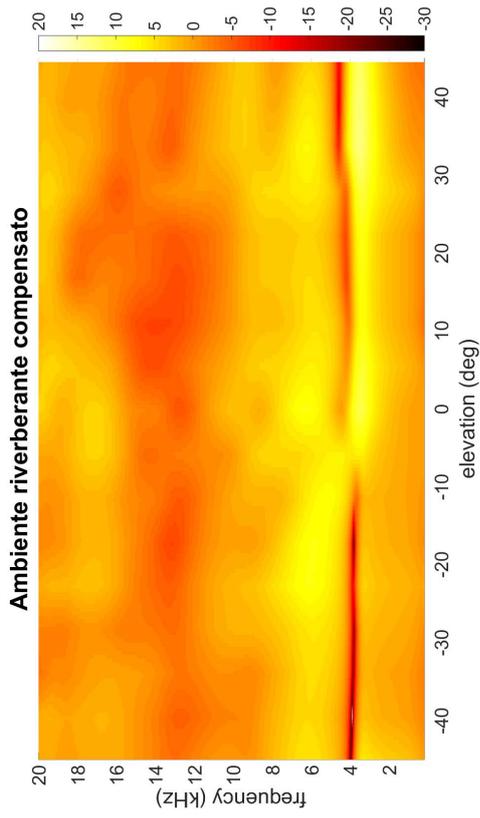
**Figura A.4:** Risultati relativi ad ambiente domestico



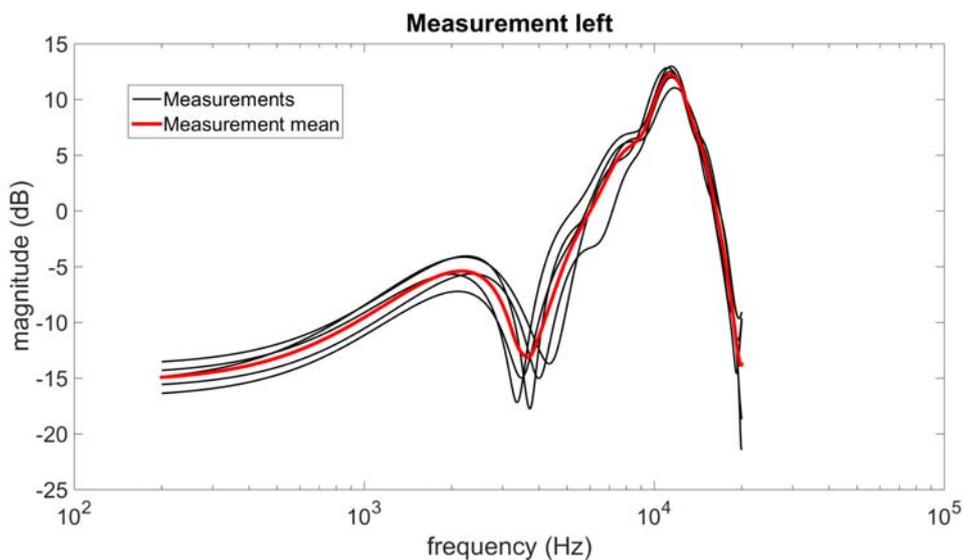
**Figura A.5:** Misurazione diffuse field acquisite in ambiente domestico con Zenfone



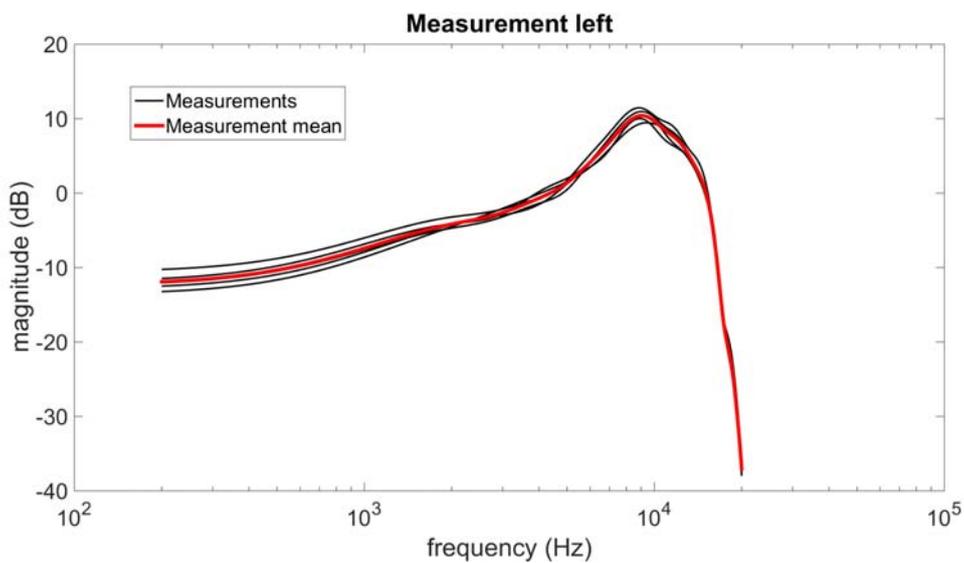
**Figura A.6:** Misurazione diffuse field acquisite in ambiente domestico con Wiko



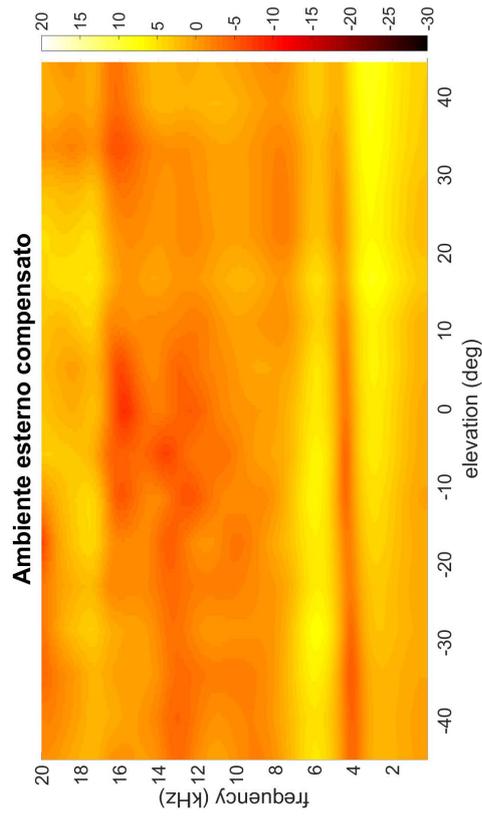
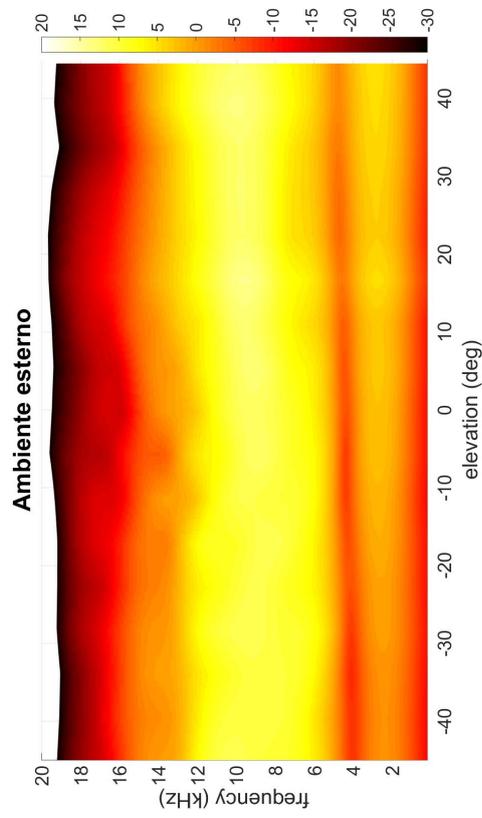
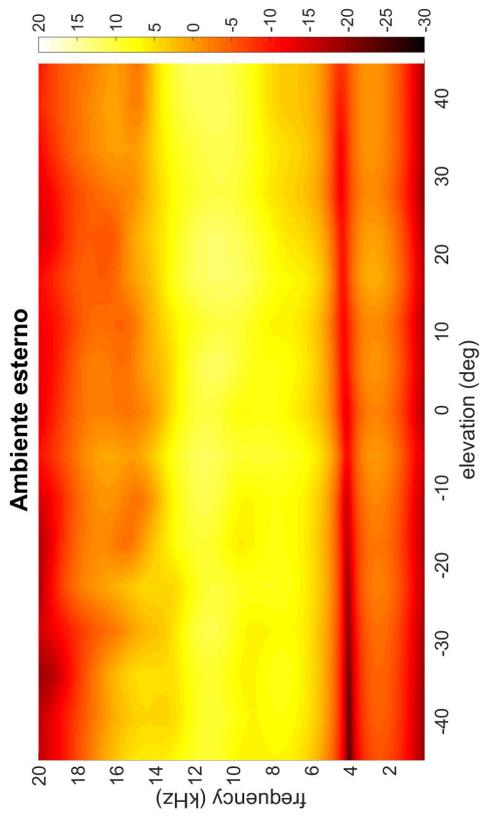
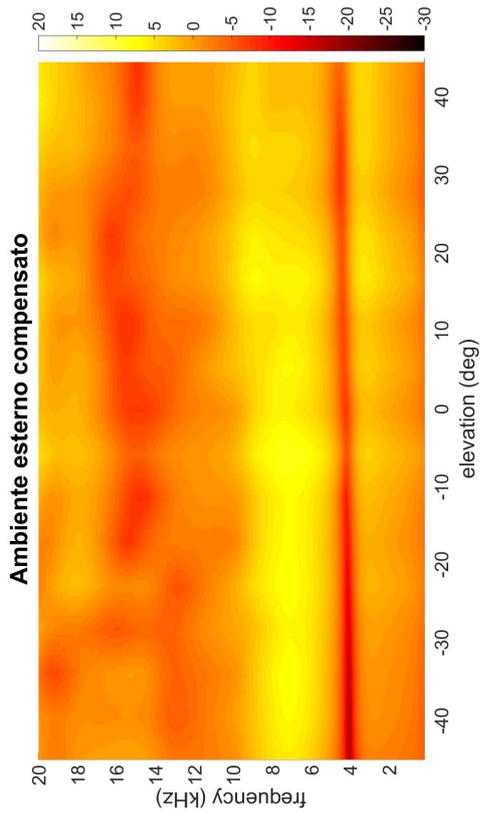
**Figura A.7:** Risultati relativi ad ambiente riverberante



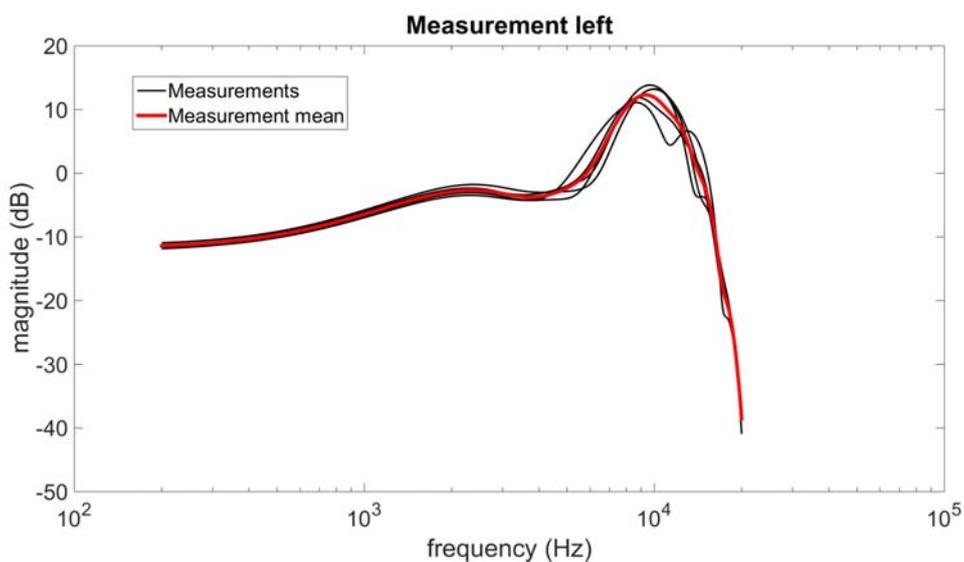
**Figura A.8:** Misurazione diffuse field acquisite in ambiente riverberante con Zenfone



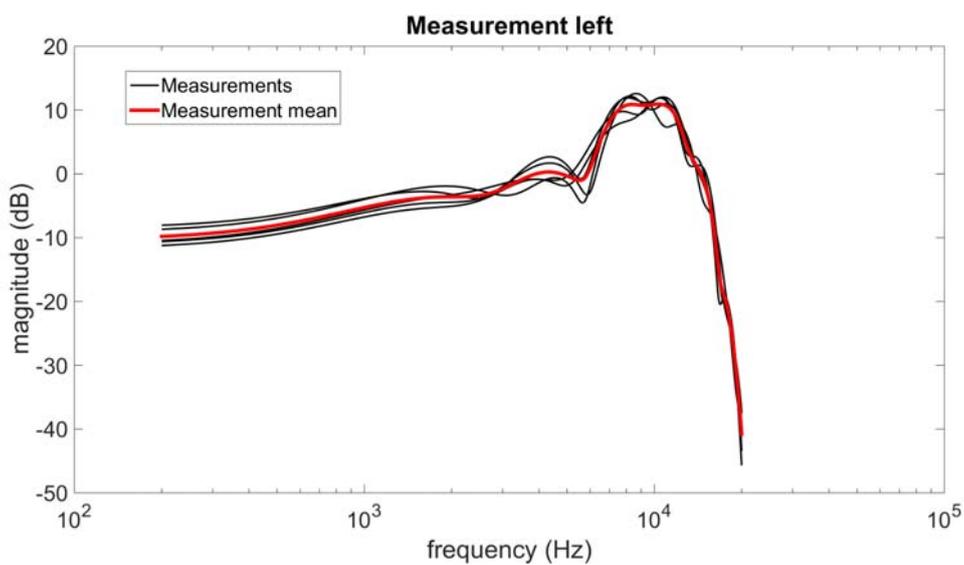
**Figura A.9:** Misurazione diffuse field acquisite in ambiente riverberante con Wiko



**Figura A.10: Risultati relativi ad ambiente esterno**



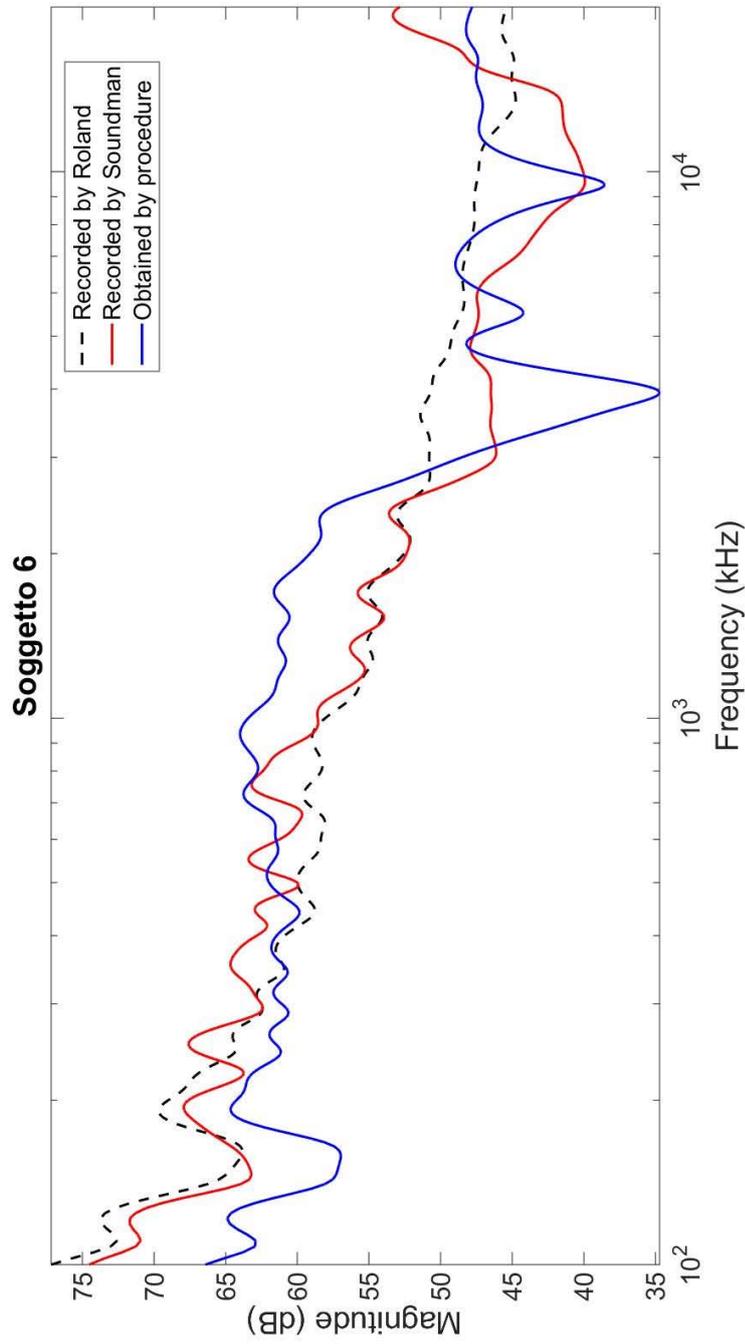
**Figura A.11:** Misurazione diffuse field acquisite in ambiente esterno con Zenfone



**Figura A.12:** Misurazione diffuse field acquisite in ambiente riverberante con Wiko

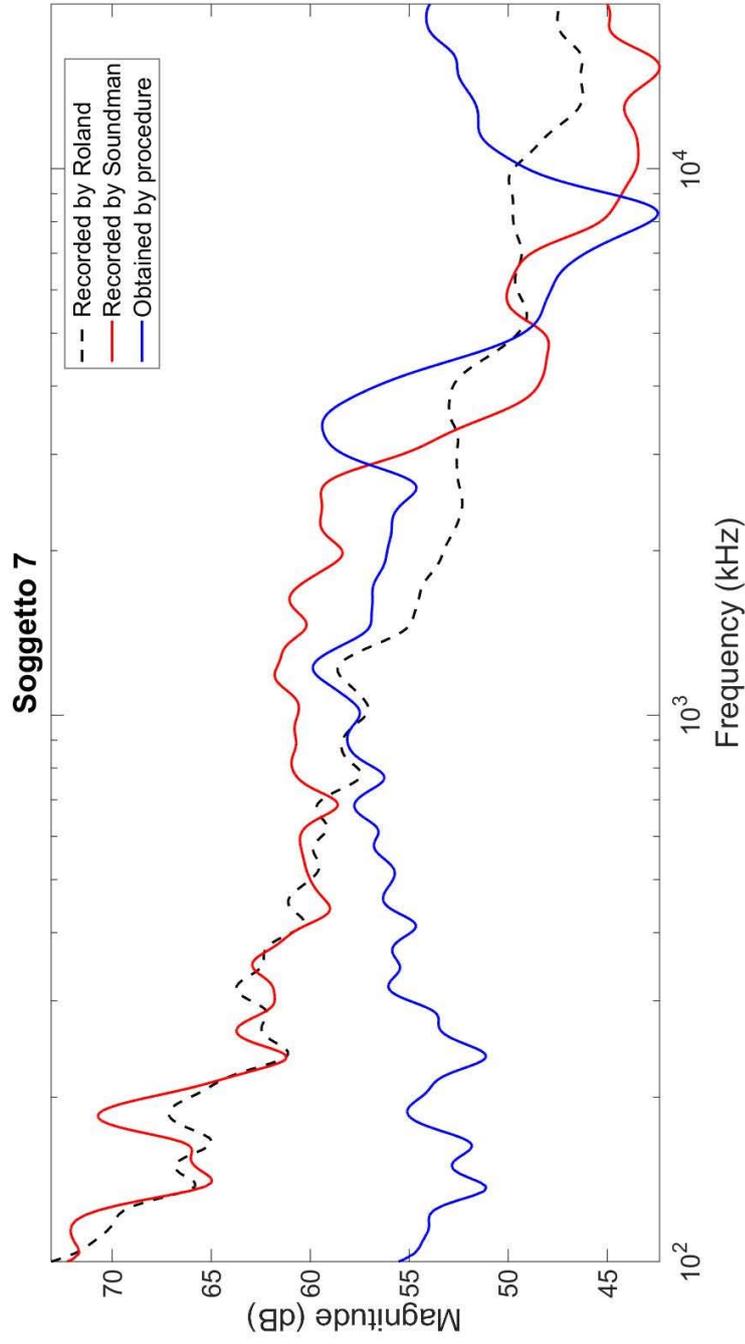
## **A.2 Risultati Cap 7**





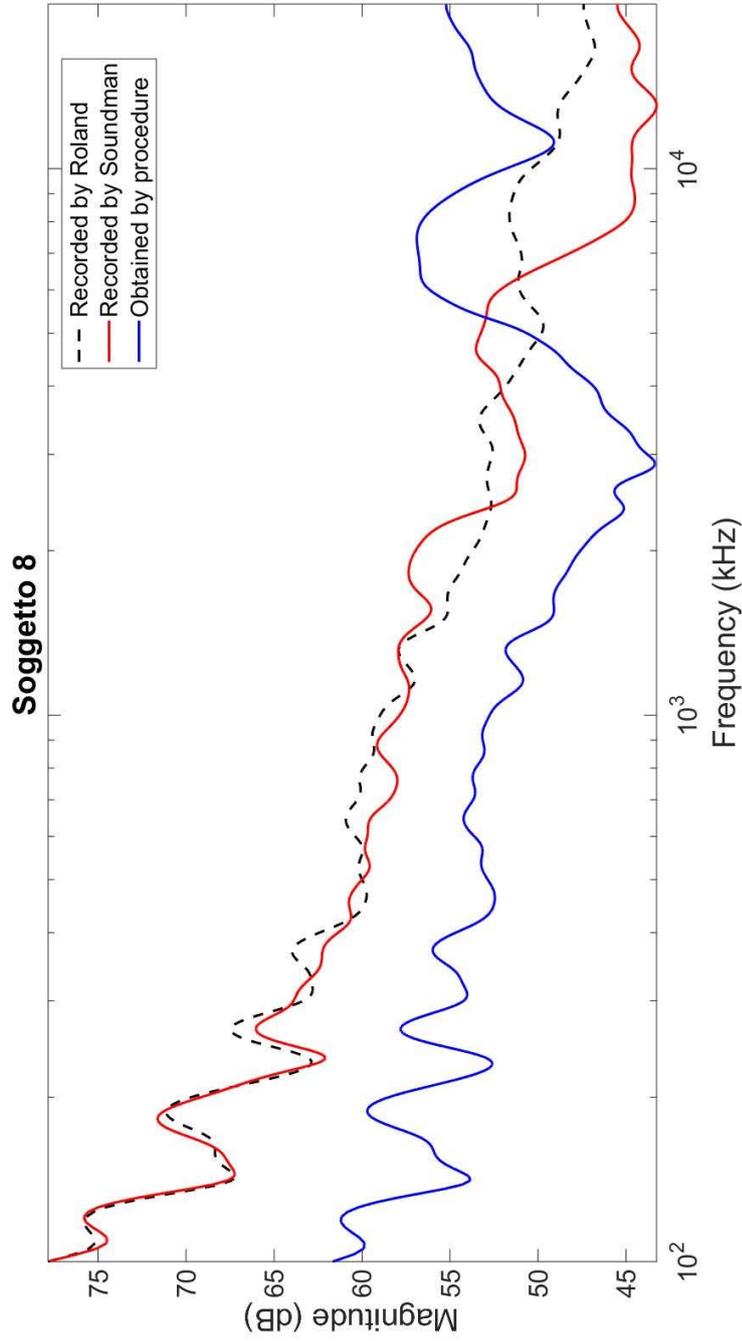
**Figura A.13:** Risultati relativi a Soggetto 6





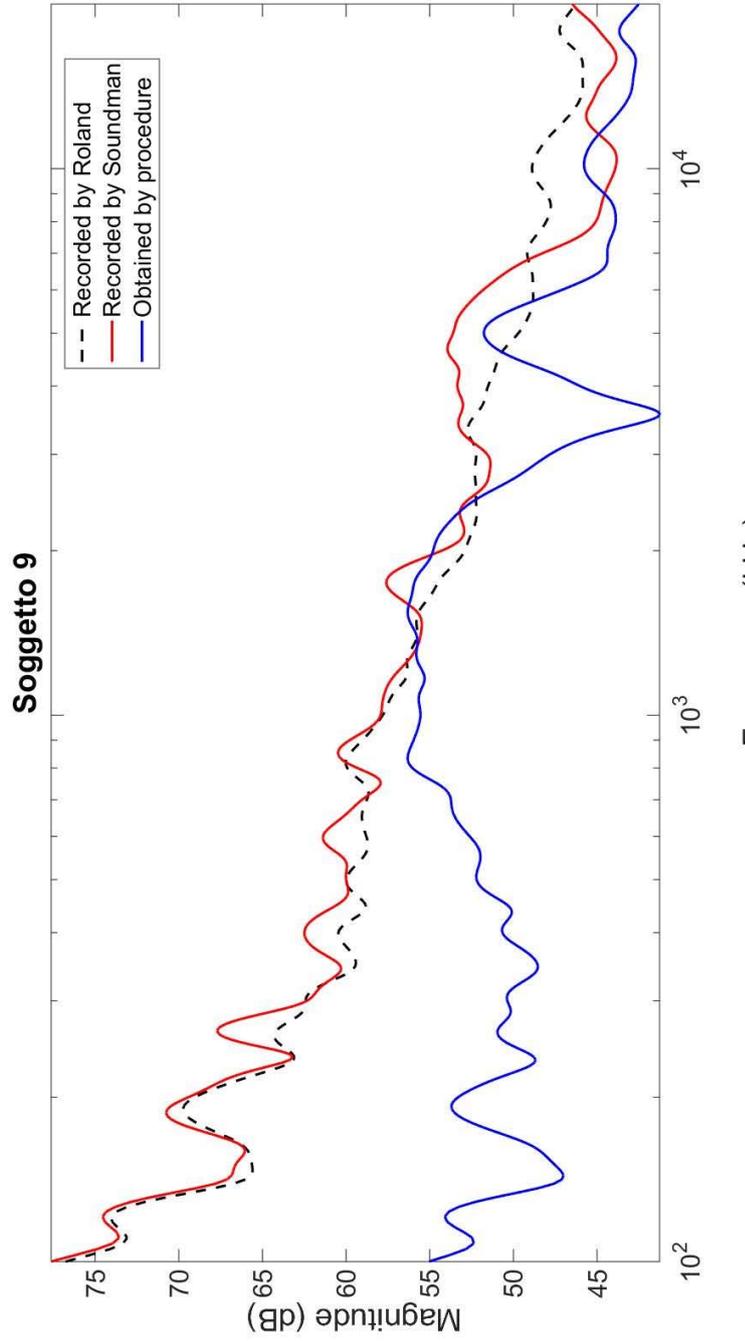
**Figura A.14:** Risultati relativi a Soggetto 7





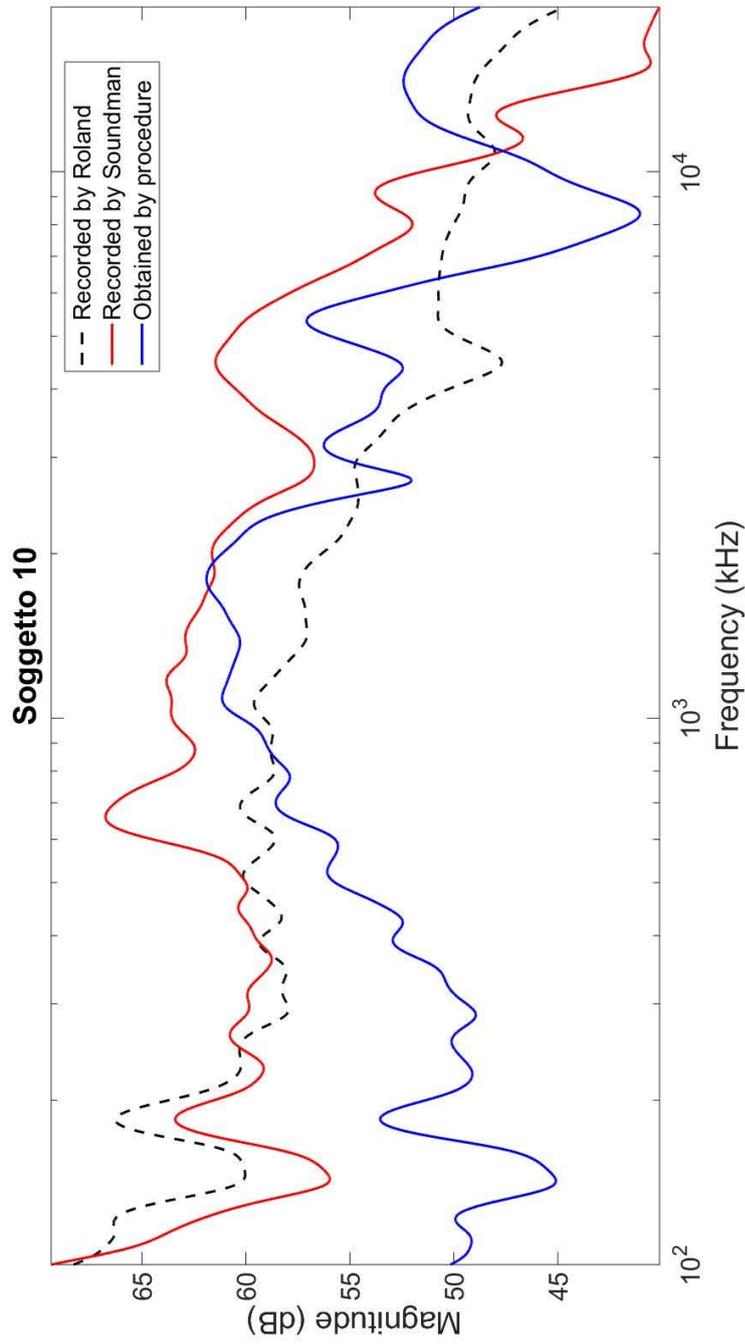
**Figura A.15:** Risultati relativi a Soggetto 8





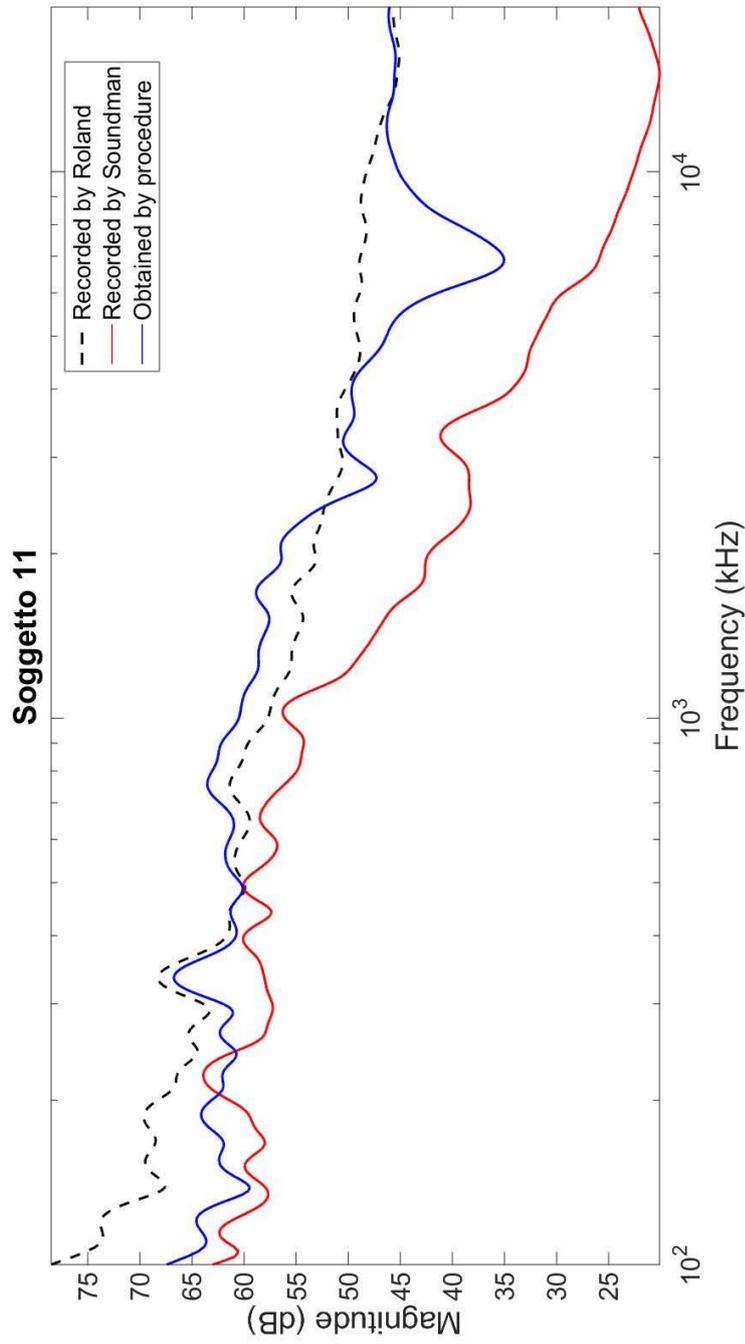
**Figura A.16:** Risultati relativi a Soggetto 9





**Figura A.17:** Risultati relativi a Soggetto 10





**Figura A.18:** Risultati relativi a Soggetto 11



# Appendice B

## Codice

### B.1 Matlab

Matlab è il software utilizzato per la realizzazione della GUI esposta nel capitolo 5 e di tutte le analisi del segnale effettuate in questa tesi. Si tratta di un linguaggio ad alto livello utilizzato in diversi ambiti, tra cui l'elaborazione dei segnali. Matlab consente di analizzare dati, sviluppare algoritmi e creare modelli integrando calcolo, visualizzazione e programmazione. Inoltre, mette a disposizione un ambiente grafico, chiamato GUIDE (GUI Design Environment) utile per la realizzazione di semplici ed immediate interfacce grafiche. Di grande utilità risultano i toolbox integrati in Matlab: in particolare il toolbox statistico, quello relativo alla manipolazione delle immagini e dell'audio. Tale linguaggio è risultato altresì adatto per l'analisi e la rappresentazione dei dati. Tutti i grafici riportati in questa tesi sono stati ottenuti mediante questo software.

### B.2 preProcessing.m

```
function preProcessing( nome )
%preProcessing
%   input: nome del soggetto
%
%   A partire dal nome del soggetto apre il relativo file
%   audio e lo separa in tre file audio separati con bande
%   [0, 2500]; [2500, 6000]; [6000, 20000] Hz
%
% @author      Giacomo Sorato
% @email      giacomo.sorato@studenti.unipd.it
% @insitute    Sound and Music Computing Group, Department of
%              Information Engineering, University of Padua
% @date       05th Oct 2016
```

```
%%%%% DEFINITION NAME
nome = 'name';

%% opening sweep
% path of sweep
    path = ['sweep_' nome '.wav'];

% opening sweep
    [s, Fs] = audioread(path);

%% Design a 8th-order lowpass Butterworth filter
% with a cutoff frequency of 2500 Hz
fc_lp = 2500;
[b1,a1] = butter(6,fc_lp/(Fs/2));
s_low = filter(b1,a1,s);

%% Design a 8th-order highpass Butterworth filter
% with a cutoff frequency of 6000 Hz
fc_hp = 6000;
[z,p,k] = butter(9,fc_hp/(Fs/2),'high');
[b2,a2] = zp2tf(z,p,k);
s_hig = filter(b2,a2,s);

%% Design a 8th-order Butterworth bandpass filter
% with cutoff frequency of 2500 and 6000 Hz.
[b3,a3] = butter(8,[fc_lp fc_hp]/(Fs/2));
s_mid = filter(b3,a3,s);

%% Saving audio file
audiowrite('proc_sweep1.wav',s_low,Fs);
audiowrite('proc_sweep2.wav',s_mid,Fs);
audiowrite('proc_sweep3.wav',s_hig,Fs);

end
```

## B.3 postProcessing.m

```
function [ output_args ] = postProcessing( args )
%postProcessing
%   input: none
%
%   Somma di tre file audio generati tramite la funzione preProcessing
%
% @author      Giacomo Sorato
% @email      giacomo.sorato@studenti.unipd.it
% @insitute    Sound and Music Computing Group, Department of
%              Information Engineering, University of Padua
% @date       05th Oct 2016

%% opening signals
[s_low, Fs] = audioread('filt_sweep1.wav');
[s_mid, Fs] = audioread('filt_sweep2.wav');
[s_hig, Fs] = audioread('filt_sweep3.wav');

%% overlap
s_left = s_low(:,1) + s_mid(:,1) + s_hig(:,1);
s_right = s_low(:,2) + s_mid(:,2) + s_hig(:,2);

s_filt = [s_left, s_right];

%% saving
audiowrite('filt_sweep.wav', s_filt, Fs);
end
```

## B.4 Filtri

### B.4.1 HPfilter.m

```
function [ output_args ] = HPfilter( cutOff, Fs)
% definition of a high pass filter

Fstop = 10;          % Stopband Frequency
Fpass = cutOff;     % Passband Frequency
if (Fpass~=0)
    Astop = 25;      % Stopband Attenuation (dB)
    Apass = 1;       % Passband Ripple (dB)

    hhp = fdesign.highpass('fst,fp,ast,ap', Fstop, Fpass, Astop, Apass, Fs);
```

```

hpFilt = design(hhp, 'equiripple', 'MinOrder', 'any', ...
               'StopbandShape', 'flat');

[H1, W1] = freqz(hpFilt);

plot(W1/pi*Fs, 20*log10(abs(H1)));
ax = gca;
ax.YLim = [-100 20];
ax.XLim = [0 40000];
ax.XTickLabel = 0:2.5:20;
xlabel('Frequency (KHz)')
ylabel('Magnitude (dB)')

else
    hpFilt=1;
end

end

```

## B.4.2 Pfilter.m

```

function [ output_args ] = Pfilter( centralF, Bandwidth, Fs )
% Definition of a second order peak filter
Np = 2;          % Order
F0p = centralF; % Center frequency
BWp = Bandwidth; % Bandwidth
if (BWp~=0)
    Ast = 20;
    hp = fdesign.peak('N,F0,BW,Ast', Np, F0p, BWp, Ast, Fs);
    pFilt = design(hp, 'cheby2', 'SOSScaleNorm', 'Linf');

    [H2, W2] = freqz(pFilt);

    %W2 = W2/pi*Fs;
    %H2 = 20*log10(abs(H2));

    plot(W2/pi*Fs, 20*log10(abs(H2)))
    ax = gca;
    ax.YLim = [-100 20];
    ax.XLim = [0 40000];
    ax.XTickLabel = 0:2.5:20;
    xlabel('Frequency (KHz)')
    ylabel('Magnitude (dB)')

```

```
else
    pFilt=1;
end
```

### B.4.3 Nfilter.m

```
function [ output_args ] = Pfilter( centralF, Bandwidth, Fs )
% notch 2 filter
Nn = 2;          % Order
F0n = centralF;  % Center frequency
BWn = Bandwidth; % Bandwidth
if (BWn~=0)
    hn = fdesign.notch('N,F0,BW', Nn, F0n, BWn, Fs);
    nFilt = design(hn, 'butter', 'SOSScaleNorm', 'Linf');

    [H3, W3] = freqz(nFilt);

    plot(W3/pi*Fs, 20*log10(abs(H3)))
    ax = gca;
    ax.YLim = [-100 20];
    ax.XLim = [0 40000];
    ax.XTickLabel = 0:2.5:20;
    xlabel('Frequency (KHz)')
    ylabel('Magnitude (dB)')

else
    pFilt=1;
end
```

## B.5 hrirsElaboration.m

```

function [ ir_window ir ] = Elaboration( data, sweep, Fs , norm )
%   Input
%       data      = signal to convert to impulse response
%       sweep     = sweep signal
%       Fs        = sample frequency
%       norm      = "1" to normalize
%
%   Output
%       ir = impulse response resulting from sweep signal
%
% @author      Giacomo Sorato
% @email       giacomo.sorato@studenti.unipd.it
% @insitute    Sound and Music Computing Group, Department of
%              Information Engineering, University of Padua
% @date        12th Apr 2016

%% padding
z = zeros([1 length(sweep)-length(data)]);
data = [data; z.'];

%% cross-correlation
onset = hrirs_onset_by_crosscorrelation( data, sweep );
% shift
data = cycshift( data, -onset );

%% IR extraction from sweep response
ir = IRSweep(data, sweep, Fs);

%% Windowing
L = round(0.5*Fs/1000);
ir_window = windowing(ir , 'hann', L);

%% 0s elimination from windowed IRs
ir_window = ir_window( find( ir_window ~= 0 ) );

%% Normalization
if norm
    ir_window = ir_window / max ( abs( ir_window ) );
end
end

```

## B.6 Ambiente.m

```

% AMBIENT
% @Author: Giacomo Sorato
% @email: giacomo.sorato@studenti.unipd.it
% @Date: 10th Nov 2016
% @Description:
%         Confronto tra:
%         - registrazione selfear in ambiente
%         - registrazione selfear in ambiente compensata
%

clear;
%% Path design

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
selector = 2; % 1=zenfone, 2=wiko
ambient = 'riverberante';

                                % controllato <- riferimento
                                % domestico
                                % esterno
                                % riverberante

Fss = 48000;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

if selector == 1
    elevation_path = ['..\ ' ambient '\SelfEar_' ambient '\'];
    diffuse_field_path = ['..\ ' ambient '\Df_' ambient '_data\'];
else
    elevation_path = ['..\ ' ambient '\SelfEar_' ambient '2\'];
    diffuse_field_path = ['..\ ' ambient '\Df_' ambient '2_data\'];
end

title = ['Ambiente ' ambient];

polang = -45:5.625:+45;
for i = 1:length(polang)
    elevation_str{i} = num2str(polang(i));
end

% numero di angoli
NElevation = length (elevation_str);

% fs resampling for uniformity
originalFs = 44100;

```

```

desiredFs = 48000;
[p,q] = rat(desiredFs / originalFs);

%% Audio reading
for i = 1:NElevation
    [rec, fs_rec] = audioread([elevation_path ...
                             elevation_str{i} '.wav']);
    if selector == 1
        rec = resample(rec,p,q);
    end
    elevation_wav{i} = rec;
end

%diffuse field procedura
for i = 1:5
    [df, fs_df] = audioread([diffuse_field_path 'Diff' ...
                             num2str(i) '.wav']);
    if selector == 1
        df = resample(df,p,q);
    end
    diffuse_field{i} = df;
end

% Sweep
[sweep1,Fs1] = audioread('sweep_48KHz_1s.wav');
if selector == 1
    [sweep2,Fs2] = audioread('sweep_48KHz_3s.wav');
else
    sweep2 = sweep1;
end

%% IR calculation

df_IR_window = [];
for i=1:5
    temp = diffuse_field{i};
    temp_i = temp(:,1);
    [df_IR_window_i df_IR_proc_i] = ...
        hrirs_elaboration_single(temp_i, sweep2, Fss, 1);
    df_IR_window = [df_IR_window; df_IR_window_i'];
end

IR_window = [];
for i = 1:NElevation
    temp = elevation_wav{i};

```

```

temp_i = temp(:,1);
[IR_window_i IR_i] = ...
    hrirs_elaboration_single(temp_i, sweep1, Fss, 1);
IR_window = [IR_window; IR_window_i'];

% converto gli angoli da stringhe a numeri
elevation_values{i} = str2num(elevation_str{i});
end

%% Initialize data for FR plot
Fmin = 200;
Fmax = 20000; % Min and max frequency in kHz
Q = Inf; % no smoothing Q = inf
log_scale = 1;
NFFT = 2*1000;

% diffuse field
[DFT_df dB_df phi_df sf_df] = freq_resp_tot(df_IR_window', ...
    Fmin, Fmax, Q, log_scale, NFFT, Fss);

% Average diffuse field
dB_df_avg = mean(dB_df.>');
semilogx(sf_df, dB_df_avg, 'k','LineWidth',1);

% Ir elevation
[DFT dB phi sf] = freq_resp_tot(IR_window',...
    Fmin, Fmax, Q, log_scale, NFFT, Fss);
semilogx(sf, dB(:,1), 'k','LineWidth',1);

%% plot
x = 1;
hrirs_plot_HRIR_by_azimuth2( dB, 0, sf, Fss, ...
    elevation_values, 0, [-45 +45], [Fmin Fmax], title, x );

title = [title ' compensato'];

hrirs_plot_HRIR_by_azimuth2( dB, dB_df_avg, sf, Fss, ...
    elevation_values, 0, [-45 +45], [Fmin Fmax], title, x+1 );

% hrirs_plot_HRIR_by_azimuth
% ( Irs
% [df_IR_w 0] 0 if don't want to use diffuse field
% Fs

```

```
% angles
% 0 = linear scale, 1 = log scale
% elevation range
% frequency range
% titolo del grafico )
```

## B.7 DfVariation.m

```

% KEMAR_CONTRIBUTE
% @Author: Giacomo Sorato
% @email: giacomo.sorato@studenti.unipd.it
% @Date: 10th Nov 2016
% @Description:
%         Patch per verificare il contributo dato dalla presenza di
%         qualcuno durante le misurazioni della diffuse field
%         Si confrontano le diffuse field calcolate alla distanza di un
%         metro in ambiente controllato con e senza la Kemar, per vedere
%         come questa influisce.
%
clear;

%% Initialization

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
selector = 2;
ambient = 'riverberante';
Fss = 48000;
N=5;
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

originalFs = 44100;
desiredFs = 48000;
[p,q] = rat(desiredFs / originalFs);

% path design
if selector == 1
    diffuse_field_path = ['..\ ' ambient '\Df_' ambient '_data\'];
else
    diffuse_field_path = ['..\ ' ambient '\Df_' ambient '2_data\'];
end

% elevation string
polang = -45:5.625:45;
for i = 1:length(polang)
    elevation_str{i} = num2str(polang(i));
end

% number of angles
NElevation = length (elevation_str);

%% Audio reading

```

```

for i = 1:N
    [df, fs_df] = audioread([diffuse_field_path 'Diff' ...

    if selector == 1
        df = resample(df,p,q);
    end
    diffuse_field{i} = df;
end

% Sweep
if selector == 1
    [sweep,Fs] = audioread('sweep_48KHz_3s.wav');
else
    [sweep,Fs] = audioread('sweep_48KHz_1s.wav');
end

%% IR calculation
df_IR_window = [];
df_IR_proc = [];
for i=1:N
    temp = diffuse_field{i};
    [df_IR_window_i df_IR_proc_i] = ...
        hrirs_elaboration_single(temp(:,1), sweep, Fss, 1);
    df_IR_window = [df_IR_window; df_IR_window_i'];
    df_IR_proc = [df_IR_proc; df_IR_proc_i'];
end

%% DFT calculation
Fmin = 200; Fmax = 20000; % Min and max frequency in kHz
Q = Inf; % no smoothing Q = inf
NFFT = 2000; % FFT lenght (oversampling)
log_scale = 1;

[DFT_l dB_l phi_l sf_l] = freq_resp_tot...
    (df_IR_window', Fmin, Fmax, Q, log_scale, NFFT, Fss);

%% AVG calculation
avg_dB_l = mean(dB_l.').';

%% SD calculation
SD = zeros(N,N);
for j = 1:N
    for i = 1:N
        SD(i,j) = compute_SD(dB_l(:,i), dB_l(:,j));
    end
end

```

```
end

%% Plot
figure (1);
p1 = semilogx(sf_l, dB_l(:,1), 'k','LineWidth',2);
if N > 1
    hold on;
    for i=2:N
        semilogx(sf_l, dB_l(:,i), 'k','LineWidth',2);
    end
    p2 = semilogx(sf_l, avg_dB_l, 'r','LineWidth',4);
    hold off;
end
h_legend1 = legend([p1 p2], 'Measurements', ...
                   'Measurement mean');

set(h_legend1,'FontSize',22);
title('Measurement left','FontSize',28);
xlabel('frequency (Hz)','FontSize',28);
ylabel('magnitude (dB)','FontSize',28);
```

## B.8 Reciprocity.m

```
% Reciproc
% @Author: Giacomo Sorato
% @email: giacomo.sorato@studenti.unipd.it
% @Date: 9th Nov 2016
% @Description:
%         Patch per verificare il principio di reciprocit
%
clear;

%% Path design
path = '..\reciprocit\Reciprocit_data\';

% stringa delle elevazioni
polang = -45:5.625:45;
for i = 1:length(polang)
    elevation_str{i} = num2str(polang(i));
end

% numero di angoli
NElevation = length (elevation_str);

%% Audio reading
for i = 1:5
    Dritti{i} = audioread([path 'Dritto' num2str(i) '.wav']);
end

for i = 1:5
    Rovesci{i} = audioread([path 'Rovescio' num2str(i) '.wav']);
end

DR = [Dritti Rovesci];

% Sweep
[sweep,Fs] = audioread('sweep_48KHz_3s.wav');

originalFs = Fs;
desiredFs = 44100;
[p,q] = rat(desiredFs / originalFs);
sweep=resample(sweep,p,q);

%% IR calculation
for i = 1:length(DR)
```

```

temp = DR(i);

[ir_windowL irL ir_windowR irR] = hrirs_elaboration ...

IR_window_L{i} = ir_windowL;
IR_window_R{i} = ir_windowR;
end

%% DFT calculation
Fmin = 50; Fmax = 20000;           % Min and max frequency in kHz
Q = Inf;                          % no smoothing Q = inf
NFFT = 2*1000;                    % FFT lenght (oversampling)
log_scale = 1;

for i=1:length(DR)
    [DFTL{i}, dBL{i}, phiL{i}, sFL{i}] = freq_resp_tot...
        ( IR_window_L{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
    [DFTR{i}, dBR{i}, phiR{i}, sFR{i}] = freq_resp_tot...
        ( IR_window_R{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
end

%% Avg

% Dritti left and right channels
dB_Dritti_L = [];
dB_Dritti_R = [];
for i = 1:(length(DR)/2)
    dB_Dritti_L = [dB_Dritti_L dBL{i}];
    dB_Dritti_R = [dB_Dritti_R dBR{i}];
end

% Rovesci left and right channels
dB_Rovesci_L = [];
dB_Rovesci_R = [];
for i = ((length(DR)/2)+1):length(DR)
    dB_Rovesci_L = [dB_Rovesci_L dBL{i}];
    dB_Rovesci_R = [dB_Rovesci_R dBR{i}];
end

% Avg of Dritti left and right channels
avg_dB_Dritti_L = mean(dB_Dritti_L.').' ;
avg_dB_Dritti_R = mean(dB_Dritti_R.').' ;

% Avg of Rovesci left and right channels
avg_dB_Rovesci_L = mean(dB_Rovesci_L.').' ;

```

```
avg_dB_Rovesci_R = mean(dB_Rovesci_R.').';

%% SD calculation

compute_SD(avg_dB_Dritti_L, avg_dB_Rovesci_L)
compute_SD(avg_dB_Dritti_R, avg_dB_Rovesci_R)

%% Plot
figure (1);
semilogx(sfL{1}, avg_dB_Dritti_L, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_Rovesci_L, 'r','LineWidth',3);
hold off;
h_legend1 = legend('Straight left', ...
                  'Opposite left');
set(h_legend1,'FontSize',28);
title('Measurement Left','FontSize',28);
xlabel('magnitude (dB)','FontSize',28);
ylabel('frequency (Hz)','FontSize',28);

figure (2);
semilogx(sfL{1}, avg_dB_Dritti_R, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_Rovesci_R, 'r','LineWidth',3);
hold off;
h_legend2 = legend('Straight right', ...
                  'Opposite right');
set(h_legend2,'FontSize',28);
title('Measurement Right','FontSize',28);
xlabel('magnitude (dB)','FontSize',28);
ylabel('frequency (Hz)','FontSize',28);
```

## B.9 SubjectContribute.m

```
% SUBJECT_CONTRIBUTE
% @Author: Giacomo Sorato
% @email: giacomo.sorato@studenti.unipd.it
% @Date: 10th Nov 2016
% @Description:
%     Patch per verificare il contributo dato dalla presenza di un
%     soggetto durante le misurazioni della diffuse field
%     Si confrontano le diffuse field calcolate alla distanza di un
%     70cm in ambiente controllato con e senza un soggetto, per vedere
%     come questo influisce.

clear;

%% Initialization

% path design
path1 = '..\reciprocit\Reciprocit_data\';
path2 = '..\controllato\Df_controllato_data\';

% elevation string
polang = -45:5.625:45;
for i = 1:length(polang)
    elevation_str{i} = num2str(polang(i));
end

% number of angles
NElevation = length (elevation_str);

% Sample frequencies are different, we need to uniform (resampling)
originalFs = 48000;
desiredFs = 44100;
[p,q] = rat(desiredFs / originalFs);

%% Audio reading

for i = 1:5
    [without_Sub{i} Fsws]= audioread([path1 'Dritto' ...

    [with_Sub{i} Fss] = audioread([path2 'Diff' ...

end

[sweep1,Fss1] = audioread('sweep_48KHz_1s.wav');
```

```

[sweep2,Fss2] = audioread('sweep_48KHz_3s.wav');
sweep2 = resample(sweep2,p,q);

DR = [without_Sub with_Sub];

%% IR calculation
for i = 1:length(DR)
    if i <= 5
        sweep = sweep2;
        Fs = 44100;
    else
        sweep = sweep1;
        Fs = 48000;
    end
    temp = DR(i);
    [ir_windowL irL ir_windowR irR] = hrirs_elaboration (...
                                                temp{1}, sweep, Fs, 1);

    IR_window_L{i} = ir_windowL;
    IR_window_R{i} = ir_windowR;
end

%% DFT calculation
Fmin = 50; Fmax = 20000; % Min and max frequency in kHz
Q = Inf; % no smoothing Q = inf
NFFT = 2*1000; % FFT lenght (oversampling)
log_scale = 1;

for i=1:10
    [DFTL{i}, dBL{i}, phiL{i}, sFL{i}] = freq_resp_tot( ...
        IR_window_L{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
    % dBspl_corrector_loud = 72.72 - dB_x(1);
    % dBL{i}=dBL_i+dBspl_corrector_loud;
    [DFTR{i}, dBR{i}, phiR{i}, sFR{i}] = freq_resp_tot( ...
        IR_window_R{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
end

%% AVG calculation

% Measurements without Kemar left and right channels
dB_NoSub_L = [dBL{1} dBL{2} dBL{3} dBL{4} dBL{5}];
dB_NoSub_R = [dBR{1} dBR{2} dBR{3} dBR{4} dBR{5}];

% Measurements with Kemar left and right channels
dB_SiSub_L = [dBL{6} dBL{7} dBL{8} dBL{9} dBL{10}];
dB_SiSub_R = [dBR{6} dBR{7} dBR{8} dBR{9} dBR{10}];

```

```
% Avg of measurements without Kemar left and right channels
avg_dB_NS_L = mean(dB_NoSub_L.'.').';
avg_dB_NS_R = mean(dB_NoSub_R.'.').';

% Avg of measurements with Kemar left and right channels
avg_dB_S_L = mean(dB_SiSub_L.'.').';
avg_dB_S_R = mean(dB_SiSub_R.'.').';

%% SD calculation
compute_SD(avg_dB_NS_L, avg_dB_S_L)
compute_SD(avg_dB_NS_R, avg_dB_S_R)

%% Plot
figure (1);
semilogx(sfL{1}, avg_dB_NS_L, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_S_L, 'r','LineWidth',3);
hold off;
h_legend1 = legend('Measurement without soubject left', ...
                  'Measurement with soubject left');
set(h_legend1,'FontSize',22);
title('Measurement left','FontSize',28);
xlabel('frequency (Hz)','FontSize',24);
ylabel('magnitude (dB)','FontSize',24);

figure (2);
semilogx(sfL{1}, avg_dB_NS_R, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_S_R, 'r','LineWidth',3);
hold off;
h_legend2 = legend('Measurement without subject right', ...
                  'Measurement with subject right');
set(h_legend2,'FontSize',22);
title('Measurement right','FontSize',28);
xlabel('frequency (Hz)','FontSize',24);
ylabel('magnitude (dB)','FontSize',24);
```

## B.10 KemarContribute.m

```

% KEMAR_CONTRIBUTE
% @Author: Giacomo Sorato
% @email: giacomo.sorato@studenti.unipd.it
% @Date: 10th Nov 2016
% @Description:
%         Patch per verificare il contributo dato dalla presenza della
%         KEMAR durante le misurazioni della diffuse field
%         Si confrontano le diffuse field calcolate alla distanza di un
%         metro in ambiente controllato con e senza la Kemar, per vedere
%         come questa influisce.
%
clear;

%% Initialization

% path design
path1 = '..\reciprocit\Reciprocit_data\';
path2 = '..\controllato\Df_controllato_Kemar_data\';

% elevation string
polang = -45:5.625:45;
for i = 1:length(polang)
    elevation_str{i} = num2str(polang(i));
end

% number of angles
NElevation = length (elevation_str);

%% Audio reading
for i = 1:5
    [without_Kemar{i} Fss] = audioread([path1 'Dritto' ...

    with_Kemar{i} = audioread([path2 'DiffK' num2str(i) '.wav']);
end

DR = [without_Kemar with_Kemar];

% Sweep
[sweep,Fs] = audioread('sweep_48KHz_3s.wav');

% Sample frequencies are different, we need to uniform (resampling)
originalFs = Fs;
desiredFs = 44100;

```

```

[p,q] = rat(desiredFs / originalFs);
sweep = resample(sweep,p,q);

%% IR calculation
for i = 1:length(DR)
    temp = DR(i);
    [ir_windowL irL ir_windowR irR] = ...
        hrirs_elaboration (temp{1}, sweep, Fs, 1);
    IR_window_L{i} = ir_windowL;
    IR_window_R{i} = ir_windowR;
end

%% DFT calculation
Fmin = 50; Fmax = 20000; % Min and max frequency in kHz
Q = Inf; % no smoothing Q = inf
NFFT = 2*1000; % FFT lenght (oversampling)
log_scale = 1;

for i=1:10
    [DFTL{i}, dBL{i}, phiL{i}, sFL{i}] = freq_resp_tot( ...
        IR_window_L{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
    [DFTR{i}, dBR{i}, phiR{i}, sFR{i}] = freq_resp_tot( ...
        IR_window_R{i}, Fmin, Fmax, Q, log_scale, NFFT, Fs);
end

%% AVG calculation

% Measurements without Kemar left and right channels
dB_NoKemar_L = [dBL{1} dBL{2} dBL{3} dBL{4} dBL{5}];
dB_NoKemar_R = [dBR{1} dBR{2} dBR{3} dBR{4} dBR{5}];

% Measurements with Kemar left and right channels
dB_SiKemar_L = [dBL{6} dBL{7} dBL{8} dBL{9} dBL{10}];
dB_SiKemar_R = [dBR{6} dBR{7} dBR{8} dBR{9} dBR{10}];

% Avg of measurements without Kemar left and right channels
avg_dB_NK_L = mean(dB_NoKemar_L.')';
avg_dB_NK_R = mean(dB_NoKemar_R.')';

% Avg of measurements with Kemar left and right channels
avg_dB_K_L = mean(dB_SiKemar_L.')';
avg_dB_K_R = mean(dB_SiKemar_R.')';

%% SD calculation
compute_SD(avg_dB_NK_L, avg_dB_K_L)

```

```
compute_SD(avg_dB_NK_R, avg_dB_K_R)

%% Plot
figure (1);
semilogx(sfL{1}, avg_dB_NK_L, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_K_L, 'r','LineWidth',3);
hold off;
h_legend1 = legend('Measurement without KEMAR left',...
                  'Measurement with KEMAR left');

set(h_legend1,'FontSize',28);
title('Measurement left','FontSize',28);
xlabel('frequency (Hz)','FontSize',28);
ylabel('magnitude (dB)','FontSize',28);

figure (2);
semilogx(sfL{1}, avg_dB_NK_R, 'k','LineWidth',3);
hold on;
semilogx(sfL{1}, avg_dB_K_R, 'r','LineWidth',3);
hold off;
h_legend2 = legend('Measurement without KEMAR right', ...
                  'Measurement with KEMAR right');

set(h_legend2,'FontSize',28);
title('Measurement right','FontSize',28);
xlabel('frequency (Hz)','FontSize',28);
ylabel('magnitude (dB)','FontSize',28);
```

# Bibliografia

- [1] Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. Acoustic selfies for extraction of external ear features in mobile audio augmented reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, pages 23–26. ACM, 2016.
- [2] Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. The selfear project: a mobile application for low-cost pinna-related transfer function acquisition. *SMC conference*, 2016.
- [3] Michele Geronazzo, Jacopo Fantin, Giacomo Sorato, Guido Baldovino, and Federico Avanzini. Selfie acustiche con il progetto selfear: un’applicazione mobile per l’acquisizione a basso costo di pinna-related transfer function. *XXI CIM conference*, 2016.
- [4] Robert Albrecht, Tapio Lokki, and Lauri Savioja. A mobile augmented reality audio system with binaural microphones. In *Proceedings of Interacting with Sound Workshop: Exploring Context-Aware, Local and Social Audio Applications*, IwS ’11, pages 7–11, Stockholm, Sweden, 2011. ACM.
- [5] Ville Riku Johannes Riikonen. User-related acoustics in a two-way augmented reality audio system. mastersthesis, Helsinki University of Technology, 2008.
- [6] Michele Geronazzo, Fabrizio Granza, Simone Spagnol, and Federico Avanzini. A Standardized Repository of Head-Related and Headphone Impulse Response Data. In *Proc. 134th Conv. Audio Eng. Society*, Rome, Italy, May 2013.
- [7] J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localization*. MIT Press, Cambridge, MA, USA, 1983.
- [8] F. Avanzini and G. De Poli. *Algorithms for Sound and Music Computing*. Creative Commons, 2009.

- [9] Henrik Mller, Michael Srensen, Jensen Friis, Boje Clemen, and Dorte Hammershi. Binaural Technique: Do We Need Individual Recordings? *J. Audio Eng. Soc.*, 44(6):451–469, 1996.
- [10] Carlos Avendano, V. Ralpgh Algazi, and Richard O. Dura. A head-and- torso model for low-frequency binaural elevation effects. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pages 179–182, 1999.
- [11] V. Ralpgh Algazi and Richard O. Dura. Headphone-based spatial sound. In *Signal Processing Magazine*, page 28(1):3342, 2011.
- [12] Nathaniel I. Durlach, Barbare G. Shinn-Cunningham, and Richard M. Held. Supernormal auditory localization. In *Presence*, page 2(2):89103, 1993.
- [13] Alex D. Little, Donald H. Mershon, and Patrick H. Cox. Spectral content as a cue to perceived auditory distance. In *Perception*, page 21(3):405416, 1992.
- [14] Soren H. Nielsen. Auditory distance perception in different rooms. In *In Audio Engineering Society Convention 92*, 1992.
- [15] C. I. Cheng and G. H. Wakefield. Introduction to Head-Related Transfer Functions (HRTFs): Representations of HRTFs in Time, Frequency, and Space. *J. Audio Eng. Soc.*, 49(4):231–249, April 2001.
- [16] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery. In *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, pages 1–8, Santorini, Greece, July 2013.
- [17] W. G. Gardner and K. D. Martin. HRTF Measurements of a KEMAR. *J. of the Acoustical Society of America*, 97(6):3907–3908, June 1995.
- [18] Wolfgang Kreuzer, Piotr Majdak, and Zhengsheng Chen. Fast multipole boundary element method to calculate head-related transfer functions for a wide frequency range. *The Journal of the Acoustical Society of America*, 126(3):1280–1290, 2009.
- [19] V. R. Algazi, R. O. Duda, D. M. Thompson, and C. Avendano. The CIPIC HRTF Database. In *Proc. IEEE Work. Appl. Signal Process., Audio, Acoust.*, pages 1–4, New Paltz, New York, USA, October 2001.
- [20] Michele Geronazzo, Simone Spagnol, Alberto Bedin, and Federico Avanzini. Enhancing Vertical Localization with Image-guided Selection of Non-individual

- Head-Related Transfer Functions. In *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2014)*, pages 4496–4500, Florence, Italy, May 2014.
- [21] Ronald T. Azuma. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):355–385, August 1997.
- [22] Nadine B. Sarter. Multimodal information presentation: Design guidance and research challenges. *International Journal of Industrial Ergonomics*, 36(5):439 – 445, 2006. Cognitive Engineering Insights for Human Performance and Decision Making Cognitive Engineering Insights for Human Performance and Decision Making.
- [23] Andrew J. Kolarik, Brian C. J. Moore, Pavel Zahorik, Silvia Cirstea, and Shahina Pardhan. Auditory distance perception in humans: a review of cues, development, neuronal bases, and effects of sensory loss. *Attention, Perception, & Psychophysics*, 78(2):373–395, 2016.
- [24] P. Hiselius. Attenuation of earplugs - objective predictions compared to subjective reat measurements. pages 764–770. *Acta Acustica united with Acustica*, 2005.
- [25] M. Tikander, A. Harma, and M. Karjalainen. Binaural positioning system for wearable augmented reality audio. pages 153–156, October 2003.
- [26] V. Valimaki, A. Franck, J. Ramo, H. Gamper, and L. Savioja. Assisted Listening Using a Headset: Enhancing audio perception in real, augmented, and virtual environments. *IEEE Signal Processing Magazine*, 32(2):92–99, March 2015.
- [27] Lauri Savioja and U. Peter Svensson. Overview of geometrical room acoustic modeling techniques. *The Journal of the Acoustical Society of America*, 138(2):708–730, August 2015.
- [28] Braxton Boren, Michele Geronazzo, Fabian Brinkmann, and Edgar Choueiri. Coloration Metrics for Headphone Equalization. In *Proc. of the 21st Int. Conf. on Auditory Display (ICAD 2015)*, pages 29–34, Graz, Austria, July 2015.
- [29] W. Owen Brimijoin, Alan W. Boyd, and Michael A. Akeroyd. The Contribution of Head Movement to the Externalization and Internalization of Sounds. *PLoS ONE*, 8(12):e83068, December 2013.
- [30] Naraji Sakamoto, Toshiyuki Gotoh, and Yoichi Kimura. On -Out-of-Head Localization- in Headphone Listening. *Journal of the Audio Engineering Society*, 24(9):710–716, November 1976.

- [31] J. S. Bradley and G. A. Soulodre. Objective measures of listener envelopment. *The Journal of the Acoustical Society of America*, 98(5):2590–2597, November 1995.
- [32] Frederic L. Wightman and Doris J. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. *The Journal of the Acoustical Society of America*, 85(2):868–878, 1989.
- [33] Tapio Lokki and Ilkka Huhtakallio. Measurements on active earplugs and effect of ear canal resonance on spectral balance. In *EuroNoise 2015*, page 21(3):405416, 2015.
- [34] Miikka Tikander. Sound quality of an augmented reality audio headset. In *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx05)*, page 21(3):405416.
- [35] Antoine Bernier and Jeremie Voix. Active musician’s hearing protection device for enhanced perceptual comfort. In *EuroNoise 2015*.
- [36] Robert Albrecht and Tapio Lokki. Electronic hearing protection for musicians. In *AES Headphone tech conference*, 2016.
- [37] Aki Hrm, Julia Jakka, Miikka Tikander, Matti Karjalainen, Tapio Lokki, Jarmo Hiipakka, and Gatan Lorho. Augmented Reality Audio for Mobile and Wearable Appliances. *J. Audio Eng. Soc.*, 52(6):618–639, 2004.
- [38] Michele Geronazzo. The external ear acoustics: a mixed structural modeling approach in virtual auditory displays. *J. of the Italian Society of Acoustics (RIA)*, 39(1):32–48, 2015.
- [39] Futoshi Asano, Yoiti Suzuki, and Toshio Sone. Role of spectral cues in median plane localization. *The Journal of the Acoustical Society of America*, 88(1):159–168, 1990.
- [40] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Estimation and Modeling of Pinna-Related Transfer Functions. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, pages 431–438, Graz, Austria, September 2010.
- [41] Sebastian Prepelitã, Michele Geronazzo, Federico Avanzini, and Lauri Savioja. Influence of voxelization on finite difference time domain simulations of head-related transfer functions. *The Journal of the Acoustical Society of America*, 139(5):2489–2504, May 2016.
- [42] Simone Spagnol, Michele Geronazzo, and Federico 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.

- 
- [43] Flemming Christensen, Pablo F. Hoffmann, and Dorte Hammershi. Measuring Directional Characteristics of In-Ear Recording Devices. In *In Proc. Audio Engineering Society Convention 134*. Audio Engineering Society, May 2013.
- [44] *Pure Data Floss manual*.